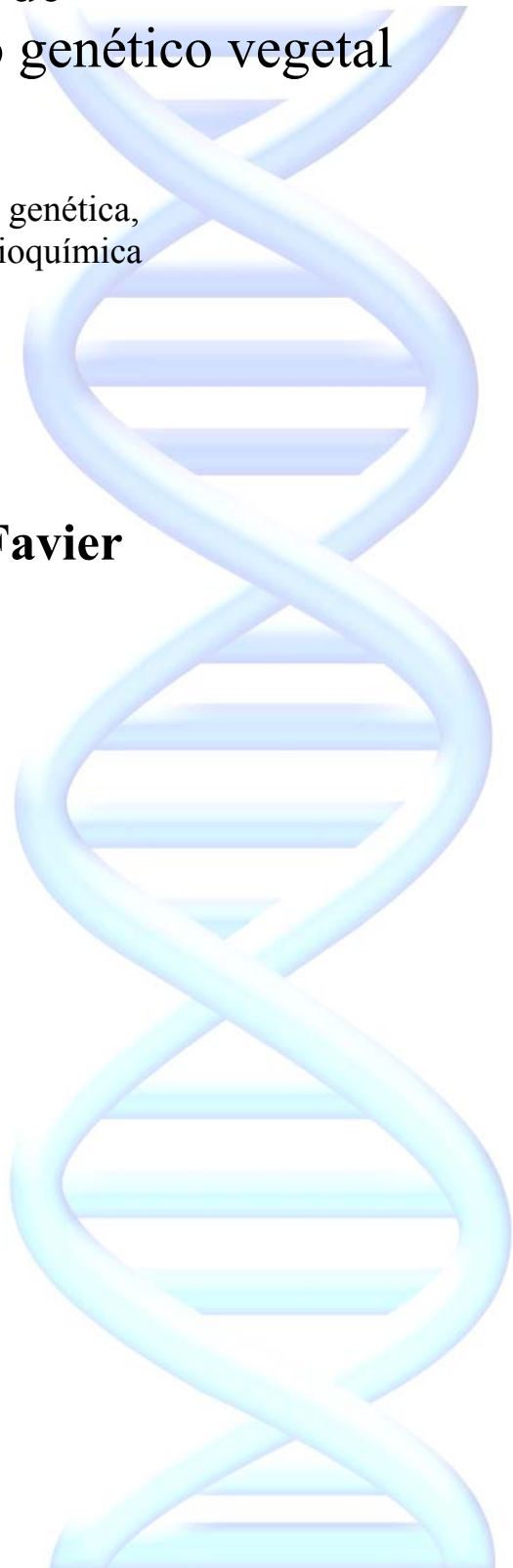


Fundamentos de mejoramiento genético vegetal

Conceptos básicos de genética,
biología molecular, bioquímica
y fisiología vegetal

Axel Tiessen Favier



Fundamentos de mejoramiento genético vegetal

PRIMERA EDICIÓN EAE

24 de Julio del 2012

Editorial EAE

ISBN: 978-3-8484-6841-6

1. Plantas – Maíz
2. Fitomejoramiento – Cultivos
3. Biotecnología – Ingeniería Genética
4. Genómica – Marcadores moleculares
5. Estadística – Bioinformática
6. Fisiología vegetal – Bioquímica

Editor



Dr. Axel Tiessen Favier

**Profesor-Investigador del
Departamento de Ingeniería Genética
CINVESTAV Irapuato
México**

www.ira.cinvestav.mx

atiessen@ira.cinvestav.mx

Tabla de Contenido

1. Notas introductorias	7
1.1. Presentación y enfoque	7
2. Conceptos básicos	9
2.1. Fundamentos de la genética.....	9
2.1.1. Genética vegetal	9
2.1.2. Factores que determinan la productividad agrícola	10
2.1.3. Del genotipo al fenotipo	14
2.1.4. Genética <i>forward</i> y genética reversa	24
2.2. Biodiversidad	25
2.2.1. Diversidad biológica	27
2.2.2. Diversidad taxonómica	32
2.2.3. Diversidad agrícola	33
2.2.4. Diversidad del maíz	34
2.3. Origen de la variabilidad genética	35
2.3.1. Mutaciones	35
2.3.2. Recombinación	39
2.3.3. Transferencia horizontal	42
2.3.4. Selección	44
2.4. El azar	46
2.4.1. Un fenómeno universal	46
2.4.2. Fuerza innovadora y fuente de diversidad	47
2.5. Evolución	49
2.5.1. Evolución biológica	49
2.5.2. Evolución dirigida	58
2.6. Filosofía científica del mejoramiento genético	61
2.6.1. Mejoramiento: Evolución en acción	61
2.6.2. Balance entre creatividad y eficiencia	67
2.6.3. Ciencia básica y aplicada	70
2.6.4. Combinación de ciencia, ingeniería y filosofía	72
2.7. Los ácidos nucleicos	75
2.7.1. Nociones básicas del DNA	75
2.7.2. Nociones avanzadas del DNA	91
3. Herramientas moleculares	97
3.1. PCR.....	97
3.1.1. Historia y principios	97
3.1.2. Técnicas y aplicaciones	101
3.2. Modificadores del DNA	101
3.2.1. Enzimas de restricción	102
3.2.2. Ligasas	102
3.2.3. Recombinasas	103
3.3. Métodos de secuenciación	106
3.3.1. Sanger	106

IV

3.3.2.	Pirosecuenciación	107
3.3.3.	Método de apareamiento y ligación	113
3.3.4.	Método de síntesis con fluoróforos	124
4.	Detección de polimorfismos genéticos	125
4.1.	Marcadores.....	125
4.1.1.	Características de los marcadores	125
4.1.2.	Tipos de marcadores genéticos	126
4.1.3.	Ejemplos de marcadores	127
4.1.4.	Estrategias de genotipaje	137
4.2.	Técnicas de comparación genómica.....	148
4.2.1.	<i>Differential Display</i> (DD)	148
4.2.2.	<i>Representational Difference Analysis</i> (RDA)	149
4.2.3.	Genome Mismatch Scanning (GMS)	152
4.3.	Ejemplos de aplicaciones.....	153
4.3.1.	Demostración de Pedigrí	153
4.4.	Haplotipos.....	154
4.4.1.	Desequilibrio por ligamiento	154
4.4.2.	Proyecto HapMap	155
4.5.	Pros y contras de las herramientas moleculares	155
4.5.1.	Ventajas de las herramientas moleculares	155
4.5.2.	Limitaciones de las herramientas moleculares	156
5.	Genómica y regulación	161
5.1.	Antecedentes de la genómica	161
5.2.	Estrategias de ensamblaje genómico	162
5.3.	Ejemplos de proyectos de genómica	165
5.3.1.	Genomas vegetales	167
5.3.2.	Genomas animales	177
5.3.3.	DNA extranuclear	182
5.3.4.	Genomas comparativos	190
6.	Herramientas bioinformáticas	191
6.1.	¿Qué es la bioinformática?	191
6.2.	Análisis de secuencias	194
6.2.1.	BLAST	194
6.3.	Bases de datos.....	198
6.3.1.	Genbank y el NCBI	198
6.3.2.	Maize GDB	200
6.3.3.	Maize Gene Index	201
6.3.4.	CIMMYT	202
6.3.5.	IRRI	203
6.4.	Paquetes de Software.....	204
6.4.1.	Fieldbook para maíz	204
6.4.2.	Alternativas gratuitas de software	205
7.	Herramientas estadísticas	209
7.1.	Conceptos básicos de estadística	209
7.1.1.	¿Para que sirve la estadística?	209

7.1.2.	Medidas centrales	209
7.1.3.	Desviación estándar y distribución normal	211
7.1.4.	Tipos de errores	212
7.2.	Diseño experimental	213
7.3.	Análisis estadísticos	216
7.3.1.	Ejemplo de análisis usando pruebas de T	216
7.3.2.	Ejemplo de análisis usando ANOVA	222
7.3.3.	Ejemplo: varianza, heredabilidad y repetibilidad	228
7.3.4.	Ejemplo de gráfica	237
7.3.5.	Ejemplo de análisis de bloques incompletos	237
7.3.6.	Ejemplo de análisis espacial	238
7.3.7.	Ejemplo de análisis con modelos lineales	238
7.4.	Software para estadística	238
7.4.1.	Uso de Excel	238
7.4.2.	R para estadística	241
8.	Biología de Sistemas	270
8.1.	Introducción	270
8.1.1.	Justificación de las ciencias bioquímicas	271
8.2.	Genómica	273
8.3.	Transcriptómica	274
8.4.	Proteómica	275
8.5.	Metabolómica	276
8.6.	Del genotipo al fenotipo	278
9.	Herramientas bioquímicas	282
9.1.	Fundamentos de enzimología	282
9.1.1.	¿Qué es una reacción química?	282
9.1.2.	¿Qué es un catalizador?	284
9.1.3.	¿Qué es una enzima?	285
9.1.4.	Características generales de una reacción enzimática	285
9.1.5.	¿Cómo funcionan las enzimas?	286
9.1.6.	Nomenclatura de las enzimas	288
9.1.7.	Regulación de la actividad enzimática	289
9.1.8.	Modelo cinético de Michaelis-Menten	295
9.2.	Fundamentos de espectrofotometría	303
9.2.1.	La radiación electromagnética	303
9.2.2.	Cromóforos	305
9.2.3.	Los auxocromos	306
9.2.4.	Espectro de compuestos biológicos	307
9.2.5.	La ley de Beer-Lambert	310
9.2.6.	Métodos enzimáticos.	312
9.3.	Análisis de macromoléculas	320
9.3.1.	Electroforesis	320
9.3.2.	Cromatografía	322
9.4.	Metodologías y equipos	324
9.4.1.	Cromatografía de líquidos (HPLC)	324

VI

9.4.2.	Cromatografía de gases (GC)	333
9.4.3.	Ejemplos de metodologías usadas para el mejoramiento de maíz	337
10.	Fisiología vegetal	349
10.1.	Introducción	349
10.2.	Asimilación de carbono	354
10.2.1.	La fotosíntesis	354
10.2.2.	Plantas C3, C4 y CAM	374
10.2.3.	Descripción de las distintas rutas fotosintéticas	375
10.3.	Tejidos vegetales	385
10.3.1.	Monocotiledóneas y dicotiledóneas	385
10.3.2.	Semillas	387
10.3.3.	Hojas	388
10.3.4.	Raíces	390
10.3.5.	Tallos	394
10.3.6.	Tejido Vascular	397
10.4.	Sistemas de transporte vascular	402
10.4.1.	Transporte por el xilema	402
10.4.2.	Transporte por el floema	410
11.	Metodologías de mejoramiento	417
11.1.	Metodologías clásicas	417
11.1.1.	Introducción	417
11.1.2.	Importancia del ambiente	417
11.1.3.	Importancia de los modos de reproducción	418
11.1.4.	Autopolinización o polinización libre	419
11.1.5.	Selección Masal	420
11.1.6.	Esquemas de selección	428
11.1.7.	Mejoramiento por pedigrí	432
11.1.8.	Selección recurrente	434
11.1.9.	Metas de mejoramiento	434
11.2.	Metodologías innovadoras	442
11.2.1.	Retrospectiva histórica del mejoramiento molecular de plantas	442
11.2.2.	Modernización de viejas herramientas	456
11.2.3.	Selección asistida por marcadores	461
11.2.4.	Nuevos esquemas de mejoramiento	463
11.3.	Retos para el futuro	467
12.	Mejoramiento molecular	468
12.1.	Desarrollo histórico del mejoramiento molecular de plantas	468
12.2.	Principios y prácticas del mejoramiento molecular de plantas	469
12.3.	Figuras de marcadores y QTLs	483
12.3.1.	Referencias	495

1. Notas introductorias

1.1. *Presentación y enfoque*

La mayoría de los libros sobre mejoramiento genético tienen un enfoque tradicional que incluye sólo de manera tangencial los desarrollos científicos modernos. A pesar de que existe un gran número de referencias bibliográficas en inglés, los libros de texto actualizados sobre mejoramiento genético en español son pocos y muy limitados. Decidimos entonces escribir un texto didáctico especialmente para los estudiantes hispano-americanos.

El punto de partida de este libro es la recapitulación de los **fundamentos y conceptos** que son necesarios actualmente para el **mejoramiento genético moderno**. Se elaboró un texto con la intención de incluir algunos **conocimientos de frontera** y mostrar su posible aplicación en la agricultura. Con este fin, se eligió al **maíz** como planta modelo por su importancia cultural, social y económica. También se hizo énfasis en las características distintivas de este cultivo, en particular las relacionadas con el fenómeno de **heterosis**. El vigor híbrido del maíz es fascinante desde el punto de vista teórico como práctico. La heterosis del maíz es un enigma científico sin resolver que además da sustento económico a muchas empresas que comercializan semilla certificada de híbridos que poseen altos rendimientos.

¿En qué se diferencian el mejoramiento clásico y el moderno? Básicamente en las herramientas, pues los principios y objetivos son los mismos. Algunas metodologías permiten realizar mejoramiento genético a diferentes escalas genotipo-fenotipo, convirtiéndose de esta manera en herramientas complementarias. El fitomejoramiento, tradicional o moderno, no son estrategias excluyentes. De manera simplificada, se puede decir que el mejoramiento moderno no es otra cosa que el refinamiento del mejoramiento clásico. Esto se logra incorporando algunas herramientas moleculares adicionales. ¿Cuáles son estas herramientas? ¿Qué se necesita saber de ellas para sacar el mayor provecho? Este libro da una primera respuesta a estas inquietudes.

El mejoramiento vegetal se inventó y se desarrolló en una época en donde no se sabía nada del DNA ni se había desarrollado la técnica de PCR. Se estandarizó en una era en la cual no se podían predecir las implicaciones de la **ingeniería genética** ni tampoco los alcances de la **genómica** y la **bioinformática**. El punto de partida de muchos libros de texto sobre fitomejoramiento son las clásicas leyes de Mendel de la herencia y las

teorías estadísticas de la genética de poblaciones. Podría considerarse que el libro más avanzado de mejoramiento clásico es el Tratado de Hallauer sobre genética cuantitativa. Sus teorías y propuestas son una excelente referencia para todos los mejoradores. Los que dominan los conceptos de Hallauer son expertos en el área. El presente libro está orientado a explicar algunos fundamentos de las metodologías modernas que amplían los criterios de selección genética para acelerar el mejoramiento.

Para tener éxito en el mejoramiento genético no se necesita entender la estructura del DNA ni las implicaciones de la genómica o la metabolómica. Sin embargo, en un mundo competitivo en donde la **rapidez**, la **eficiencia** y la **economía** son aspectos que deben tomarse en cuenta, es necesario incorporar todas las herramientas que se han desarrollado en los últimos años para facilitar el trabajo y obtener mejores resultados. Es por ello que en este libro se decidió iniciar con los ácidos nucleicos y los fundamentos moleculares de la herencia. Partiendo de estas bases pretendemos acercarnos a los diferentes temas de genética, bioquímica y fisiología vegetal, temas relevantes para incrementar el rendimiento o la calidad de un cultivo. Por último abordamos las metodologías clásicas de mejoramiento, y las particularidades de los métodos de cruzamientos, evaluación y selección.

2. Conceptos básicos

2.1. Fundamentos de la genética

2.1.1. Genética vegetal

La genética es una disciplina de las ciencias biológicas, por medio de la cual se generan nuevos conocimientos sobre la herencia y descendencia de los seres vivos. También generará información sobre el funcionamiento molecular de las células vivas. El **ácido desoxirribonucleico (DNA)** por sus siglas en inglés de *DeoxyriboNucleic Acid*) es muy importante en este contexto. La **genética** nos ayuda a revelar la importancia del DNA como la molécula portadora de la información biológica. Más que ser una ciencia meramente básica, la genética también tiene un componente muy aplicado de **ingeniería**, es decir, del uso de tecnología (conocimientos, métodos y materiales) que permite diseñar y modificar productos o procesos. Esto ha permitido desarrollar un conjunto de metodologías para manipular el DNA de una especie. Uno de los objetivos de la ingeniería genética es proporcionar herramientas moleculares para modificar los caracteres productivos de las especies biológicas. Con ello se pretende lograr que algunos genes sean más útiles y aprovechables para el ser humano (**biotecnología**).

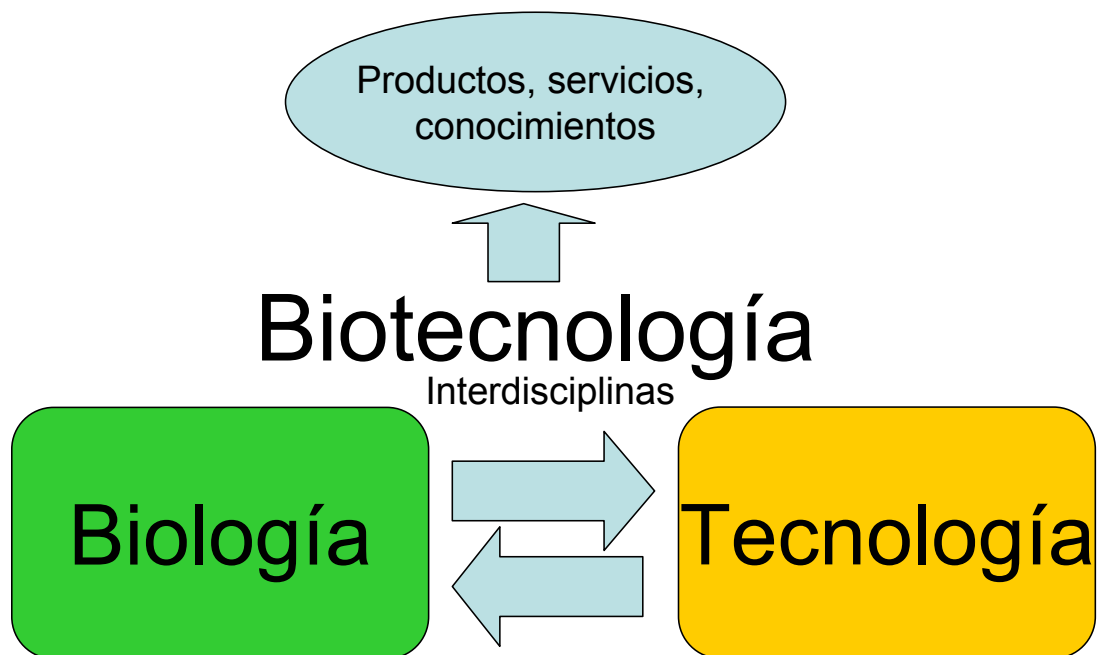


Figura 2-1. La biotecnología como una ciencia interdisciplinaria con dos vertientes: biología para la tecnología y viceversa.

En especial, la **ingeniería genética vegetal** está enfocada a mejorar la productividad de las plantas y los microorganismos benéficos. Pero, ¿cómo lograrlo de manera concreta? ¿Cómo se puede incrementar el rendimiento de un cultivo como el maíz? La tarea no es fácil. Requiere de técnicas diversas, dedicación, tiempo y esfuerzo. Para llegar a ser expertos en el tema hay que profundizar nuestros conocimientos sobre las bases del mejoramiento genético. Primero hay que entender todos los fundamentos científicos sobre el tema. ¿Qué factores influyen en el rendimiento agrícola? Empezemos a analizarlos uno por uno.

2.1.2. Factores que determinan la productividad agrícola

La producción agrícola en términos de **cantidad** y **calidad** dependen de dos grandes conjuntos de factores:

Factores externos abióticos + bióticos	Factores internos genéticos + epigenéticos
El ambiente incluye factores físicos y químicos, denominados factores abióticos, como son: la disponibilidad de agua, nutrientes, luz, temperatura, etc. El ambiente externo también incluye factores biológicos de interacciones ecológicas entre diferentes especies, como por ejemplo, microorganismos simbióticos, insectos polinizadores, virus y diversos patógenos. Estos factores se denominan factores bióticos.	La genética incluye una serie de factores internos sobre la herencia del DNA, como; genes, alelos, heterosis, poliploidia, epistasia, etc. Los factores genéticos incluyen la información codificada en la secuencia nucleotídica del DNA. También existe una serie de factores que no dependen de la secuencia primaria del DNA (factores epigenéticos), pero que si se heredan y también influyen en la forma que se expresa y se manifiesta el fenotipo.

La producción agrícola se puede incrementar al optimizar todas las variables de estos dos grandes conjuntos de factores (variables externas e internas). Es muy importante considerar los dos conjuntos de factores de forma integral, ya que el rendimiento depende del factor que es más limitante en ese momento.

Por ejemplo, si la limitante es la cantidad de agua que se le aplica a un cultivo, entonces no tiene caso incrementar la fertilización de nitrógeno más allá de un cierto límite. Lo mismo ocurre con los factores genéticos y ambientales ¿De qué sirve hacer mejoramiento genético de una forma sofisticada si las prácticas agrícolas no son adecuadas para las variedades que se están desarrollando? Por otro lado, no tiene caso implementar la más alta tecnología agrícola si lo que se pretende es utilizar variedades criollas que no han sido mejoradas para esas condiciones óptimas, por ejemplo, la siembra a una densidad mayor a 80 mil plantas por hectárea. Por lo regular, las variedades criollas son de plantas muy grandes, que comúnmente se siembran a densidades menores a 50 mil plantas por hectárea. Es decir, el mejoramiento genético de plantas tiene que ir de la mano con el manejo agronómico.

El mejorador tiene que hacer un trabajo multidisciplinario. Tiene que saber tanto de genética como de agronomía, patología, fisiología, biología molecular y bioquímica. No sólo debe dominar las ciencias naturales, sino que a veces también es muy importante saber algo de economía, sociología, psicología y de política para tener un mayor impacto. En la práctica, muy pocos científicos dominan todas las disciplinas necesarias, por lo que el mejoramiento genético es una labor que debe realizarse en equipo. Se requiere un grupo de expertos con un enfoque interdisciplinario. No es la simple adición de personas de múltiples disciplinas, sino la sinergia de diversas metodologías la que conduce al éxito.

¿Cómo podemos optimizar las variables externas?

El ambiente se puede acondicionar a través del manejo agrícola. Esto se logra, por ejemplo, por medio de la fertilización, el riego, el control integral de plagas, el uso de invernaderos, etc.

La **agricultura protegida**, es decir, el cultivo de plantas dentro de invernaderos, es una de las técnicas más eficientes para incrementar los rendimientos agrícolas. En los invernaderos se puede optimizar la humedad relativa, se riega y fertiliza adecuadamente, se regula la temperatura, se incrementa el CO₂ y se protegen los cultivos de lluvias, heladas, plagas y patógenos. Si comparamos, a campo abierto se cosechan 200 toneladas de tomate por hectárea por año, mientras en un invernadero se llegan a producir hasta 600 toneladas por año. En algunos invernaderos de Holanda ya se está rompiendo el record de mil toneladas por hectárea por año.

Ahora que sabemos como se optimizan las variables externas por medio del manejo agronómico, nos queda por preguntar: ¿Cómo se optimizan las variables internas de la genética? El potencial genético de una especie se puede incrementar al combinar genes y acumular alelos favorables. Esto se logra a través del **mejoramiento genético por medio de selección y recombinación**. Este es un proceso que debe de ser constante y continuo, con ciclos iterativos.

Muchas veces la influencia del ambiente es mucho más fuerte que el factor genético. De hecho, la mayor limitante para la productividad de los cultivos

vegetales es la **disponibilidad de agua y nitrógeno**. En promedio, los agricultores solo obtienen del 5 al 20% del máximo potencial genético de un cultivo.

Por ejemplo, el maíz puede producir hasta 16 toneladas de grano por hectárea, mientras que el rendimiento promedio en México es de 2.4 ton/ha. Esto se debe a que la mayoría de los agricultores siembran en tierras y localidades donde no hay suficiente nitrógeno ni agua. Algunos agricultores no obtienen los máximos rendimientos por falta de recursos o conocimientos para implementar buenas prácticas agrícolas. Si bien es cierto que se puede incrementar el rendimiento a través de su potencial genético, en el caso de países con baja tecnificación agrícola como México, los factores ambientales son por mucho los factores limitantes para la productividad de maíz a nivel nacional.

Para incrementar la producción de grano en México y Latinoamérica en general, se necesita optimizar el manejo agrícola para obtener mayores rendimientos. Sin embargo, los fertilizantes son costosos, el agua es escasa y las plagas atacan los cultivos. Es por eso que la genética si es importante, sobre todo, si dos agricultores con el mismo suelo, insumos y manejo agrícola comparan sus rendimientos.

Optimización de factores	
externos agrónomos y agricultores	internos mejoradores y científicos
Ejemplos: • Agricultura de conservación • Labranza mínima • Riego y fertilización • Hidroponía y ferti-irrigación • Biofertilizantes e inóculos • Control integral de plagas • Cultivos orgánicos • Lombricultura • Invernaderos	Ejemplos: • Mejoramiento poblacional • Selección recurrente • Heterosis • Mejoramiento por pedigrí • Marcadores moleculares • Fenotipo de precisión • Bioestadística y bioinformática • Injertos y clones • Organismos genéticamente modificados

El uso de semilla no-mejorada o si-mejorada puede marcar la diferencia entre una **agricultura rentable** o no rentable. Puede significar el sustento o la ruina para un campesino y su familia. La genética de los cultivos puede

derivar en hambre o calidad de vida para pueblos enteros. El mejoramiento genético del maíz es un trabajo muy importante. Muchas personas dependen de este cultivo para su alimentación y sustento. El mejoramiento de especies vegetales es particularmente atractivo en **países megadiversos** como los nuestros en Latinoamérica. El territorio de México cuenta con una amplia diversidad de climas, ecosistemas y suelos, lo que ha dado origen a una gran riqueza de especies, tanto animales como vegetales. Es una gran ventaja contar con una enorme riqueza genética y debemos de cuidarla como un patrimonio de toda la humanidad. También es cierto que debemos de sacarle provecho a esa ventaja competitiva que nos ha dado la naturaleza. ¿De qué nos sirve tener mucha diversidad biológica, si las condiciones de vida de una mayoría son de pobreza y marginación? La **biotecnología vegetal** nos puede ayudar a encontrar los mecanismos para que la diversidad biológica y la riqueza natural sean una fuente de bienestar para toda nuestra sociedad.



Figura 2-2. Biotecnología y diversidad biológica, requieren de un trabajo con múltiples niveles de observación y complejidad.

2.1.3. Del genotipo al fenotipo

2.1.3.1. Visión histórica

Hoy en día, uno de los temas más importantes de las ciencias biológicas es determinar la **relación que existe entre los genes y sus funciones específicas**. Es decir, entender cómo se llega **del genotipo al fenotipo** y viceversa. La forma de comprender este vínculo ha cambiado considerablemente en los últimos 3 siglos.

En el siglo 19, en la época de Gregor Mendel y Charles Darwin, el vínculo entre el genotipo y el fenotipo se postulaba sin saber nada sobre el DNA. Parecía un modelo muy simple. Lo que se veía era el fenotipo, pero ya se intuía que había algo más llamado **gen**. Los genes eran unas instancias ocultas que se heredaban a los descendientes. Se sabía que existían dos genes para cada característica, y que uno veía de la madre, y el otro del padre. Cada gen podía tener diferentes variantes, a los que llamaron **alelos**. También se sabía que unos alelos eran **dominantes** sobre otros que eran **recesivos**. Las leyes de herencia de Mendel explicaban como segregaban las características fenotípicas en las generaciones filiales F1 y F2.

Tarea 2-1

Repase las leyes de herencia de Mendel. Haga un cuadro donde explique la segregación Mendeliana en la generación F2.

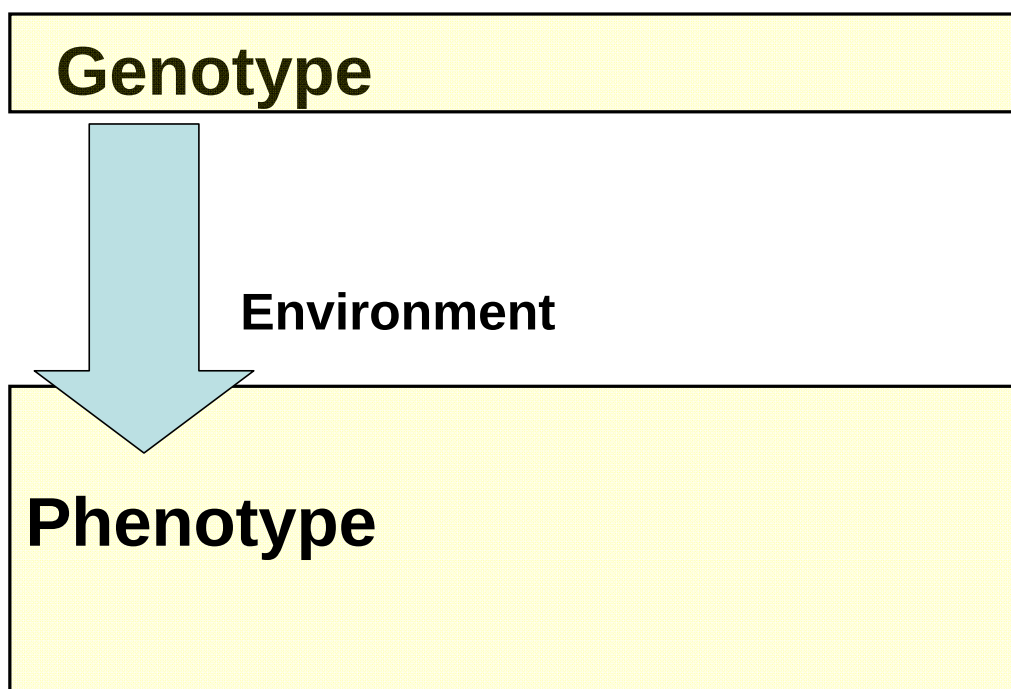


Figura 2-3. Vínculo entre el fenotipo y el genotipo como lo entendía Gregorio Mendel y otros científicos en el siglo 19. Un fenotipo es

causado por un gen. La forma en que se heredaban esas características esta descrita por las leyes de la genética mendeliana. Esto aplicaba muy bien para el color de las flores o la forma de las semillas de chícharo. Se pensaba que el ambiente solo tenía una influencia en los casos de fenotipos mas complejos.

Mas tarde, a mediados del siglo 20 se estudiaron las proteínas y el DNA. Se descubrió que el DNA era el portador de la información genética. También se encontró que el DNA se transcribía a RNA, y que este se traducía a una secuencia específica de amino ácidos según un **código genético universal**. Nació entonces el **dogma central de la biología molecular**, que postula que la información fluye en una sola dirección, del genotipo al fenotipo. Esto implicaba que no se pueden heredar las características adquiridas durante la vida. Otra forma de ver esto, es que el DNA es una molécula dictadora que determina todo lo que esta subordinado a nivel molecular.

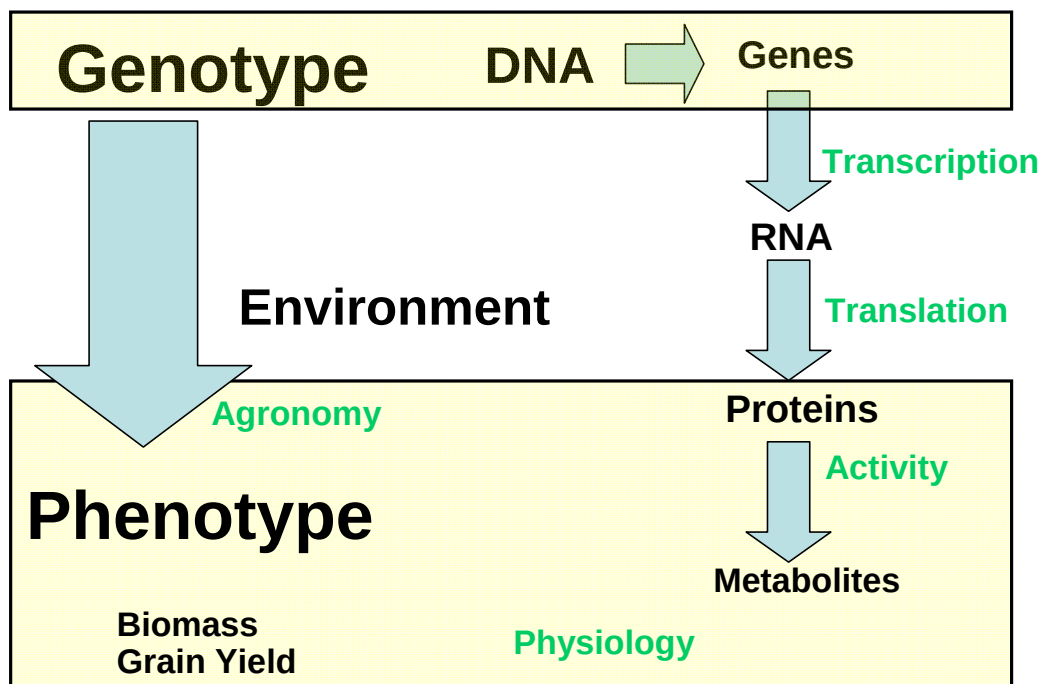


Figura 2-4. Vínculo entre el fenotipo y el genotipo como lo entendían los científicos en el siglo 20. El DNA se replica y se hereda a las siguientes generaciones. La información del DNA se transcribe primero al mRNA y después se traduce al idioma de las proteínas con 20 amino ácidos. Las proteínas pueden formar enzimas que catalizan reacciones bioquímicas para transformar metabolitos. Las enzimas funcionan dentro de rutas metabólicas que están compartimentalizadas en la célula vegetal. Por medio de la fotosíntesis se fija el dióxido de carbono y se forman carbohidratos como sacarosa y almidón. Lo que finalmente cosechamos en forma de biomasa vegetal es el resultado de la acumulación de metabolitos en ciertos órganos de

la planta. El rendimiento depende entonces no solo de los genes, sino de la fisiología y la agronomía, y estas a su vez están influenciadas por el medio ambiente.

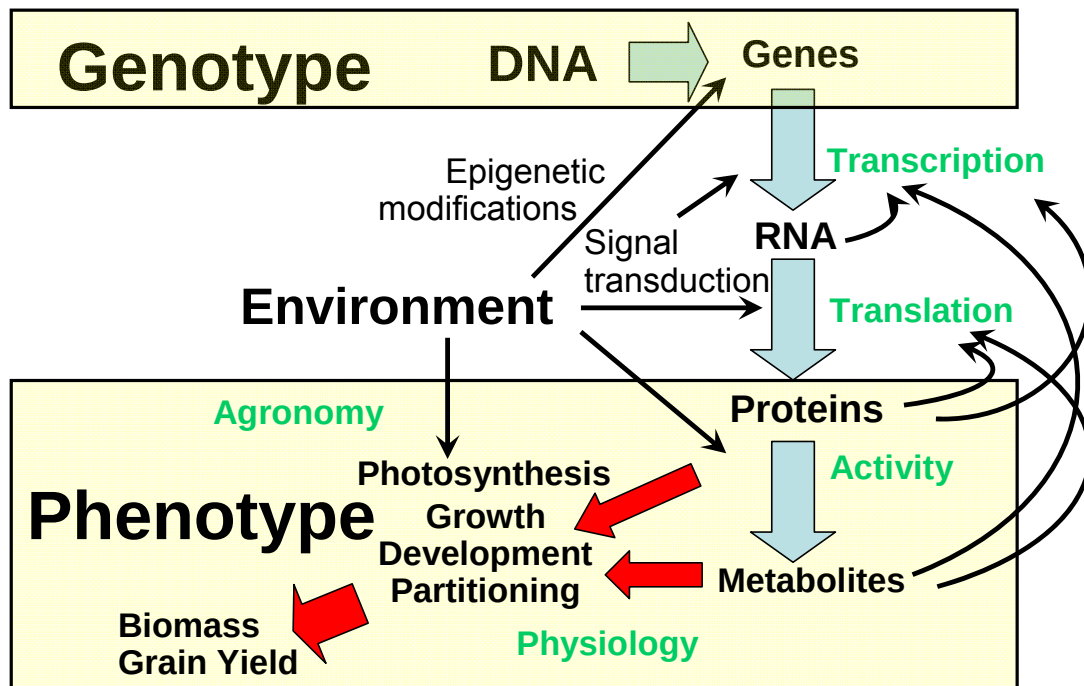


Figura 2-5. Vínculo entre el fenotipo y el genotipo como lo están empezando a entender los científicos del siglo 21. El modelo simplista se ha venido haciendo cada vez más complejo. Ya no solo se conoce un flujo de información en una sola dirección, sino que también hay mucha retroalimentación. Por ejemplo, la epigenética demuestra que el DNA puede ser modificado por medio el ambiente. Los microRNA pueden afectar la expresión de los genes de manera retrograda (*gene silencing*). El flujo de información es bidireccional, ya que algunos RNAs se pueden convertir en DNA por medio de la enzima transcriptasa reversa (RT). El ambiente puede afectar los procesos moleculares en todos los niveles, desde la transcripción, hasta la actividad enzimática. Todo se ha mas vuelto mas complicado. Con este último modelo se puede entender porque el rendimiento de grano del maíz no es un simple reflejo de los genes, y por consiguiente, el mejoramiento molecular no puede ser tan simple como solo secuenciar el DNA a nivel genómico.

De la **genómica estructural** se está pasando cada vez más a la **genómica funcional**. La secuenciación de genomas completos es sólo el principio de una tarea más amplia dentro de la **biología de sistemas**.

2.1.3.2. Modelos genéticos

En el siglo pasado se debatió mucho acerca de la influencia del ambiente y los genes. Esta discusión no se limitaba a características de las especies vegetales, sino también al ser humano. Algunos decían que nuestra vida estaba determinada por el ambiente, otros opinaban que los genes lo predeterminaban todo. Hoy en día, los expertos todavía no se han puesto de acuerdo acerca de algunas de nuestras habilidades. ¿Cómo están determinadas?, ¿es el talento innato adquirido por medio de la herencia o es lo que aprendemos por medio de la experiencia y la educación?, ¿somos el resultado de nuestros genes o de nuestras vivencias? Lo mismo ocurre con las plantas, ¿qué es lo que determina la apariencia y el fenotipo de una planta?

Para discutir los posibles modelos genéticos es muy importante contar con definiciones muy claras de los conceptos utilizados.

Fenotipo (F): El fenotipo es el conjunto de caracteres que se manifiestan visiblemente a nivel del individuo o población. Estos pueden ser parámetros agronómicos cualitativos o cuantitativos (color, rendimiento, etc.). El fenotipo es lo que podemos observar, medir, cuantificar o cosechar. El fenotipo puede ser algo muy simple como el color del grano, pero también puede ser una característica tan compleja como la inteligencia o el comportamiento sexual en los seres humanos.

Ambiente (E del inglés *Environment*): El ambiente es el conjunto de variables externas que influyen sobre el desarrollo y las funciones de un organismo. Muchas veces el ambiente controla la expresión de los genes o las proteínas. Por ejemplo: la temperatura afecta la función de la célula, promueve la actividad de algunas enzimas o cambia el patrón de acumulación de metabolitos. El ambiente incluye todos los factores **abióticos** (parámetros químicos y físicos) y también factores **bióticos** de interacciones biológicas y ecológicas. Algunas variables ambientales se pueden medir, pero muchas veces no se pueden controlar en el campo.

Genotipo (G): El genotipo es la suma de los genes y combinación de alelos que tiene un determinado individuo o variedad. En términos moleculares es la información genética codificada en el DNA que está presente en el núcleo, los plástidos y mitocondrias. A nivel de especie se habla del **genoma**, que incluye todos los genes y su potencial de expresión. A nivel de población se habla del **pool genético**,

mientras que a nivel individual se habla de la **combinación de alelos** que están presente en determinado genotipo.

A lo largo de la historia se han postulado varios modelos que tratan de explicar los factores que influyen el fenotipo de un individuo.

Fenotipo = E	Modelo ambientalista reduccionista
Fenotipo = G	Modelo genetista reduccionista
Fenotipo = E + G	Modelo aditivo simple

El modelo ambientalista asume que todas las características dependen principalmente del ambiente. Por ejemplo, podríamos asumir que todas las personas tienen exactamente las mismas habilidades y aptitudes (mismo fenotipo), y que es solo una cuestión cultural el hecho que los niños jueguen con carritos y pelotas, mientras que las niñas prefieran las muñecas. Digamos que en ese caso, el fenotipo (el comportamiento de los niños) es una imposición del ambiente, de su entorno, moldeándolos a un rol social preestablecido. En el caso de las plantas pudiéramos pensar en algo similar, en donde el fenotipo (aspecto de la planta) dependa solo del ambiente, y que la genética no juegue ningún papel importante. El modelo genetista predice lo contrario: el aspecto de la planta (forma de la hoja) depende solo del genotipo, independientemente del ambiente en que se desarrolle. Esos son casos extremos de modelos reduccionistas. En la mayoría de los casos la verdad se encuentra justo en medio.

El modelo aditivo simple asume que el factor genético y el factor ambiental son totalmente independientes y actúan de forma lineal. Muchas veces este es el caso mas frecuente, sin embargo, no siempre. Algunas veces, el factor genético puede tener una interacción no-lineal con el ambiente. Un mismo alelo puede tener un efecto fenotípico diferente, dependiendo del ambiente en que se exprese. A esta interacción que no necesariamente es aditiva ni lineal se le llama interacción genotipo-ambiente $G \times E$. (*Genotype x Environment*)

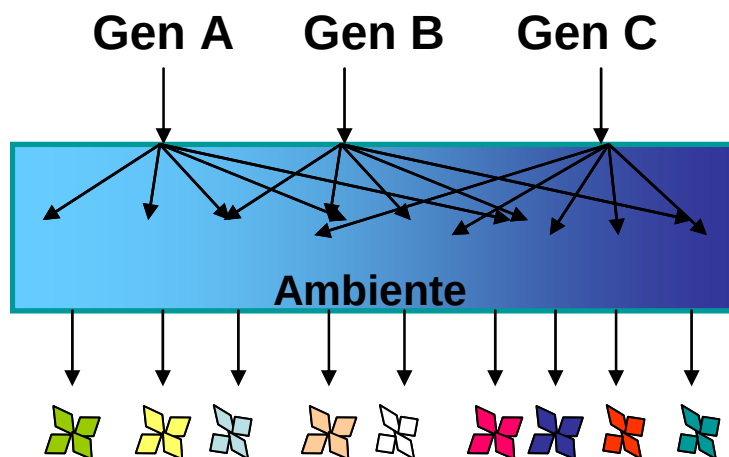


Figura 2-6. Modelo de determinación fenotípica que muestra cómo los genes y el ambiente interactúan en el desarrollo para producir un fenotipo. El ambiente es como un prisma que puede cambiar la forma en que un gen se expresa a nivel de fenotipo. El mismo gen puede tener un efecto diferente dependiendo del ambiente. La combinación de genes y la interacción con el ambiente pueden dar lugar a fenotipos complejos. Esta complejidad significa que el fenotipo es más que la simple suma de los efectos individuales de los genes.

De cierta forma, la interacción genética no-lineal contradice la noción que generalmente nos enseñan sobre los genes, derivada de los experimentos con chícharos de Gregor Mendel y las mosquitas de la fruta de Thomas Morgan. Por lo regular, un alelo que determina el color rojo o blanco de los ojos de una mosquita, va a funcionar de la misma forma, independientemente del medio en que se desarrolle el individuo. Sin embargo, este parece ser sólo el caso para los caracteres bioquímicos simples. Esta relación simple es mucho más común para las bacterias que para los organismos superiores. Para los caracteres fenotípicos con mayor complejidad, como el rendimiento de grano en maíz, existen muchos ejemplos de que los genes no siempre actúan de la misma forma y que existe un tipo de herencia que es más complicada de lo que generalmente se menciona en los libros de genética clásica.

Los investigadores que tienen experiencia en mejoramiento genético reconocen que es un fenómeno de expresión complejo, por eso hacen sus evaluaciones de rendimiento en diferentes localidades y en diferentes años. De esta forma, pueden hacer un análisis estadístico que permite calcular los efectos de G, E y GxE para tratar de identificar aquellos alelos que tienen un efecto genético positivo y una baja interacción con el ambiente. A esto se refieren los mejoradores cuando hablan de “potencial” de rendimiento y “estabilidad” de rendimiento.

Los libros de biología molecular muchas veces explican el concepto de la información genética de manera muy simple. Por ejemplo: diciendo que a partir de un gen se forma una sola proteína, que siempre es la misma, que tiene una determinada y única función. Es por eso, que la biología molecular postula que al determinar la secuencia del DNA, se sabrá que proteína se va a formar y por ende se puede predecir la función y el fenotipo. Recientemente se ha encontrado que a través del mecanismo de *splicing* diferencial, un solo gen puede dar lugar a varias proteínas diferentes. Si ya sabemos que una secuencia de DNA puede dar origen a varias proteínas, no estamos lejos de comprender que un gen puede tener diferentes funciones. Una sola proteína puede tener diferentes efectos dependiendo del ambiente o de las demás proteínas con las que se encuentre asociada.

En términos prácticos, la interacción GxE puede significar que un alelo puede ser favorable para condiciones de riego normal, pero perjudicial bajo condiciones de sequía.

Por ejemplo: un gen que incrementa la densidad de aperturas estomáticas (número de estomas) en las hojas verdes puede ayudar a que la planta tenga

una mayor asimilación de dióxido de carbono. Esto será benéfico para la tasa fotosintética y el rendimiento de grano en condiciones de alta humedad; sin embargo, esta misma característica puede ser perjudicial si la planta se expone a una baja humedad relativa y condiciones limitantes de agua. El efecto de un gen depende del ambiente, de forma que un mismo alelo puede ser benéfico o desfavorable según las condiciones particulares en que se exprese.

Para explicar esta diferencia de una respuesta genética simple se debe de incluir el factor GxE en la ecuación. En resumen se puede establecer un modelo más completo como:

$$\text{Fenotipo} = E + G + G \times E$$

Este es el modelo que en la actualidad se aplica con mayor frecuencia para el análisis de varianzas y detección de **QTLs** (del inglés *Quantitative Trait Locus*). No obstante, hay que tomar en cuenta que este modelo todavía es un poco reduccionista. Un factor que siempre está presente en todos los experimentos es el azar, y por lo tanto, se debe incluir un término de **variabilidad residual** que llamaremos **R** (del inglés *randomness*). Sin embargo, el factor R es un tanto intangible y para muchos muy molesto. Por lo regular se asume que es constante, pequeño y tiene una media de cero y una distribución normal, por lo que en forma común se ignora (a veces injustificadamente) y se elimina de las ecuaciones.

Otro factor que también influye en el fenotipo, es la interacción no lineal entre genes (interacción GxG). Esta interacción se expresa en forma de epistasis, heterosis y factores epigenéticos.

Por ejemplo: un gen puede ser benéfico en cierto fondo genético, mientras que en otro puede ser detrimental. Un alelo puede funcionar si está presente en estado homocigótico, pero tener un efecto diferente en estado heterocigótico y viceversa.

En el caso del maíz, el efecto de heterosis es tan marcado (>200%), que casi todos los parámetros agronómicos importantes se tienen que evaluar en híbridos y no en líneas homocigóticas. Este fenómeno se refiere a la interacción de gen con gen y se le puede incluir en la ecuación como interacción GxG.

En el presente libro, enfocado a los mejoradores modernos, usaremos un modelo más completo como:

$$\text{Fenotipo} = E + G + G \times E + G \times G + R$$

En el capítulo 7.3 explicaremos con mayor detalle como distinguir entre las diferentes variables. Daremos ejemplos de que experimentos y cálculos se deben de hacer para estimar la contribución de cada parámetro.

2.1.3.3. Control de variabilidad

La variabilidad fenotípica es uno de los obstáculos más grandes para el mejoramiento genético convencional. En los ensayos de campo siempre hay mucha variabilidad. La influencia predominante del ambiente representa un problema, ya que genotipos iguales en ambientes diferentes pueden tener fenotipos muy diferentes, mientras que genotipos diferentes en un mismo ambiente pueden tener fenotipos muy parecidos. Es decir, el factor del ambiente puede ocultar a los demás factores genéticos (cuando E y R son mucho mas grandes que G).

$$E \quad G \quad G \times E \quad G \times G \quad R$$

Los mejoradores de maíz tienen que identificar diferencias genéticas entre variedades de menos de 1 ton/ha. Sin embargo, la variabilidad de campo muchas veces es mayor a 2 ton/ha, mientras que las diferencias entre una condición (riego normal) y otra condición (sequía) causa variaciones de mas de 8 ton/ha ¿Cómo detectar diferencias sutiles con un ruido de fondo altísimo? Bajo condiciones de enorme variabilidad no-genética es muy difícil identificar los alelos de las variedades superiores.

Para detectar mejor las diferencias genéticas, tenemos que disminuir la influencia de los demás factores que afectan el fenotipo. Interesa que el azar sea lo más pequeño posible ($R \approx 0$) para tener resultados más significativos. También interesa que el ambiente sea lo más uniforme posible ($E \approx \text{constante}$) para que la relación fenotipo/genotipo sea lo menos complejo posible ($F \approx E$). Esta uniformidad del ambiente ayuda a identificar mejor a los genes favorables, y así realizar las mejoras genéticas deseadas, ya que la selección por lo general se hace sobre el fenotipo (**selección convencional**).

Una de las ventajas de los marcadores moleculares es que usan la información del DNA y permiten así una **selección genotípica** directa. A esto se le llama selección asistida por marcadores moleculares (MAS, por sus siglas en inglés de *Marker Assisted Selection*).

Para saber si G tiene importancia y es cuantitativo para un caracter determinado en una población, primero tenemos que minimizar los efectos de E y de GxE. Si el ambiente E tiene la influencia mínima posible, el fenotipo F dependerá básicamente del genotipo G. En esas condiciones podremos

determinar si el fenotipo tiene cierta variabilidad y si tiene una cierta distribución estadística (puede ser una distribución binomial o normal). Las distribuciones binomiales se derivan de caracteres monogénicos con efectos muy marcados, mientras que las distribuciones polinomiales y distribuciones normales se explican por un gran número de genes involucrados (poligénicos), cada uno con efecto menor y aditivo (caracteres cuantitativos). Hay que recordar que el factor R, que por definición es lo que no se puede controlar, siempre va a estar presente y nunca se puede eliminar por completo.

2.1.3.4. Matriz de parentescos

Para determinar si G tiene una influencia importante sobre el fenotipo podemos comparar el carácter de interés en individuos que comparten genes (padres-hijos, hermanos completos, medios hermanos, etc.). Hay que establecer una **matriz de parentescos** (ver Figura 2-7) y relacionarlo con el fenotipo. En función del parentesco es posible saber el porcentaje de alelos que comparten los individuos por probabilidad.

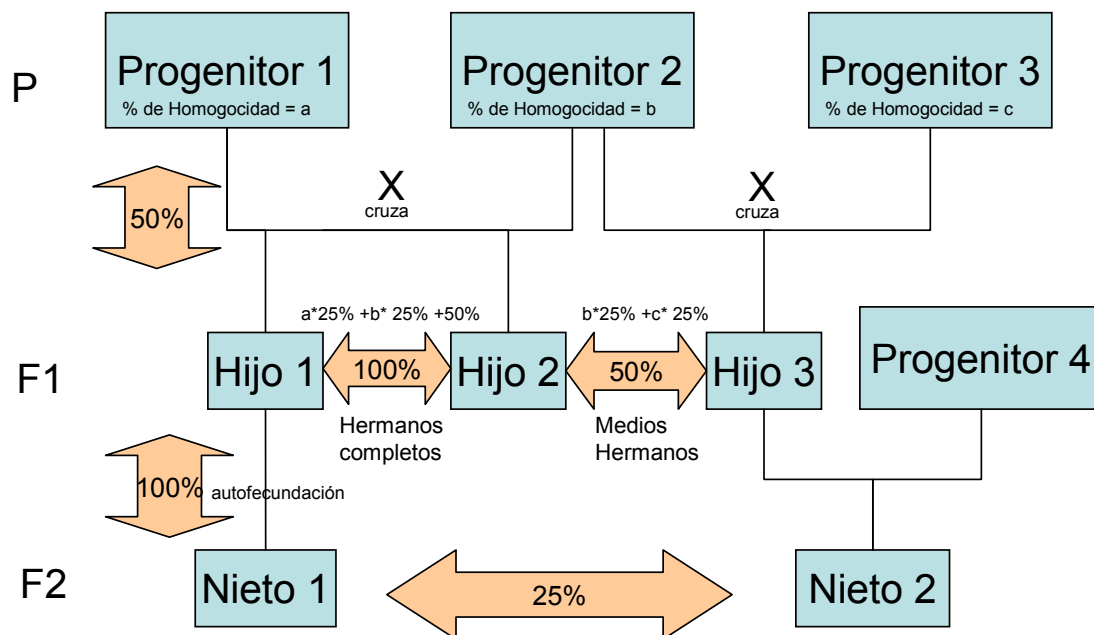


Figura 2-7. Matriz de parentescos.

Progenitor-Descendiente = En caso de una cruza simple, entonces los hijos (F1) compartirán 50% de los alelos con cada uno de sus padres. En caso de una autofecundación con un solo progenitor homocigótico, se compartirán 100% de los alelos.

Si el progenitor es heterocigótico, aunque sea autofecundación, sólo se compartirán 50% de los alelos.

Hermanos completos (misma Si los dos progenitores son heterocigóticos, es decir, cada uno

madre y padre). En caso de que los padres sean homocigóticos, la generación F1 tendrá 100% de alelos compartidos (primera ley genética de Mendel). Esta es la razón por la que los híbridos son genéticamente homogéneos. En la siguiente generación F2 habrá segregación (por ejemplo del tipo 3:1 o 1:2:1).

Medios hermanos (misma madre, pero diferente padre) = 50% alelos compartidos.

tienen dos alelos diferentes para un total de cuatro alelos, entonces los hermanos completos compartirán 50% de sus alelos, pero sólo compartirán el 25% de las combinaciones alelicas diploides.

Si los progenitores son heterocigóticos (la madre tiene dos alelos y los diferentes padres también tienen alelos diferentes), entonces los medios hermanos compartirán 50% de sus alelos.

El parecido que tienen hermanos completos, en comparación con el que tienen los hijos (F1) con sus padres (F0), es un indicador de la heterocigotidad de los padres y de la herencia de un carácter.

Ejemplo-Maíz: Supongamos que sembramos las semillas de una mazorca autofecundada (hermanos completos de un mismo padre y madre). Si vemos que las plantas varían mucho y el carácter que nos interesa tiene una distribución aleatoria (curva normal) entonces, podemos decir que en este caso el factor R es grande, y el factor G influye poco (asumiendo que los factores E y GxE son constantes para todos, ya que las plantas se sembraron en el mismo surco, al mismo tiempo y bajo las mismas condiciones). En cambio, si comparamos las mazorcas de dos plantas diferentes y vemos que los dos grupos de plantas se distinguen claramente, entonces eso nos dice que los genes si son importantes. Estudiando la distribución de los hijos es como podremos detectar cual progenitor es mejor genéticamente. Para saber la importancia de los genes es necesario una **matriz de parentivos** y un estudio de los **fenotipos** (que permitirá ver qué progenitor tiene mejores genes). Si hay coherencia entre el parentesco y el fenotipo, se puede decir que la genética si tiene un peso importante en el carácter.

2.1.4. Genética *forward* y genética reversa

Históricamente, la **genética directa (*forward genetics*)** es la metodología más antigua para hacer estudios genéticos. Por lo regular, primero se observaba un fenotipo, y si era muy interesante se buscaba entonces el gen o los genes que eran responsables de esa característica. Esto se lograba por medio de mapeo hasta encontrar el *locus* genético, para después determinar la secuencia del gen responsable. A esa forma de proceder se le llama genética *forward* (dirección hacia delante, en inglés) (Figura 2-8).

Antes era muy caro y laborioso determinar la secuencia nucleotídica de un gen. Pero hace una década se empezó a usar cotidianamente con una mayor rapidez y exactitud. Cuando los costos de secuenciación bajaron, se volvieron muy accesibles para la mayoría de los investigadores, y por lo tanto se empezó a producir mucha información nucleotídica. Fue entonces que se identificaron muchos más genes que fenotipos. Así inicia la **genética reversa (*reverse genetics*)**. Esta parte de un gen determinado con una función desconocida y trata de ver cual es el efecto de ese gen cuando se interrumpe o se cambia su patrón de expresión. Una forma de alterar el gen es a través de mutagénesis química, con metilsulfonato (EMS), transgénesis insercional (transposones o T-DNA) o también por medio de la sobreexpresión con un promotor diferente.

Algunos científicos usan los conceptos de genética directa y genética reversa. Sin embargo, los términos en inglés (*forward & reverse*) son mucho más comunes en los artículos científicos. Más allá del idioma, lo importante es entender las diferencias entre las metodologías que se utilizan para cada una de esas dos estrategias genéticas.

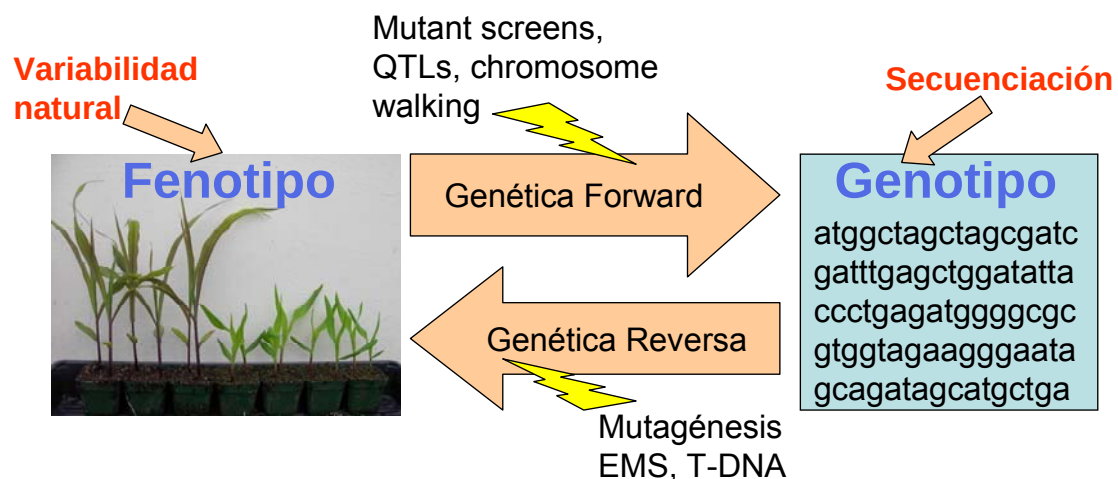


Figura 2-8. Genética *forward* y genética *reverse*.

2.2. Biodiversidad

Ningún individuo es exactamente igual a otro; incluso dentro de una misma familia se encuentra una fuerte variación. Se calcula que actualmente existen unas 450 mil especies vegetales y más de dos millones de especies animales. Entendemos por diversidad biológica o **biodiversidad**, la variedad de formas de vida que habitan la tierra. A este conjunto se la llama la **biosfera**. La diversidad se compone no sólo de un elemento, sino de la variación y la abundancia relativa de especies. Las medidas de diversidad consideran estos dos factores: **riqueza** de especies, que es el número de especies; y **distribución**, esto es, en qué medida son abundantes las poblaciones de cada especie. Se pueden clasificar según los niveles de organización en:

Diversidad genética:

Cada individuo de una especie posee una composición genética fruto de la evolución de millones de años. En el genoma está escrito el potencial de cada individuo, provocando la gran diversidad existente incluso dentro de una misma especie (*pool* genético).

Diversidad epigenética:

Una de los mecanismos por medio del cual esas diferencias ambientales pueden heredarse a las siguientes generaciones es por medio de cambios epigenéticos en el genoma. Esto se refiere principalmente a los patrones de metilación del DNA. Los cambios epigenéticos también se refieren a la acetilación de las histonas y los modificadores de la cromatina.

Diversidad fenotípica:

La misma composición genética puede expresarse de diferente manera según el ambiente y la historia del individuo, generando así plasticidad y diversidad fenotípica. Esta es la razón por la que los gemelos, aunque tengan los mismos genes, a veces sean diferentes y se comporten cada uno en su forma particular.

Diversidad de poblaciones:

Variabilidad entre grupos de individuos que están en una determinada región. Dentro de una misma especie, pueden existir subpoblaciones con determinadas características. Por ejemplo, todos los humanos pertenecen a la misma especie, sin embargo, en África, Asia y Europa existen poblaciones que son muy distintas unas de otras (color de piel, estatura, etc).

Diversidad de especies:

A la diversidad global del planeta contribuyen por una parte las especies ubicuitas (universales) y las especies endémicas. Existen muchas especies

que se encuentran muy extendidas y que en cada zona aparecen como una raza o subespecie, pero siempre dentro de la misma especie. Las especies endémicas son aquellas cuya distribución geográfica se limita a un área muy localizada. La diversidad se debe tanto a las especies universales (en menor medida) como a las especies endémicas (en mayor medida).

Diversidad de ecosistemas:

Viene dada por la multitud de ecosistemas que integran la tierra. En este nivel de diversidad existe cierta imprecisión por la ambigüedad del concepto de ecosistema y la dificultad de delinear un ecosistema de otro vecino.

Un ecosistema es un sistema natural vivo que está formado por un conjunto de organismos vivos (**biocenosis**) y el medio físico en donde se relacionan, **biotopo**. Un ecosistema es una unidad compuesta de organismos interdependientes que comparten el mismo hábitat. Los ecosistemas suelen formar una serie de cadenas tróficas que muestran la interdependencia de los organismos dentro del sistema. El concepto de ecosistema tiene en cuenta las complejas interacciones entre los organismos (por ejemplo plantas, animales, bacterias, algas, protistas y hongos, entre otros) que forman la comunidad (biocenosis) y los flujos de energía y materiales que la atraviesan.

Un concepto parecido al de ecosistema es el de bioma, que es una zona ecológica, climática y geográficamente definida, en donde se dan condiciones climáticas similares y comunidades similares de plantas, animales y organismos del suelo. Los biomas se definen basándose en factores tales como las estructuras de las plantas (árboles, arbustos y hierbas), los tipos de hojas (como maleza de hoja ancha o delgada), la distancia (bosque, floresta, sabana) y el clima (temperatura, humedad, etc). A diferencia de las ecozonas, los biomas no se definen por genética, taxonomía o semejanzas históricas y se identifican con frecuencia con patrones especiales de sucesión ecológica y vegetación clímax. Ejemplos de biomas son la selva tropical, el desierto, la tundra, etc.

Los biomas también pueden considerarse como mega-ambientes para el cultivo de una especie agrícola. Actualmente se distinguen diversos mega-ambientes principales para el mejoramiento de maíz:

- Regiones templadas
 - o Templado de ciclo corto (Iowa, Illinois, Kansas, Alemania, Inglaterra)
 - o Templado de ciclo más largo (Florida, Francia, España)
- Regiones subtropicales
 - o Valles altos >2000 mts de altura (Etiopía, Nepal, Edo. de México, Hidalgo)

- o Subtrópico (800-2000 mts de altura. India, Kenia, Jalisco, Guanajuato)
- o Trópico <800mts de altura (Colombia, Brasil, Veracruz, Yucatán)

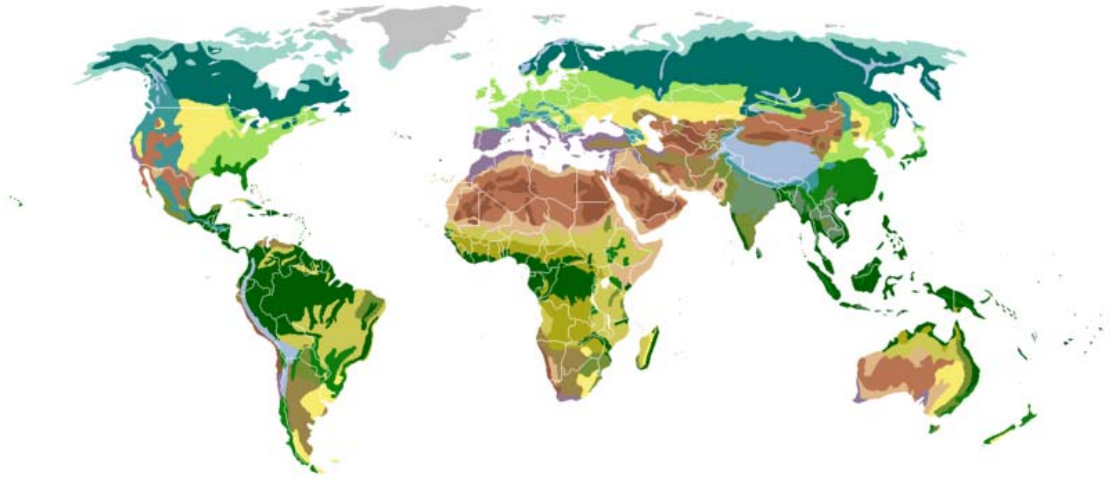


Figura 2-9. Mapa de biomas terrestres. Los diferentes colores del mapa representan diferentes biomas con características distintivas. Por ejemplo, las regiones tropicales están de color verde oscuro.

2.2.1. Diversidad biológica



Figura 2-10. Diversidad de organismos eucariontes (animales, plantas, hongos y protistas).



Figura 2-11. Diversidad de hortalizas y frutas.



Figura 2-12. Diversidad de especies vegetales.

Tarea 2-2

Refuerce sus conocimientos de taxonomía y sistemática. Averigüe el nombre común y científico de algunas de las especies que se muestran en las figuras anteriores.

Tarea 2-3

Refuerce sus conocimientos de botánica. Haga una lista de todas las familias de especies vegetales que conozca.

Tarea 2-4

Haga un esquema de la estructura floral de las siguientes familias de plantas: *poacea*, *brassicacea* y *fabacea*. ¿Cuántos sépalos, pétalos, estambres y carpelos tienen las flores de esas familias?

Tarea 2-5

¿Cuál es la diferencia botánica entre flor e inflorescencia? Identifique y distinga la inflorescencia y las flores individuales del maíz

Tarea 2-6

¿Cuál es la diferencia botánica entre hoja simple y hoja compleja?

Tarea 2-7

¿Cuál es la diferencia botánica entre las espinas que se derivan de una hoja, de un tallo o de una raíz? ¿A qué se refieren los biólogos cuando hablan de órganos homólogos?

Tarea 2-8

¿Qué especies de gramíneas tienen ahuates, espinas o tricomas? ¿Para qué sirven? Los tricomas del maíz ¿cómo se producen?

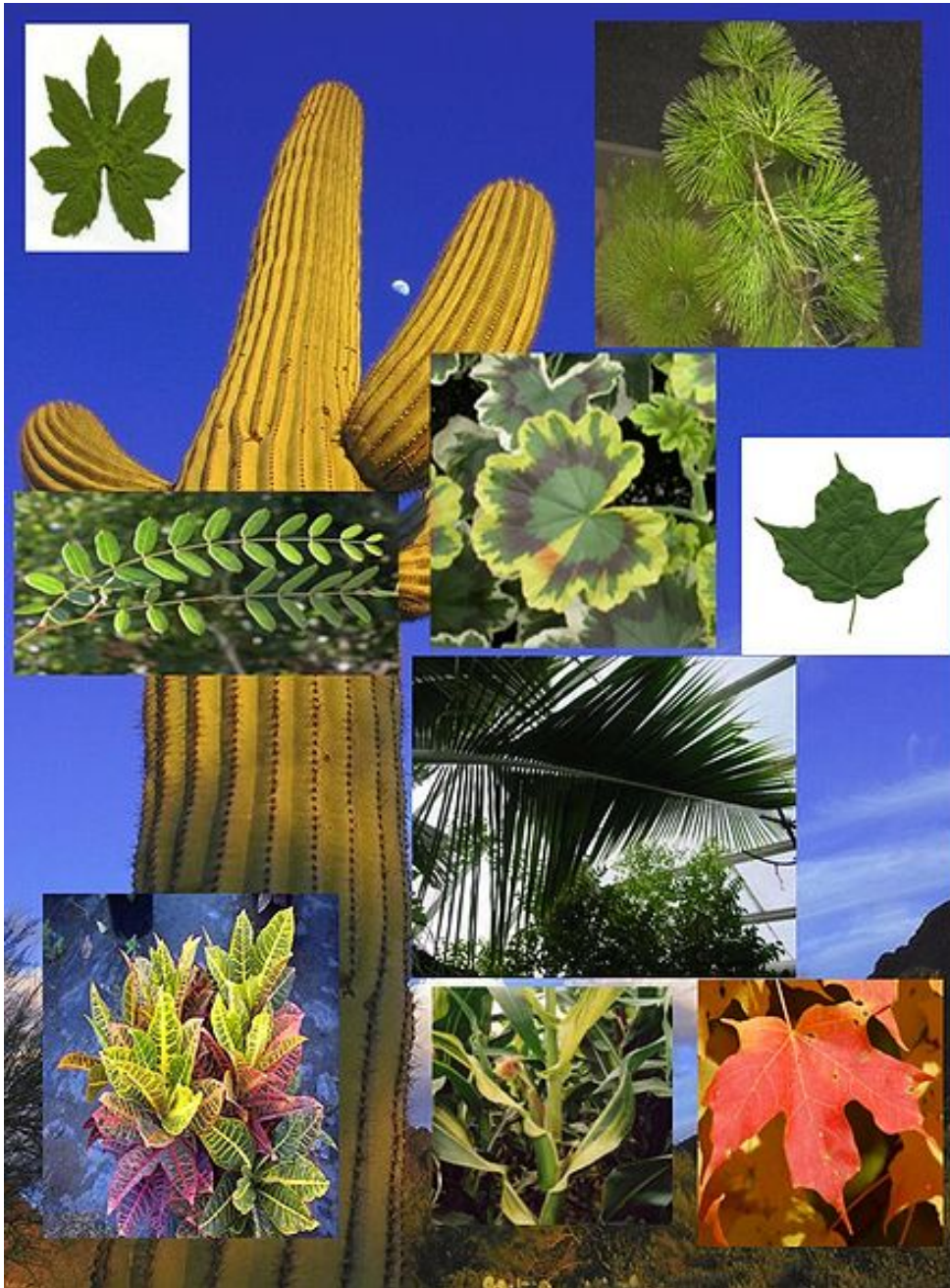


Figura 2-13. Diversidad de formas y funciones de hojas vegetales.

Tarea 2-9

¿A qué se refieren los científicos cuando dicen que México es un país megadiverso? Investigue el origen de algunas especies agrícolas de consumo diario. Averigüe en su biblioteca o por internet cuáles fueron los centros de domesticación de las siguientes plantas: maíz, trigo, arroz, frijol, chile, caña de azúcar, jitomate, arroz, aguacate, cacao, nopal, papa y tabaco.

Tarea 2-10

Muchas plantas de origen mesoamericano se cultivan hoy en día en todo el mundo. Sin embargo, los animales domesticados como el perro, el gato, el cerdo, el caballo, la gallina, la vaca, la cabra, la oveja y muchos animales

son originarios de otros continentes. Sólo el pavo (el guajolote) es de origen mesoamericano y se ha vuelto importante a nivel mundial. Es decir, de México proceden muchas plantas, pero muy pocos animales con importancia económica. Postule una hipótesis de porque esto fue así y haga una recopilación de información que le ayude a sustentarla.

Tarea 2-11

Lea el capítulo sobre biodiversidad escrito por Kevin J. Gaston. Consultelo en el siguiente sitio:

http://s3.amazonaws.com/mongabay/conservation-biology-for-all/Conservation-Biology-for-All_Chapter-2.pdf

Haga un resumen en español de los conceptos principales.

Tarea 2-12

Dibuje el ciclo de vida de una planta con flores. Explique el fenómeno de la doble fecundación.

Tarea 2-13

Defina la palabra dimorfismo sexual ¿Está relacionado con el término polimorfismo genético? ¿Cuál es la importancia del dimorfismo en la evolución?

Tarea 2-14

¿Cuáles son las diferencias entre individuos, familias, poblaciones y comunidades? Si no lo sabe, postule una hipótesis y defiéndala con argumentos.

Tarea 2-15

Para hacer entre profesor y alumnos.

Escoja una población de plantas y animales que conozca cerca de donde usted vive. Enliste las diferencias morfológicas entre individuos y encuentre los dimorfismos si es que los hay.

Tarea 2-16

¿Cuál de los siguientes organismos tiene un genoma plastídico: la diatomea, la víbora, la seta, el maíz, la bacteria *Escherichia Coli*, la garrapata, el tripanosoma? Si no conoce todas las especies, consulte un libro de taxonomía.

Tarea 2-17

Si la especie A tiene más DNA por núcleo que la especie B, ¿tiene A necesariamente más genes que B? Explique ¿Qué factores determinan la cantidad de genes o de DNA de una especie?

Tarea 2-18

Imagine que tiene cuatro variedades de maíz. Los números 1, 2 y 3 son de color azul y el número 4 es blanco. Las cruza siguientes fueron hechas y la progenie obtenida fue evaluada:

1 x 2	→	todo azul
1 x 3	→	$\frac{3}{4}$ azul y $\frac{1}{4}$ blanco
1 x 4	→	$\frac{1}{2}$ azul y $\frac{1}{2}$ blanco

¿Qué fenotipo es dominante, el azul o el blanco? Ahora ponga los genotipos de los cuatro maíces progenitores en una lista. Use la siguiente nomenclatura: A para azul y a para blanco.

Prediga los tipos de progenie y proporciones de la cruza 2x3, 2x4 y 3x4.

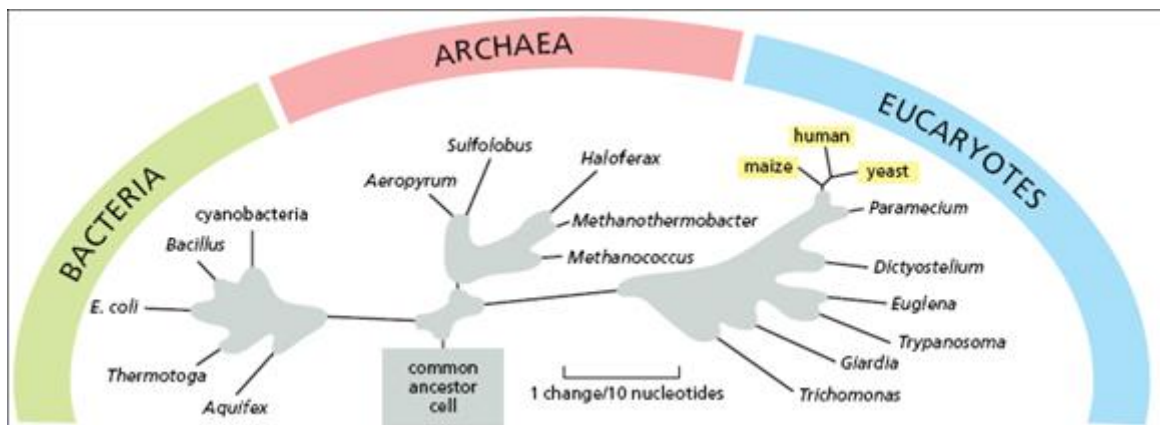
2.2.2. Diversidad taxonómica

Figura 2-14. Árbol de la vida a partir de un ancestro común.



Figura 2-15. Algunos ejemplos de especies de los tres dominios.

2.2.3. Diversidad agrícola

En la agricultura mundial se emplean menos de 300 especies de las 400 mil disponibles. Más del 80% de la alimentación mundial depende de tan solo de 3 especies vegetales (maíz, trigo y arroz). Muchos científicos están muy preocupados por la conservación de la diversidad biológica y genética de las especies silvestres. Algunas de las especies que todavía no han sido domesticadas podrían proporcionar genes muy valiosos. Podría ser un tesoro insustituible para el desarrollo sostenible de la humanidad durante los próximos milenios. Algunos de los genes que se podrían incorporar en variedades futuras son aquellos que son responsables de resistencias a enfermedades y tolerancia a factores abióticos adversos. También están aquellos genes que podrían generar nuevos colores, sabores y olores, hasta ahora inimaginados. Es por ello que la conservación debe de ser una estrategia sistemática de almacenamiento, clasificación, caracterización, unida con bioprospección e introgresión. Pero no se trata sólo de conservar y de almacenar. Un banco de germoplasma no tiene ninguna utilidad si no se usa también para caracterización (fenotipo) y mejoramiento genético.

2.2.4. Diversidad del maíz



Figura 2-16. Colores y formas de mazorcas de variedades criollas.

El maíz es una de las especies biológicas con mayor diversidad. Esta riqueza tan grande puede tener varios orígenes. Por un lado, el genoma de maíz es de los más plásticos y dinámicos, debido a la cantidad de transposones que contiene, además de la alta tasa de mutación y recombinación. No fue una coincidencia de que los elementos genéticos móviles fueran descubiertos por primera vez en maíz.

Tarea 2-19

Averigüe quien fue Bárbara McClintock ¿Qué fue lo que ella descubrió? ¿Qué es un elemento genético movable? ¿Por cuáles meritos científicos le otorgaron el Premio Nobel?

Tarea 2-20

Visite el sitio de la Fundación Nobel en Estocolmo e investigue las contribuciones que han hecho los científicos más famosos en las áreas de química y de medicina.

<http://nobelprize.org/>

Escoja 5 premios Nobel y haga un resumen de los logros científicos con sus propias palabras.

Fue hasta después de los años 70's que en otras especies también se descubrieron los transposones. Otra razón de que sea un cultivo tan diverso, es que el maíz depende 100% del ser humano para la dispersión de su semilla y su desarrollo. Los pueblos indígenas mesoamericanos generaron una gran cantidad de razas. Estas por lo regular se distinguen por el tipo y color de grano y no tanto por el aspecto de la planta. Esto ha hecho que algunos genes responsables de la forma de la mazorca, del tamaño del grano y de algunos caracteres bioquímicos como la síntesis de almidón y de la acumulación de metabolitos (proteínas, sacarosa, lípidos, carotenos, antocianinas, etc) tengan una huella de selección muy fuerte por la domesticación y la formación de razas con características distintas. Por ejemplo, muchas veces se distingue una raza de maíz criolla por ser de grano dulce, o de color azul o amarillo, siendo que estos fenotipos pueden ser determinados por un solo gen.

Tarea 2-21

Haga una revisión de genes de maíz que se hayan descubierto por medio del efecto que se observa cuando se mutan y se inactivan. Inicie con una búsqueda en la página Web de MaizeGDB. Haga una lista de los genes/enzimas que se conozcan para la coloración de las semillas.

2.3. Origen de la variabilidad genética

2.3.1. Mutaciones

Una parte de la diversidad genética se genera por mutaciones puntuales durante el ciclo celular. Por lo regular, el DNA se replica dentro del núcleo generando copias idénticas.

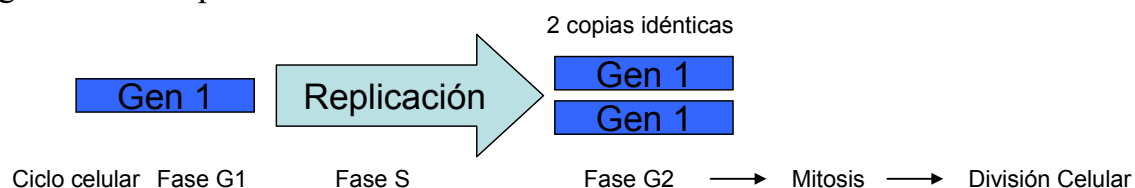


Figura 2-17. Replicación de los genes durante el ciclo celular. La fase S es la etapa en la que se da la síntesis del DNA. La mitosis es la parte del ciclo

celular en la que el núcleo se divide en dos. Muchas mutaciones ocurren durante la replicación del DNA (Fase S).

Por efecto del azar (fenómenos aleatorios) se pueden cometer errores al copiar el DNA. Una de las formas en que el azar se manifiesta es a través de las fluctuaciones cuánticas a nivel molecular que causan errores en el copiado del DNA. Una de las proteínas responsables se llama polimerasa. Es una enzima que tiene una fidelidad de copiado mayor a 99%, pero menor que 100%. Al cometerse un error de copiado (menor del 0.1% de las veces) se genera una nueva variante del gen como se observa en la Figura 2-18. A esto se le llama un polimorfismo. Esta mutación puntual puede generar un nuevo alelo que puede tener una mejor o peor eficiencia que el original. Estas mutaciones se van acumulando durante sucesivas generaciones, de forma que con el tiempo, el alelo puede adquirir una función diferente.

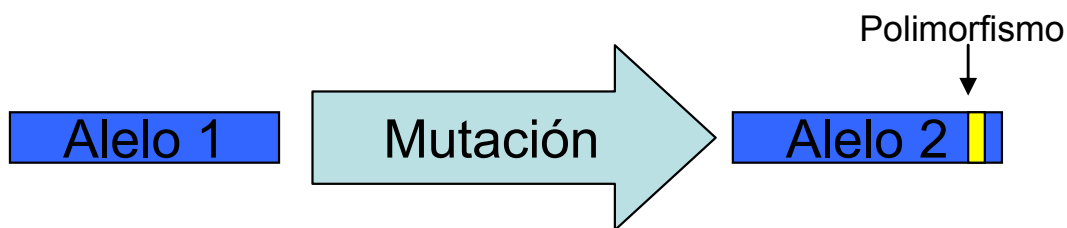


Figura 2-18. Las mutaciones dan origen a nuevas variantes de genes.

Si el proceso mutagénico se combina con una duplicación de ese gen (por ejemplo, un alelo en el cromosoma 1 se puede copiar e insertar en el cromosoma 4, obteniéndose dos copias del mismo alelo en distintos loci del genoma). Por medio de mutaciones acumulativas las diferentes variantes se diversifican cada vez más. Los alelos evolucionan de manera independiente. Con el transcurso del tiempo se obtienen genes que pertenecen a una misma familia, pero que han adquirido funciones diferentes. A estas variantes duplicadas ya no se les llama alelos sino familias de genes. Ver Figura 2-19.

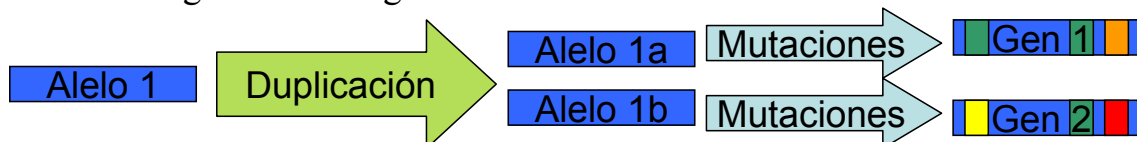


Figura 2-19. La duplicación y mutación divergente genera familias de genes.

El proceso de duplicación de genes es muy común en los organismos superiores. Cuando se secuenció el genoma completo de *Arabidopsis thaliana* en el año 2000, se advirtió que existían muchos bloques de genes repetidos en diferentes partes del genoma. La Figura 2-20 muestra los 5 cromosomas de *Arabidopsis* (líneas horizontales blancas). Las líneas verticales de colores muestran los grupos de genes que están repetidos en

diferentes cromosomas. Se puede observar que no son pocos, sino miles de genes que están duplicados en esta planta.

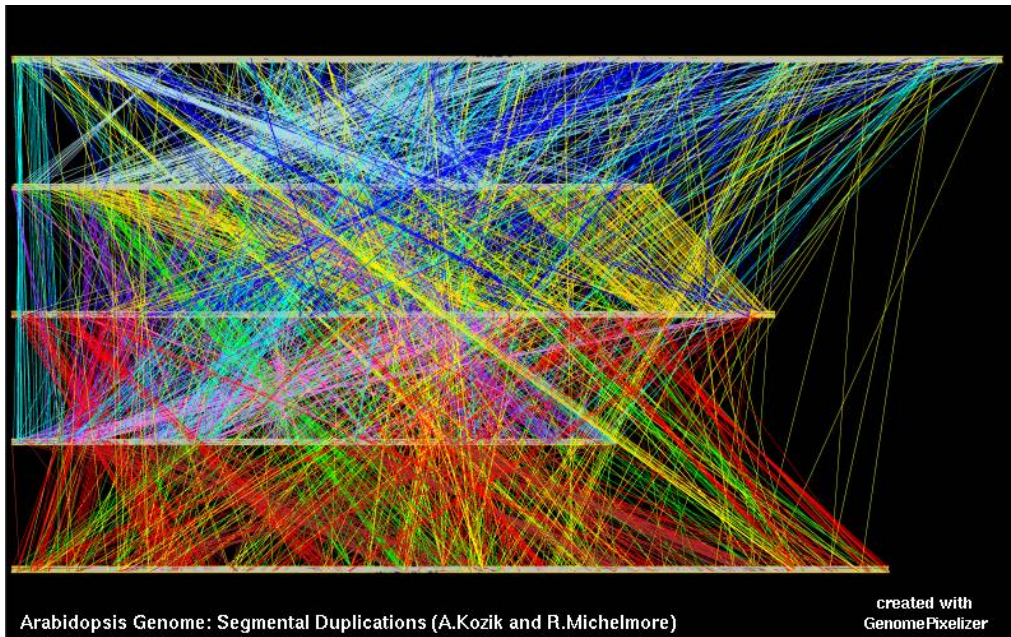


Figura 2-20. Duplicaciones segmentales del genoma de *Arabidopsis*.

Tarea 2-22

Consiga los siguientes dos artículos y lealos:

- Blanc, G., Hokamp, K., and Wolfe, KH. (2008). A Recent Polyploidy Superimposed on Older Large-Scale Duplications in the Arabidopsis Genome. *Genome Res.* 2003 13: 137-144
- The Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408. p796
- Traduzca al español el siguiente texto:

Abstract:

The flowering plant *Arabidopsis thaliana* is an important model system for identifying genes and determining their functions. Here we report the analysis of the genomic sequence of *Arabidopsis*. The sequenced regions cover 115.4 megabases of the 125-megabase genome and extend into centromeric regions. The evolution of *Arabidopsis* involved a whole-genome duplication, followed by subsequent gene loss and extensive local gene duplications, giving rise to a dynamic genome enriched by lateral gene transfer from a cyanobacterial-like ancestor of the plastid. The genome contains 25,498 genes encoding proteins from 11,000 families, similar to the functional diversity of *Drosophila* and *Caenorhabditis elegans* the other sequenced multicellular eukaryotes. *Arabidopsis* has many families of new proteins but also lacks several common protein families, indicating that the sets of common proteins have undergone differential expansion and contraction in the three multicellular eukaryotes. This is the first complete genome sequence of a plant and provides the foundations for more comprehensive comparison of conserved processes in all eukaryotes, identifying a wide range of plant-specific gene functions and establishing rapid systematic ways to identify genes for crop improvement.

Si bien, antes se pensaba que el genoma de un organismo era estático y fijo, hoy en día se sabe que existe muchísima plasticidad durante la evolución genómica. Por ejemplo: en maíz se han secuenciado segmentos de cromosomas de diferentes variedades. Al comparar los segmentos supuestamente homólogos se han encontrado que puede haber diferencias muy grandes, hasta 30% de los genes pueden faltar o sobrar en alguna de las variedades. En la Figura 2-21 se muestra el DNA de dos variedades que comparten tres genes, mientras que el gen 3 y 5 sólo están presentes en una de las variedades, en la otra variedad tiene genes únicos.

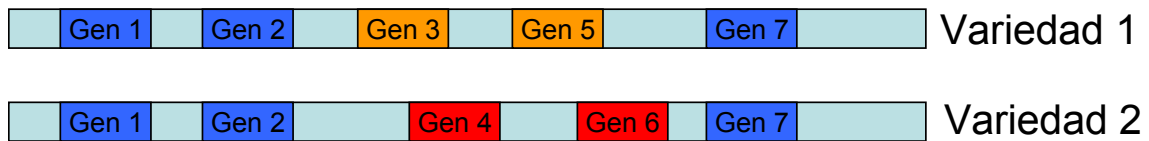


Figura 2-21. Algunas líneas homocigóticas de maíz tienen diferentes genes en regiones homologas de sus cromosomas.

Tarea 2-23

Consiga el siguiente artículo y lealo:

Brunner, S., Fengler, K., Morgante, M., Tingey, S., and Rafalskia A. (2005). Evolution of DNA Sequence Nonhomologies among Maize Inbreds. *Plant Cell* 17, p 343–360.

Abstract:

Allelic chromosomal regions totaling more than 2.8 Mb and located on maize (*Zea mays*) chromosomes 1L, 2S, 7L, and 9S have been sequenced and compared over distances of 100 to 350 kb between the two maize inbred lines Mo17 and B73. The alleles contain extended regions of nonhomology. On average, more than 50% of the compared sequence is noncolinear, mainly because of the insertion of large numbers of long terminal repeat (LTR)-retrotransposons. Only 27 LTR-retroelements are shared between alleles, whereas 62 are allele specific. The insertion of LTR-retrotransposons into the maize genome is statistically more recent for nonshared than shared ones. Most surprisingly, more than one-third of the genes (27/72) are absent in one of the inbreds at the loci examined. Such nonshared genes usually appear to be truncated and form clusters in which they are oriented in the same direction. However, the nonshared genome segments are gene-poor, relative to regions shared by both inbreds, with up to 12-fold difference in gene density. By contrast, miniature inverted terminal repeats (MITEs) occur at a similar frequency in the shared and nonshared fractions. Many times, MITES are present in an identical position in both LTRs of a retroelement, indicating that their insertion occurred before the replication of the retroelement in question. Maize ESTs and/or maize massively parallel signature sequencing tags were identified for the majority of the nonshared genes or homologs of them. In contrast with shared genes, which are usually conserved in gene order and location relative to rice (*Oryza sativa*), nonshared genes violate the maize colinearity with rice. Based on this, insertion by a yet unknown mechanism,

rather than deletion events, seems to be the origin of the nonshared genes. The intergenic space between conserved genes is enlarged up to sixfold in maize compared with rice. Frequently, retroelement insertions create a different sequence environment adjacent to conserved genes.

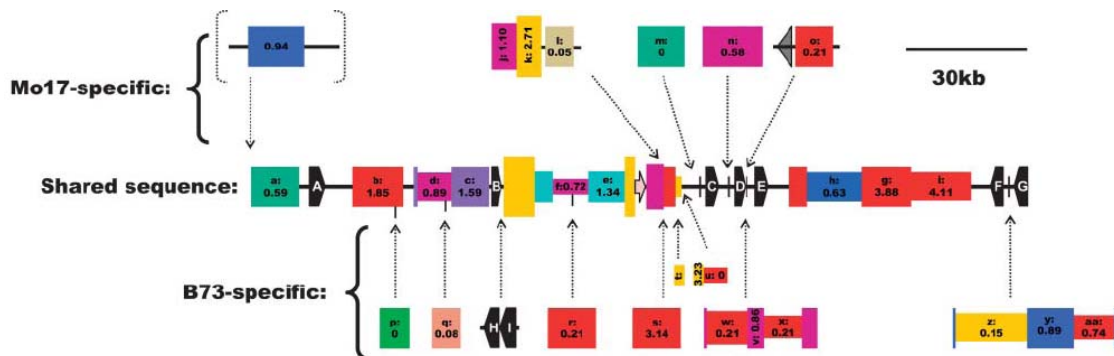


Figura 2-22. Diagrama que muestra como algunos genes estan presentes solo en la línea Mo17, mientras que otras secuencias sólo están en B73. Figura tomada de Brunner *et al.* (2005).

2.3.2. Recombinación

Además de las mutaciones, otro mecanismo que contribuye de manera muy significativa a la formación de nuevos alelos y genes es la recombinación. Durante la meiosis los dos alelos de un padre diploide se aparean y pueden intercambiar segmentos (*crossing over*), formándose así dos nuevos alelos, como se muestra en la Figura 2-23.

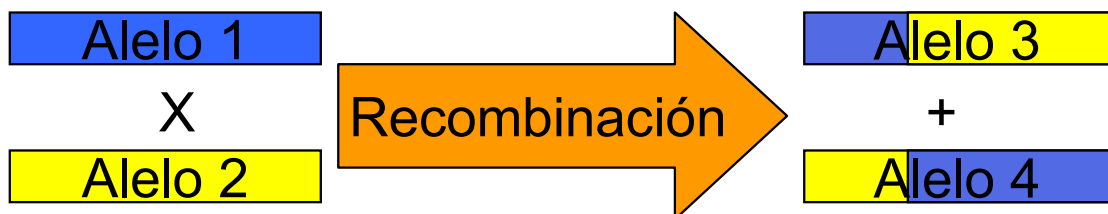


Figura 2-23. La recombinación genética crea nuevos alelos.

Un atributo de este mecanismo es que permite generar variantes con mayor probabilidad de ser ventajosas. Una mutación al azar tiene una gran probabilidad de ser letal, mientras que la recombinación puede generar una variante que combine las ventajas del 5' de un alelo, con las características superiores de la región 3' de otro alelo. De esta forma se puede generar un alelo recombinante mejorado (un superalelo).

De hecho, parece ser que la organización de los genes en intrones y exones, facilita este tipo de recombinaciones entre diferentes alelos, pues así se pueden combinar y piramidizar módulos de diferentes proteínas y, de esta manera, generar genes con funciones novedosas. A este concepto evolutivo se le llama **Exon-Shuffling** (Figura 2-24).

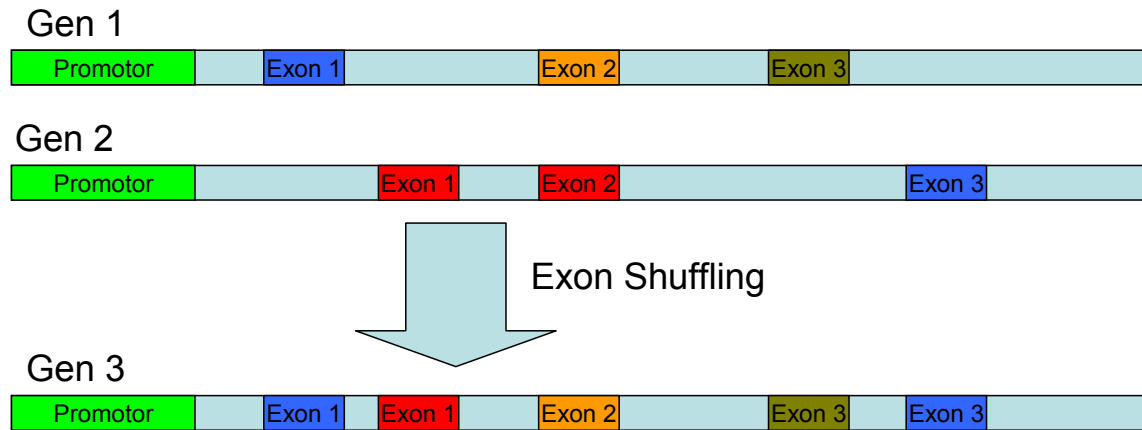


Figura 2-24. Exon shuffling por recombinación.

Tarea 2-24

Consiga los siguientes artículos y lealos:

Patthy L (1999) Genome evolution and the evolution of exon-shuffling - a review. *Gene* 238: 103-114

Morgante M, Brunner S, Pea G, Fengler K, Zuccolo A, Rafalski A (2005) Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. *Nature Genetics* 37: 997-1002

Liu MY, Grigoriev A (2004) Protein domains correlate strongly with exons in multiple eukaryotic genomes - evidence of exon shuffling? *Trends in Genetics* 20: 399-403

Es curioso notar que el genoma de las bacterias se caracteriza por la ausencia de exones e intrones, mientras que los genomas de organismos superiores, como el maíz o el ser humano, están repletos de ellos. Si analizamos la historia evolutiva, podríamos constatar que el genoma de *Zea mays* ha evolucionado mucho más rápido que el genoma de una bacteria como *Escherichia coli*. ¿A qué se debe esto?

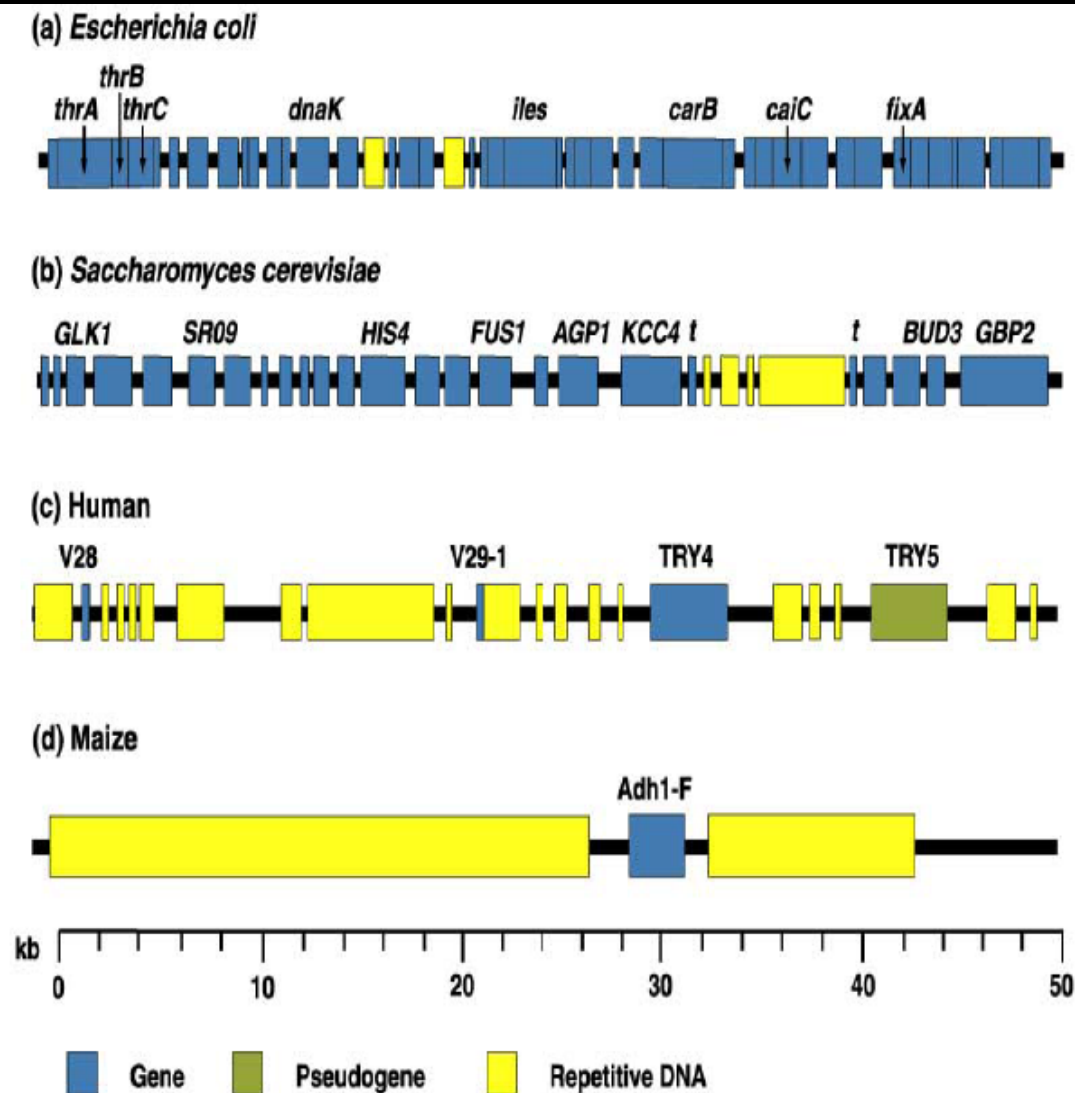


Figura 2-25. Comparación del genoma de *E. coli*, *S. cerevisiae*, humano y maíz. Se muestran las diferencias en la densidad génica y secuencias repetitivas.

Nadie lo sabe a ciencia cierta. Tal vez sea que los intrones y el “DNA basura” de los organismos superiores facilita la recombinación de exones y de esta forma se facilita la **innovación genética**. Otros procesos que podrían contribuir a este fenómeno, es la ploidía y los procesos de transferencia genética por vía sexual. La sexualidad incrementa la tasa de recombinación mucho más allá de lo que las bacterias pueden lograr por medio de la **conjugación bacteriana**. Todo esto combinado, es lo que ha permitido que algunos organismos evolucionen más rápido que otros, aún cuando los ciclos generativos sean mucho más largos. Por ejemplo, un ciclo generativo de *E. coli* dura menos de 1 h mientras que un ciclo de maíz dura más de 3 meses. Y sin embargo, pareciera que el maíz ha evolucionado más que la bacteria *coli* en los últimos cien años.

Esta es una noción evolutiva relevante para el mejoramiento genético, ya que cualquier mecanismo que nos ayude a acelerar la innovación genética facilita nuestra tarea de mejoramiento, ya sea, incrementando nuestra probabilidad de que la variante sea ventajosa o ayudándonos a lograrlo de manera más rápida o eficiente.

Tarea 2-25

Formule una teoría/hipótesis que explique las diferencias entre las tasas evolutivas de bacterias, animales y plantas. ¿Qué experimento podría hacer para probar su hipótesis?

2.3.3. Transferencia horizontal

En la sección anterior vimos que un gen se puede duplicar dentro de una misma especie. El gen duplicado puede quedarse cerca del gen original o transferirse entre los diferentes cromosomas (ver Figura 2-20). ¿Podríamos entonces pensar que este proceso de duplicación y transferencia genética también se lleva a cabo entre los genomas de diferentes especies? No es difícil imaginar que sí puede suceder. A este fenómeno se le llama **transferencia genética horizontal** (Figura 2-26) y debemos distinguirlo de la **transferencia genética vertical** (Figura 2-27) que se refiere al proceso en donde los progenitores transfieren una parte de sus genes a sus hijos de manera sexual (fenómeno de herencia).

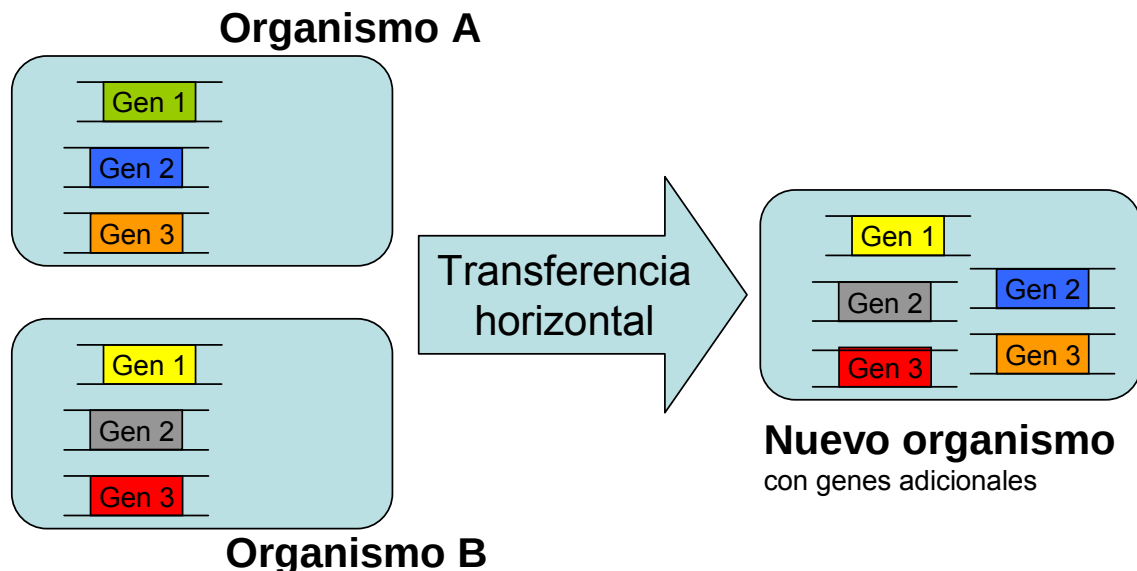


Figura 2-26. Transferencia genética horizontal.

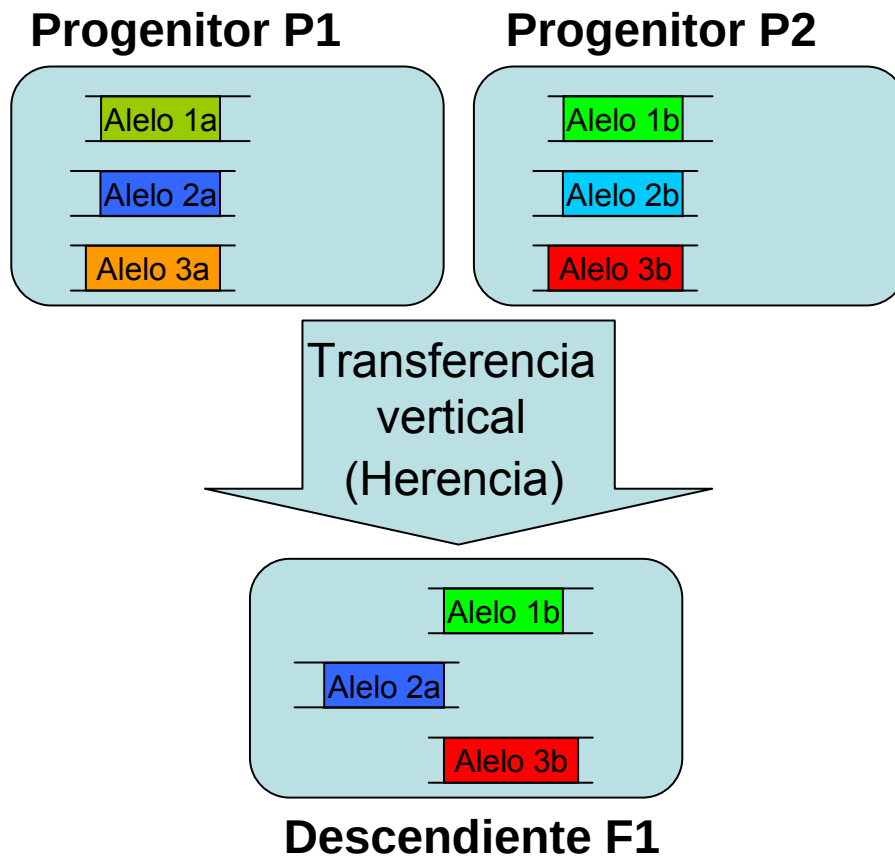


Figura 2-27. Transferencia genética vertical.

La transferencia genética horizontal es un proceso muy común en las bacterias y los microorganismos que comparten un mismo hábitat. Esto se debe, a que las bacterias tienen mecanismos especiales para transferir el DNA entre diferentes células. Este proceso se denomina **conjugación bacteriana** y se lleva a cabo por medio de estructuras llamadas *sex pilli*. Sin embargo, las bacterias también pueden adquirir DNA de fuera de la célula, en un proceso conocido como **transformación genética** y es uno de los pilares metodológicos con mayor importancia de la biología molecular.

Tarea 2-26

Averigüe por internet cuales son los protocolos de transformación de bacterias mas usados en la actualidad por los laboratorios de investigación. Haga un comparativo entre el protocolo de choque térmico (*heat shock*) y el protocolo de electroporación. ¿Cuál es más fácil? ¿Cuál es más eficiente? ¿Qué equipo requiere cada protocolo?

Tarea 2-27

Consiga los siguientes artículos y lealos:

Koonin EV, Makarova KS, Aravind L (2001) Horizontal gene transfer in prokaryotes: Quantification and classification. *Annual Review of Microbiology* 55: 709-742

Philippe H, Douady CJ (2003) Horizontal gene transfer and phylogenetics. *Current Opinion in Microbiology* 6: 498-505

Eisen JA (2000) Horizontal gene transfer among microbial genomes: New insights from complete genome analysis. *Current Opinion in Genetics & Development* 10: 606-611

Richards TA, Soanes DM, Foster PG, Leonard G, Thomson CR, Talbot NJ (2009) Phylogenomic analysis demonstrates a pattern of rare and ancient horizontal gene transfer between plants and fungi. *Plant Cell* 21: 1897-1911

2.3.4. Selección

La selección natural afecta de gran manera la cantidad y la calidad de la variabilidad genética. En realidad no es un proceso por el cual se produzcan nuevas variantes, sino todo lo contrario. Por medio de la selección se reduce y se restringe enormemente la diversidad genética. Casi el 99% de la diversidad de especies que han existido en nuestro planeta se han extinguido a causa de la selección natural. Es decir, la selección es una **fuerza que elimina** y de esta forma moldea la diversidad genética que sobrevive y puede reproducirse. Podemos decir entonces que las mutaciones son parte de una **fuerza creativa**, mientras que la selección es una **fuerza destructiva** y que la evolución se da por el balance entre estas dos grandes fuerzas, que en cierto sentido son opuestas (ver Figura 2-28).

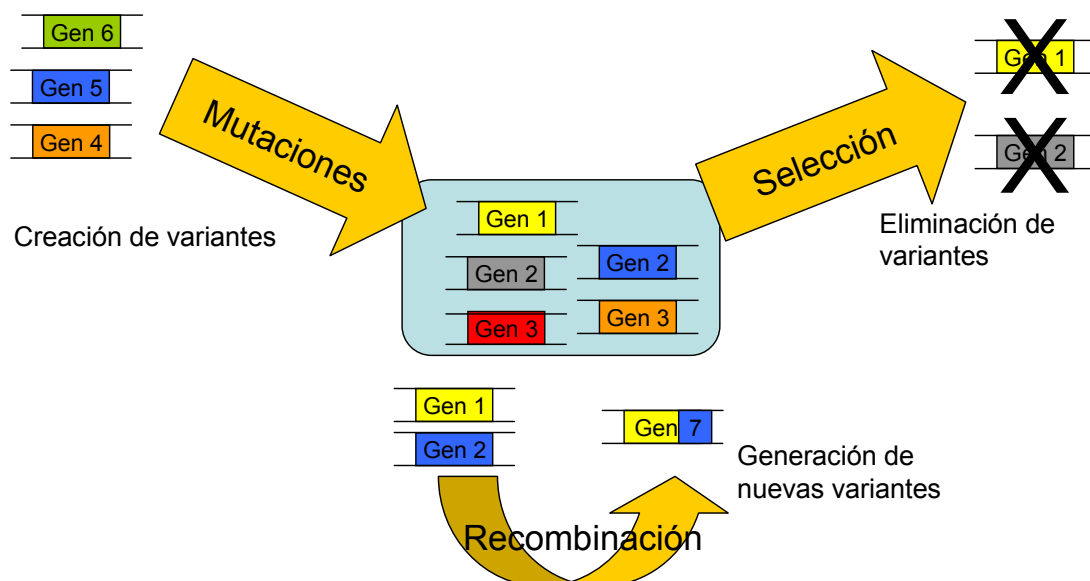


Figura 2-28. Fuerzas evolutivas de mutación, recombinación y selección.

La **selección natural** puede dividirse en varias categorías:

- La sexual, ocurre cuando los organismos son más atractivos para el sexo opuesto debido a sus características. Los alelos favorables se reproducen más y aumentan la frecuencia de estas características en el patrimonio genético común.
- La ecológica, ocurre en el resto de las circunstancias como: habilidad para obtener o procesar alimento; capacidad de esconderse, huir o defenderse; capacidad para resistir fluctuaciones ambientales; etc.

La selección en muchos casos implica una muerte antes de llegar a la etapa reproductiva. Cabe recalcar que la muerte del individuo después de la etapa reproductiva no se puede considerar como selección natural. Cuando un individuo ya ha procreado, sus alelos ya se han transferido a la siguiente generación. La muerte de individuos sin capacidad de reproducción (por ejemplo: abejas y hormigas obreras) tampoco se considera selección natural.

El papel central de la selección natural en la teoría de la evolución ha dado origen a una fuerte conexión con el estudio de la ecología. Las interacciones ecológicas entre diferentes organismos son uno de los factores determinantes para el *fitness* y el éxito evolutivo. Esto es lo que Darwin llamo *struggle for life* y Haeckel después describió como *survival of the fittest*.

La selección natural altera las frecuencias de los alelos de diferentes formas:

- La **selección negativa** elimina las mutaciones perniciosas de una población. A este fenómeno también se le ha llamado selección purificadora, porque sólo permite que sobrevivan un número limitado de variantes.
- La **selección positiva** aumenta la frecuencia de variantes benéficas, puesto que los organismos portadores tienen más éxito y producen más descendencia, que a su vez sobrevive y puede reproducirse y por consiguiente, esa ventaja selectiva de generación en generación causa un cambio en las frecuencias génicas de la población.
- La **selección de balanceo** mantiene ciertas variaciones dentro de una población, debido a que los alelos en algunos casos tienen un efecto de *fitness* positivo y en otros casos un efecto de *fitness* negativo. La selección de balanceo se da a través de mecanismos tales como:
 - La sobredominancia o **vigor híbrido**. Este fenómeno se debe a que un alelo en estado heterocigoto puede ser ventajoso, pero cuando se encuentra en estado homocigoto puede ser desfavorable.

- La **selección dependiente de la frecuencia**, debido a que el efecto de un alelo puede ser ventajoso cuando solo está presente con muy baja frecuencia que cuando está presente con mayor frecuencia en la población, ya que ese alelo no confiere una ventaja de *fitness*.

Las mutaciones que no se ven afectadas por la selección natural, son llamadas **mutaciones neutras**. Su frecuencia en la población está dictada por su **tasa de mutación**, por la **deriva genética** y el **flujo genético**. Se entiende que la secuencia de DNA de un organismo, en ausencia de selección, sufre una acumulación de mutaciones neutras. El efecto probable de las mutaciones es la diversificación. Un gen que no está bajo selección será destruido por las mutaciones acumuladas. Este es un aspecto de lo que se pudiera llamar “degradación genómica”.

2.4. El azar

2.4.1. Un fenómeno universal

El azar tiene muchos nombres y muchas caras. Algunos lo llaman aleatoriedad, suceso fortuito o fenómeno estocástico. Otros lo llaman serendipia o suerte. En nuestras vidas nos confrontamos diariamente con fenómenos de incertidumbre. ¿Irás a llover hoy? ¿Y si me deja el autobús? ¿Llegaré tarde al trabajo por un accidente de tránsito? Existe un dicho alemán que establece que en nuestras vidas no podemos tener certeza de nada; sólo la muerte es cierta, necesaria e inevitable. En los fenómenos naturales podemos preguntarnos, ¿qué tanto se da por azar y qué tanto se da por necesidad? Es decir, ¿es un evento que se repite muchas veces siguiendo una ley preestablecida, o es algo único, azaroso y fundamentalmente impredecible?

En un sentido científico, podemos decir que el azar es el fenómeno que se interpone a las leyes mecánicas de la naturaleza. El azar es el **error** que interfiere con la **repetibilidad**, es decir, el azar es la incertidumbre que rompe con la regularidad de los resultados. Por ejemplo: si partimos exactamente de las mismas condiciones y repetimos un experimento con las mismas variables, nuestra expectativa es que el resultado sea exactamente el mismo. Las mismas premisas bajo las mismas reglas deben de dar las mismas conclusiones. Es como en las matemáticas: uno más uno siempre da dos, ¿o no? El azar es ese fenómeno observable de la vida real que rompe con esa regularidad. El azar es como la excepción a la teoría de las matemáticas. Si no fuera por el azar, todo sería mucho más fácil de pronosticar. El azar es algo inesperado e impredecible, generando variantes en los resultados, de forma que un día, uno más uno dan dos, mientras que en otra ocasión, uno más uno dan tres. Es como si fuera un resultado

fortuito. Por lo regular decimos que esa variación del resultado se da de forma espontánea. Pero, ¿a qué se debe esa espontaneidad? ¿Cuál es la causa? ¿Es un capricho de la naturaleza? ¿Es algo mágico? Si analizamos los fenómenos fortuitos con cuidado, podemos concluir que son igual de misteriosos que el azar. Tal vez sea sólo otra palabra para el mismo fenómeno ¿Podríamos decir que las fluctuaciones se dan por una fuerza real que no entendemos del todo, pero que llamamos azar? No suena tan descabellado. El azar ¿Es la causa inicial? Sí, tal vez sí la sea. Es una fuerza que rompe con la inercia de los resultados siempre iguales. Es una fuerza inicial, que a su vez ¿no tiene fin, motivo o razón? Podríamos decir de forma trascendental, que es una **causa sin causa**, una causa inicial o podríamos simplemente decir, que es azar.

Tarea 2-28

Haga una revisión sobre el tema del azar. Lea las siguientes referencias:

Hausdorff JM, Peng CK (1996) Multiscaled randomness: A possible source of 1/f noise in biology. *Physical Review E* 54: 2154-2157

Vogt G, Huber M, Thiemann M, van den Boogaart G, Schmitz OJ, Schubart CD (2008) Production of different phenotypes from the same genotype in the same environment by developmental variation. *Journal of Experimental Biology* 211: 510-523

Eble GJ (1999) On the dual nature of chance in evolutionary biology and paleobiology. *Paleobiology* 25: 75-87

HendersonSellers B (1996) Towards a conceptual model of randomness. *Ecological Modelling* 85: 303-308

Axel Tiessen Favier (2005) *The Power of Randomness. The fifth force of the universe.* Deutsche Wissenschafts-Verlag (DWV). Baden Baden, Germany. ISBN 3-935176-41-4

2.4.2. Fuerza innovadora y fuente de diversidad

En la física, el **concepto de fuerza** está definido como la **capacidad de alterar el estado de un objeto**. Por ejemplo, si un automóvil se encuentra en reposo (velocidad cero) una fuerza aplicada a lo largo de una distancia (una cantidad de energía), es capaz de cambiar el estado de reposo, generando así una variación en la velocidad del automóvil. La aceleración del automóvil será directamente proporcional a la fuerza que se aplique e inversamente proporcional a la masa inerte del automóvil. En la física, el concepto de **inercia** refleja esa noción de que todo permanece igual hasta que se aplica una fuerza que logra un cambio.

Usando los mismos conceptos de inercia y fuerza, podemos entonces aplicarlos al fenómeno molecular del azar. Un átomo radioactivo no va a cambiar su estado hasta que la fuerza del azar cause un efecto cuántico inesperado e impredecible. Ese cambio azaroso hace que el átomo se

descomponga emitiendo una partícula radioactiva y transformándolo en otro elemento. Es así como un átomo de carbono 14 se puede convertir en un átomo de nitrógeno 14 al emitir una partícula radioactiva beta. Una de los fenómenos que mas nos pueden impresionar en la naturaleza son los saltos cuánticos, cuando algo a se transforma en otra cosa, cuando una molécula dejo de ser para convertirse en otra.

¿Podemos entonces decir que el azar es la fuerza que rompe con la inercia del átomo? Es la fuerza que genera un cambio de estado. Sin el azar no habría los decaimientos radioactivos. Si no fuera por esa fuerza misteriosa, el átomo de carbono 14 seguiría siendo un átomo de carbono 14 por toda la eternidad. Si no fuera por el azar, todo se quedaría en el mismo estado inerte. Sin una fuerza externa, no puede haber cambios de inercia. Es por ello que en la física, el azar debe considerarse como una fuerza.

¿Qué se puede decir del azar en el contexto químico y biológico? ¿Es el azar la principal fuente de diversidad? ¿Podríamos decir que el azar es una **fuerza diversificadora** de los elementos químicos en el universo? Tal vez si. Por ejemplo, el azar evita que todos los átomos del universo sean solo de helio o de carbono, sino que constantemente se estén generando átomos de diferentes elementos y no siempre los mismos. Si hubiera una ley universal sobre la formación de los elementos, entonces todas las partes del universo deberían tener la misma composición química. Pero las observaciones que se han hecho es que si hay grandes diferencias entre los elementos que están presentes en una región u otra del espacio. ¿Porque la Tierra tiene una composición química tan diferente de Venus o de Marte? Si continuamos analizando otros ejemplos de la cosmología o de la física cuántica, nos sorprenderíamos de la multitud de fenómenos que tienen que ver con el azar y la incertidumbre.

En la biología, el azar también tiene un papel predominante. Un alelo va a seguir siendo el mismo alelo, hasta que un agente causante, es decir una fuerza mutagénica, haga que ese alelo cambie. Es decir, si la ley de la complementariedad de los nucleótidos del DNA debe de generar copias idénticas de los genes durante la replicación, el azar es esa fuerza, que interrumpe esa regularidad y genera variantes diferentes a las esperadas. El azar es la fuerza que altera la inercia de un gen a no cambiar. Actualmente la visión que predomina en la comunidad científica es que todas las **mutaciones biológicas se dan por el azar**. Esas mutaciones son las que generan polimorfismos genéticos que se van acumulando con el tiempo y de esa forma van incrementando la diversidad. Es decir, si analizamos el origen de la diversidad biológica, llegaremos tarde o temprano a la conclusión de que el azar es una fuerza innovadora que genera cambios continuos e impredecibles.

¿Cómo surgió la vida en nuestro planeta? Tal vez sí fue un evento químico y biológico bastante improbable. ¿Se dio de manera espontánea? ¿Se dio por necesidad a causa de una ley universal? Hasta ahora no hay evidencias científicas de que el inicio de la vida sea repetible. En el famoso experimento de Miller se generaron diversas moléculas orgánicas, pero ni una sola molécula de DNA y mucho menos una célula viva. ¿Fue un milagro el inicio de la vida en la Tierra primitiva? ¿Pudo haber sido por un capricho del azar? Tal vez no podamos contestar todas esas preguntas de manera científica, pero si podemos formular una hipótesis bastante tangible: **el azar es el origen de la diversidad en el universo**. Podemos darle muchas vueltas al asunto como filósofos, pero lo que nos correspondería como científicos, es hacer experimentos para comprobar o descartar esa hipótesis.

Tarea 2-29

Para realizar entre profesor y alumnos. Se sugiere trabajo en equipos y discusión plenaria.

¿Se le ocurre algún experimento que se pueda hacer para comprobar alguna hipótesis sobre el azar? Piense en algún sistema modelo, diseñe un experimento, escriba el protocolo y hágalo. Si no puede hacer el experimento en la vida real, concíballo en su imaginación. Discuta sus ideas con otros compañeros, estudiantes y maestros.

2.5. Evolución

2.5.1. Evolución biológica

Generalmente se denomina evolución a cualquier **proceso de cambio** en el tiempo. La palabra evolución se usa en el ámbito social, económico y biológico. Esto no implica que sea un cambio para bien o para mal, ni para mejor o peor. Es simplemente cambio. Es por ello que la evolución no debe tener ninguna connotación hacia una dirección fija o específica. Una de las mejores formas de representar el camino impredecible de la evolución es por medio de una famosa caricatura sobre la evolución del hombre.

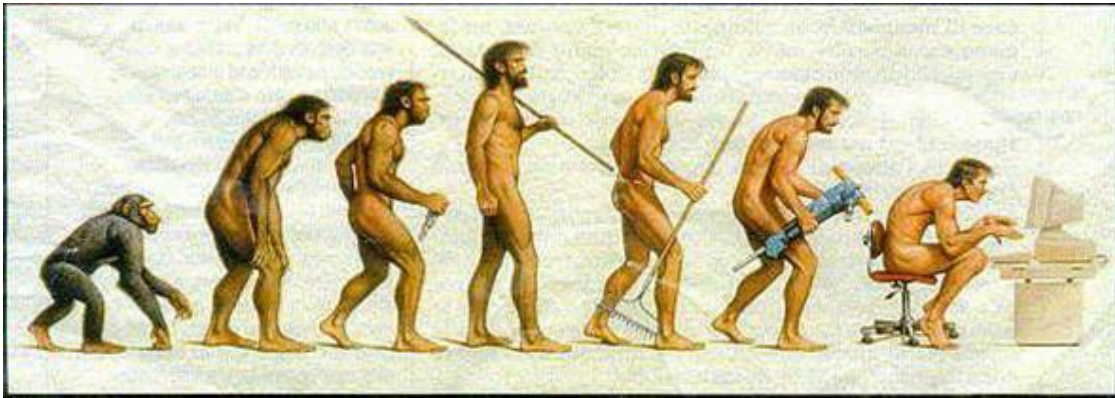


Figura 2-29. Caricatura de la evolución que representa un cambio de apariencia a lo largo del tiempo. La evolución puede ser hacia una dirección como hacia otra.

En el contexto de las ciencias de la vida, la **evolución** es un cambio, que puede llevar a la aparición de nuevas especies, a la adaptación a distintos ambientes o a la aparición de novedades morfológicas. La evolución biológica es el proceso continuo de transformación de las especies a través de cambios producidos en sucesivas generaciones y que se ve reflejado en el cambio de las frecuencias alélicas de una población. Es decir, es un cambio en el perfil genético de una población de individuos.

A menudo existe cierta confusión entre hecho evolutivo y teoría de la evolución. Se denomina **hecho evolutivo** a un suceso que está sustentando por abundante evidencia. Se refiere a hechos observables, comprobables e independientes. Por ejemplo, el hecho de que los seres vivos están emparentados entre sí y han ido transformándose a lo largo del tiempo. Esto por lo tanto se le llama hecho científico. La **teoría de la evolución** es el modelo científico que describe la transformación evolutiva y explica sus causas. ¿Cuál es la causa última de la evolución? La evolución como teoría puede tener algunas imperfecciones, pero la evolución como hecho es bastante obvia e irrefutable. Por ejemplo, puede ser que las teorías evolutivas actuales todavía le asignen pesos diferentes a las diferentes causas de la evolución, es decir, que le asignen una importancia diferente a las diversas fuerzas motrices, como selección natural, selección artificial, selección sexual, migración, mutación, deriva genética, azar, etc. Pero por otro lado, es un hecho evolutivo que los perros y los lobos tuvieron un ancestro en común. Hay suficiente evidencia para sustentarlo y se puede comprobar de manera experimental.

Casi nada en la biología tiene sentido, si no se mira desde la perspectiva de que los organismos se originaron unos de otros y que las especies se transforman con el tiempo. Si no se tiene esa noción de que cada generación hereda ciertas características de sus progenitores, pero que están acompañadas de algunos cambios, no tendría sentido intentar hacer mejoramiento genético. Las personas que dicen tener dudas sobre la teoría de la evolución, no es que no

crean en la evolución, sino más bien que tienen dudas sobre los procesos causales y las razones finales. Por ejemplo, la evolución ¿tiene un fin? ¿Existe una razón superior? ¿Hay diseño en la naturaleza? ¿Podemos hablar de un propósito? ¿Es el azar o una voluntad misteriosa? ¿Puede haber una inteligencia divina dirigiendo la evolución?

Tarea 2-30**Para realizar entre profesor y alumnos.**

¿Conoce algunos ejemplos de evolución de algún organismo? Haga una búsqueda bibliográfica y presente algunos indicios de evolución. Discuta la evidencia de manera crítica con otros compañeros, estudiantes y maestros.

Tarea 2-31

Busque información sobre alguna teoría científica sobre el origen de la vida. Puede empezar por la propuesta por Oparin y Haldane. ¿Qué es lo que propusieron? ¿En qué se basaron para hacer sus afirmaciones? ¿Con qué experimentos se podría confirmar esta teoría?

Tarea 2-32

¿Qué aportaciones hizo Stanley L. Miller y Harold Urey al conocimiento sobre el origen de la vida? Investigue los detalles sobre su famoso experimento.

Tarea 2-33

¿Qué obras ha escrito el biólogo mexicano Antonio Lazcano Araujo sobre el origen de la vida? Investigue más sobre sus artículos y compártalo con sus compañeros.

2.5.1.1. Teorías originales

El **Creacionismo**, es la visión de que en un grado u otro, los seres vivos tienen un creador personal, consciente y divino (léase Dios). El Creacionismo original asume que las especies son inmutables, es decir, que no cambian con el paso de las generaciones. En algún momento se dio un solo acto de creación que generó a todas las especies que conocemos actualmente. Es una posición religiosa o filosófica que no puede probarse experimentalmente y por tanto no es una teoría científica. No obstante, en el marco de la cultura popular anglosajona, algunos se esfuerzan por presentarla como tal. Pero la comunidad científica en su conjunto considera tales intentos como una forma de propaganda religiosa de carácter medieval en pleno siglo XXI. Una vertiente mas avanzada del Creacionismo, parte de múltiples actos de intervención divina, a partir de cuales se van creando nuevas especies de forma continua, a la par de que otras especies van desapareciendo por medio de selección y catástrofes.

El **Lamarckismo** es la suposición de que el fenotipo de un organismo puede dirigir de alguna forma el cambio del genotipo en sus descendientes. Es una posición teórica ya casi indefendible, en la medida en que es incompatible con lo que sabemos sobre la herencia y también porque todos los intentos por encontrar pruebas han fracasado. Sin embargo, a nivel molecular, todavía existen algunas dudas sobre la capacidad de las células de alterar su información genética para adaptarse al medio. Como ejemplos se pueden mencionar la inducción de mutaciones en las bacterias expuestas a condiciones ambientales adversas, o el sistema inmunológico de los humanos, en donde se inducen ciertos cambios en las regiones hipervariables para la formación de anticuerpos más selectivos después de la exposición previa con un antígeno patógeno. Todavía no se puede contestar con certeza si existe o no algún fenómeno de adaptación en la naturaleza, es decir, que se dé una modificación de los genes en dependencia del medio o la experiencia previa del organismo.

Robert **Chambers** fue el primer científico en postular una hipótesis que contradecía la noción generalizada de su época. La visión que predominaba era que el universo y, en lo particular, las especies biológicas eran inmutables. Esta hipótesis de cambios continuos la postuló en su libro "*Vestiges of the natural history of creation*" que fue publicado en Londres en 1844, mucho antes que el libro famoso de Darwin. Este libro fue un *bestseller* de la época. Contenía ideas tan radicales para ese entonces, que su autor decidió publicarlo de manera anónima, para no ser víctima de una inquisición del tipo religiosa. El problema más grave de las teorías de este tipo era que ponían en peligro la justificación divina del orden establecido; de ahí que, en general, los conservadores se opusieran con uñas y dientes a

ellas. Fue sólo hasta después de su muerte que se supo quien había sido el autor original de esa obra.

Charles **Darwin** y Alfred Russel **Wallace** propusieron la **selección natural** como principal mecanismo de la evolución. El libro más famoso de Darwin, "*On the Origins of species by means of Natural Selection*", fue publicado en **1859**, mucho después de que Darwin hubiera publicado sus memorias del viaje a Suramérica en los barcos de su majestad inglesa durante los años 1826-1836 (*Voyages of the Adventure and Beagle. Darwin 1839*). Hoy en día se ha generado el mito de que Darwin inventó la teoría de la evolución durante su estancia en las islas Galápagos. Si uno lee las memorias de su viaje en el Beagle, publicadas en 1839, son pocas las alusiones a su posterior teoría. Es probable que Darwin haya sido mas inspirado por los mejoradores y los criadores de pichones de la época. Darwin también fue apresurado por la carta que le envió otro naturalista que trabajaba en lo que hoy en día es Nueva Guinea, Java, Borneo e Indonesia.

Es notable que el joven Wallace se le adelantara a Darwin en la publicación de sus ideas. En 1855 Wallace publicó un artículo titulado "*On the law which has regulated the introduction of new species*", donde defendía el hecho de la evolución, aunque sin atribuirle una causa. Tres años más tarde, un nuevo artículo "*On the tendency of varieties to depart indefinitely from the original type*", proponía la selección natural como el mecanismo explicativo de la **transmutación** de las especies. Wallace remitió el artículo a Darwin para su revisión. Cuando este lo leyó, se encontró con lo que calificó como el mejor resumen imaginable de las ideas que él mismo llevaba gestando trabajosamente desde hacía más de veinte años. Tras consultar con Charles Lyell, Darwin realizó una presentación pública de su manuscrito ante la Sociedad linneana de Londres el 1 de julio de 1858, acreditando a Wallace como codescubridor.

Cita completa del trabajo original:

Darwin, C. R. and A. R. Wallace. 1858. *On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection*. [Read 1 July] Journal of the Proceedings of the Linnean Society of London. Zoology 3 (20 August): 46-50. Communicated by Sir CHARLES LYELL, F.R.S., F.L.S., and J. D. HOOKER, Esq., M.D., V.P.R.S., F.L.S

Es interesante notar que el término de evolución originalmente se usaba para referirse a la historia humana en un contexto social. El término que se usaba para el cambio de las especies biológicas a lo largo del tiempo era **transmutación**. En esa época el concepto de mutación se refería al cambio

en la geología y no como se usa actualmente en contexto biológico. Hoy en día nadie habla ya de transmutación, aunque en esencia, las mutaciones son las verdaderas causantes de que se formen nuevas especies y no la selección natural como postuló Darwin originalmente. La selección natural sólo se limita a eliminar individuos que no sobreviven en la lucha por la vida (*struggle for life*) y, por consiguiente, no se puede eliminar algo que no se ha creado anteriormente. En otras palabras: sin previa mutación que genera diversidad, no puede haber selección.

En la época de Darwin, los científicos no conocían cómo se heredaban las características. No se sabía nada de los genes ni del DNA. Darwin tampoco conocía la fuente de las variaciones en los organismos individuales, pero sí observó que siempre estaban presentes y además, que parecían ocurrir aleatoriamente. En trabajos posteriores otros científicos atribuyeron la mayor parte de estas variaciones a las mutaciones en los genes. Es por ello que hoy en día, un título mas adecuado para un libro de evolución sería algo como: *On the origins of species by means of transmutation, and the mechanism of adaptation by means of natural selection.*

La **Teoría de la Evolución de Darwin** ha sido enriquecida desde su formulación en el siglo XIX, por avances en otras disciplinas como: la biología molecular, la genética del desarrollo, embriología o la paleontología. Wallace y Darwin fueron los primeros en postular un mecanismo evolutivo por medio de la selección natural. La teoría más popular hoy en día es la síntesis de las ideas de muchos científicos.

En la actualidad sabemos que en algunos aspectos Darwin se equivocó. Darwin fue un pionero del pensamiento para su época, aunque en realidad Robert Chambers fue mucho más adelantado con su libro de *Vestiges*. Hoy en día sabemos que las ideas en el libro de Chambers tuvieron muchas fallas. De igual forma sería erróneo e impreciso decir que Darwin tuvo razón en todo. De hecho, más que los biólogos, son los sociólogos e historiadores que entienden como algunos de los grandes científicos tuvieron razón en parte, pero que sus teorías originales mostraron algunos errores que poco a poco se han ido modificando y refinado. ¿Cómo iba a saber Darwin algo sobre las mutaciones del DNA, si ni siquiera se conocía nada de los genes en su época? Es importante tener presente que los grandes descubridores, pensadores o sabios, fueron y son humanos, capaces de equivocarse y con frecuencia lo hacen. Nos podemos dar cuenta de ello si después de varios siglos estudiamos todos los aspectos de sus teorías con máximo detalle. Lo mismo que de Lineus o Darwin en la biología podemos decir de Kepler, Copérnico o Newton en la física. Eso no significa que las teorías originales sean erróneas, simplemente que estuvieron incompletas y que algunos hechos hoy en día se pueden interpretar mucho mejor. Podemos decir que las teorías científicas también evolucionan y se enriquecen con el paso del tiempo, las generaciones de nuevos científicos y la aparición de mentes brillantes.

Tarea 2-34

Consiga el libro o en su caso por Internet, el texto original de Charles Darwin (*On the origins of species by means of natural selection*). Puede encontrarlo en la página web del proyecto Gutenberg (www.gutenberg.org). Otro sitio donde puede encontrar todas las obras originales de Darwin es: <http://darwin-online.org.uk>. Lea cuidadosamente la introducción de *Origins* y el resumen y trate de ver si encuentra algunas diferencias con respecto a la teoría actual. ¿Cuál es el verdadero papel de la selección natural? ¿Es una fuerza creadora que genera nuevas especies o es una fuerza destructiva que elimina variantes? ¿A qué se refiere Darwin cuando habla de mutación? ¿A quién le daría más crédito científico: a Chambers, a Darwin o a Wallace? ¿Por qué?

Es interesante notar que gran parte de la argumentación del texto de Darwin no proviene de las islas Galápagos como generalmente se cree, sino principalmente de la domesticación de animales y del mejoramiento genético. Es decir, Darwin usó el concepto de **selección artificial** que ya se conocía en su época y postuló que un mecanismo similar funcionaba en la naturaleza. Eso fue a lo que le llamo **selección natural**. Lo revolucionario de su texto no fue el concepto de selección *per se*, sino que dijo que existía una selección sin la supervisión del ser humano (selección natural). Para muchas personas de la época esto era inconcebible, ya que el concepto de selección implicaba que alguna inteligencia superior debía estar supervisando este proceso, así como los mejoradores escogían sus pichones favoritos y tomaban decisiones de cómo debían hacer las cruzas entre individuos. Históricamente, el ser humano tiene mayor experiencia empírica sobre el mejoramiento artificial que sobre la evolución natural. Esto está documentado, puesto que Darwin dedicó varios libros y capítulos a la crianza de palomas, ovejas, etc.

Tarea 2-35

Consulte la página <http://darwin-online.org.uk> y haga una lista de los trabajos originales de Darwin que tuvieron como tema la domesticación y el mejoramiento de las especies.

2.5.1.2. Teoría actual (síntesis moderna)

Actualmente, la teoría de la evolución combina las propuestas de Darwin y Wallace, las interpretaciones de Huxley, con las leyes de Mendel y otros avances genéticos posteriores; por eso es llamada Síntesis Moderna o **Teoría Sintética de la Evolución**. Los principales contribuyentes a esta síntesis fueron Thomas Hunt Morgan, R. A. Fisher, Theodosius Dobzhansky, J.B.S. Haldane, Sewall Wright, William Donald Hamilton, Cyril Darlington, Julian Huxley, Ernst Mayr, George Gaylord Simpson, y G. Ledyard Stebbins.

Tarea 2-36

Consiga y lea algunas de las publicaciones que definieron la síntesis moderna de la evolución. ¿Qué aportaciones hizo cada uno?

- Fisher, R. (1930) *The Genetical Theory of Natural Selection* [ISBN 0-19-850440-3](#).
- Haldane, J. B. S. *The Causes of Evolution*, Longman, Green and Co., 1932; Princeton University Press reprint, [ISBN 0-691-02442-1](#)
- Huxley, J. S. *Evolution: The Modern Synthesis*, Allen and Unwin, 1942 [ISBN 0-02-846800-7](#)

En el seno de esta teoría, la evolución se define como un cambio en la frecuencia de los alelos en una población a lo largo de las generaciones. Este cambio puede ser causado por uno o varios mecanismos diferentes: selección natural, deriva genética, mutación, migración (flujo genético).

La evolución biológica es un fenómeno natural, real, observable y comprobable empíricamente. La **Síntesis Evolutiva Moderna** es una teoría robusta y ampliamente aceptada por la comunidad científica. Aunque algunos detalles son cuestionados por algunos investigadores. Actualmente proporciona explicaciones y modelos matemáticos sobre los mecanismos generales de la evolución o los fenómenos evolutivos, como la adaptación o la especiación. Como cualquier teoría científica, sus hipótesis están sujetas a constante crítica y comprobación experimental.

Dobzhansky, uno de los fundadores de la síntesis evolutiva moderna, la definió del siguiente modo: "La evolución es un cambio en la composición genética de las poblaciones. El estudio de los mecanismos evolutivos corresponde a la genética poblacional."

La síntesis moderna de la evolución se basa en tres aspectos fundamentales:

- La ascendencia común de todos los organismos a partir de un único ancestro.
- El origen de nuevos caracteres en un linaje evolutivo.

- Los mecanismos por los que algunos caracteres persisten mientras que otros desaparecen.

Las mutaciones introducen nuevas variaciones genéticas, siendo la principal fuente de diversificación y de evolución. En la teoría sintética, la mutación tiene el papel de generar diversidad genética sobre la cual actúa la selección natural y también la deriva génica. Las mutaciones que afectan a la eficacia biológica del portador, y por tanto son objeto de la selección natural, pueden ser deletéreas (negativas), beneficiosas (positivas) o bien neutras. Las mutaciones beneficiosas son las menos frecuentes, aunque se conocen muchos ejemplos que afectan a rasgos específicos, como la resistencia a enfermedades o a estrés, la longevidad, el tamaño, la capacidad para metabolizar nuevas sustancias, etc. La mayor parte de las mutaciones son neutras; no afectan las oportunidades de supervivencia y reproducción de los organismos y se acumulan con el tiempo a una velocidad más o menos constante. Algunas mutaciones son deletereas, por ejemplo: cuando se interrumpe un gen que es importante para mantener a una célula viva. En esos casos se dice que la mutación es letal, y por consiguiente, desaparece rápidamente del fondo genético de la población por la presión selectiva tan alta.

La mayoría de los biólogos creen que la adaptación ocurre fundamentalmente por selección natural en etapas, mediante la acumulación de variaciones genéticas ventajosas con efectos relativamente pequeños. Las macromutaciones, por el contrario, producen efectos drásticos, fuera del intervalo de variación normal de la especie. Se ha propuesto que quizá hayan sido responsables de ciertos rasgos adaptativos o de la aparición de novedades evolutivas, aunque, dado que algunas mutaciones suelen tener efectos muy nocivos o letales, esta vía se considera actualmente poco frecuente, pero no por eso menos importante.

Las duplicaciones génicas ocurren cuando las familias multigénicas están presentes en muchos genomas; por ejemplo en las globinas. Pueden surgir por duplicación de todo el genoma o pequeños bloques o genes. Los mecanismos de duplicación incluyen: entrecruzamiento desigual, errores de replicación, transposición, poliploidía, etc. Por ejemplo: el 45% del genoma humano está compuesto de **transposones**. En maíz, se cree que el genoma se duplicó antes de ser domesticado, por lo que se dice que es una especie autotetraploide.

La selección natural también consiste en la reproducción diferencial de los individuos, según su dotación genética, y generalmente como resultado del ambiente. Existe selección natural cuando hay diferencias en eficacia biológica entre los individuos de una población, es decir, cuando su

contribución en descendientes es desigual. La eficacia biológica puede desglosarse en componentes como: la supervivencia (la mortalidad diferencial es la tasa de supervivencia de individuos hasta la edad de reproducción), la fertilidad, la fecundidad, etc.

Tarea 2-37

Lea los siguientes artículos:

Koonin EV (2009) Darwinian evolution in the light of genomics. *Nucleic Acids Research* 37: 1011-1034

Nowak MA, Ohtsuki H (2008) Prevolutionary dynamics and the origin of evolution. *Proceedings of the National Academy of Sciences of the United States of America* 105: 14924-14927

Carroll RL (2000) Towards a new evolutionary synthesis. *Trends in Ecology & Evolution* 15: 27-32

Boto L, Doadrio I, Diogo R (2009) Prebiotic world, macroevolution, and darwin's theory: A new insight. *Biology & Philosophy* 24: 119-128

2.5.2. Evolución dirigida

Algunas personas tienen una visión progresista de la evolución. Por ejemplo: de que el destino de las bacterias primitivas es la de evolucionar hacia convertirse en organismos multicelulares complejos. Es decir, de que la evolución consiste en progresar de lo simple como una bacteria a lo complejo como un ser humano. Algunos consideran a la especie humana como la cúspide de la evolución y muchas veces esto está tan arraigado dentro de nosotros que lo reflejamos en nuestra religión y cultura.

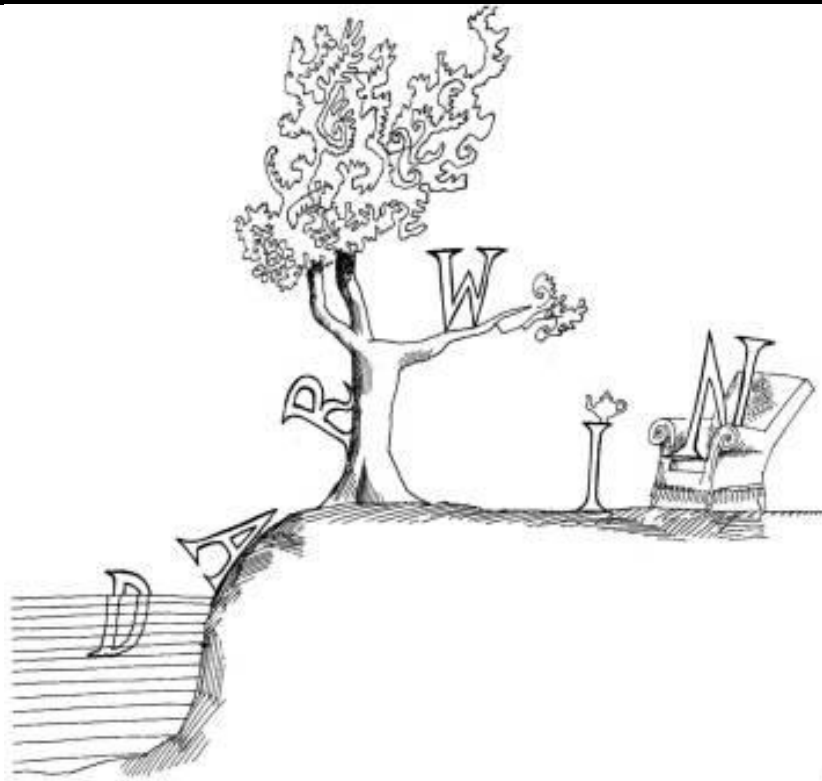


Figura 2-30. Caricatura de la evolución progresiva de Darwin. La figura muestra como la vida surgió del mar, para después poblar la tierra, subirse a los árboles, para después bajar de las ramas, y finalmente, pasar a sentarse en un sillón para tomar el té y reflexionar sobre todo ello.

Sin embargo, otros piensan que esto es un punto de vista antropocéntrico, más que científico. La evidencia científica actual no es suficiente para sacar una conclusión definitiva. Sin embargo, la evolución natural no parece tener una dirección específica, ni un camino predeterminado. No sabemos si la evolución tiene una dirección, pero si es que la tiene, entonces no tiene un rumbo fijo predeterminado. Es decir, no es un rumbo definido para siempre, sino que el rumbo puede cambiar con el tiempo y según las circunstancias. La **historia evolutiva** está llena de **accidentes**, cataclismos y sorpresas. No se puede decir que exista una presión selectiva universal. Hay organismos que con el tiempo se han hecho más complejos o grandes, pero también hay muchos casos contrarios de organismos que se han hecho menos complejos o pequeños. Lo que si se puede constatar es que existe una fuerza continua de innovación que genera siempre nueva diversidad genética. Esta diversidad biológica tiende a ocupar todos los nichos ecológicos libres (Figura 2-31).

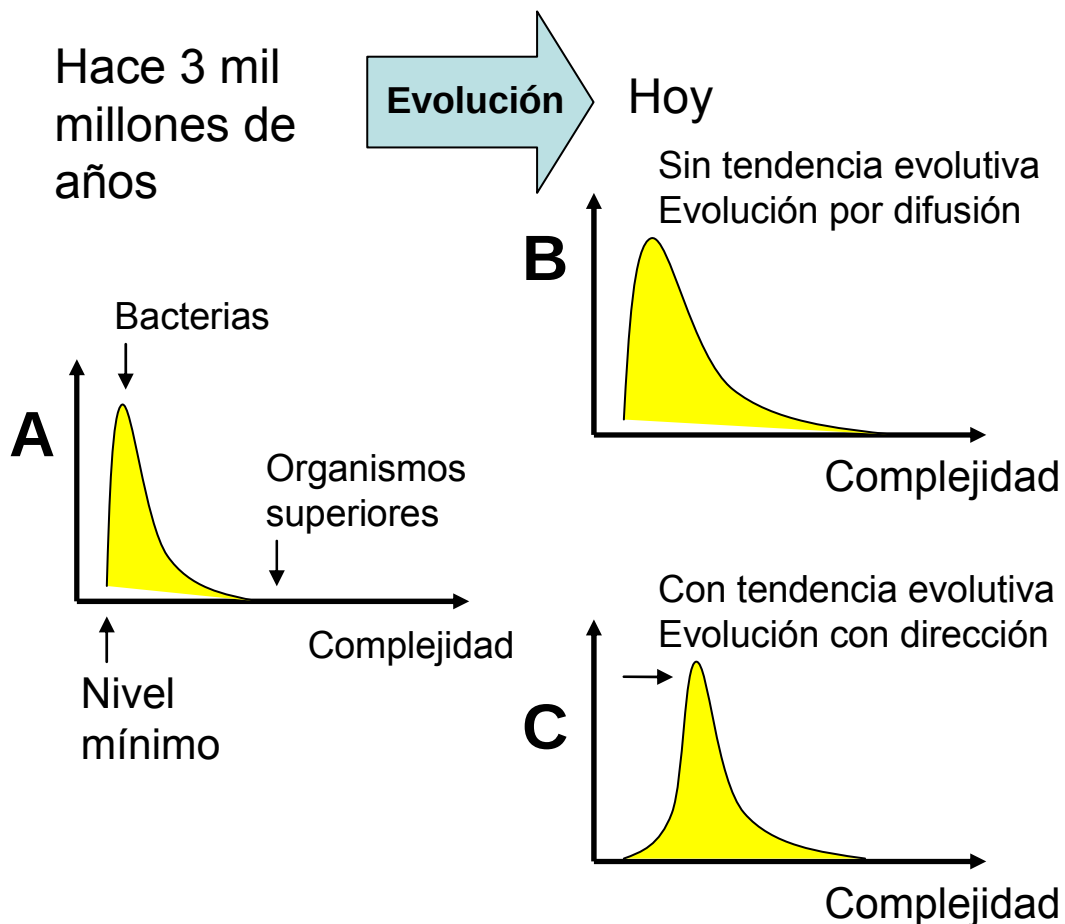


Figura 2-31. Curvas de distribución según diferentes tendencias evolutivas. El panel A muestra la situación original hace mas de mil millones de años. El panel B muestra un ejemplo de evolución sin una tendencia fija, es decir, evolución por difusión. En este caso el promedio no cambia. El panel C muestra un ejemplo de cómo se comportaría la distribución, si existiera una tendencia evolutiva que hiciera que todas las especies se vuelvan cada vez más complejas. En este caso, el promedio si cambia. Idea basada en los argumentos de Stephen Jay Gould (*Full House: The spread of excellence from Plato to Darwin*).

Many people wrongly believe that the process of evolution has a preferred direction—a tendency to make organisms more complex and more sophisticated as time goes by. Those who believe in evolution's drive towards progress often demonstrate it with a series of organisms that appeared in different eons, with increasing complexity, e.g., "bacteria, fern, dinosaurs, dog, man". Gould explains how these increasingly complex organisms are just one hand of the complexity distribution, and why looking only at them misses the entire picture—the "full house". He explains that by any measure, the most common organisms have always been, and still are, the bacteria. The complexity distribution is bounded at one side (a living organism cannot be much simpler than bacteria), so an unbiased random walk by evolution, sometimes going in the complexity direction and sometimes going towards simplicity (without having an intrinsic preference to either), will create a distribution with a small, but longer and longer tail at the high complexity end.

Tarea 2-38

Trate de recabar información biológica que le permita distinguir si la curva de distribución de las especies se comporta más como el recuadro de arriba o el de abajo en la Figura 2-31. Es decir, trate de responder por si mismo si existe alguna tendencia evolutiva universal que aplique para algunos grupos de seres vivos.

Si existiera una tendencia evolutiva universal, entonces uno debería ver que todos los organismos de ese grupo se han desarrollado en esa dirección, y es por eso que el promedio también debería de cambiar con el tiempo. Es decir, la evolución con dirección se debe observar por medio de la comparación de los promedios y no de los extremos. Mediante la evolución “por difusión” el promedio general no cambia. Simplemente es que la curva de distribución se hace más amplia. La evolución por difusión se debe observar a través de la comparación de los límites de los extremos. Si antes no había organismos tan complejos, con el tiempo han surgido algunos cuantos por el simple hecho de que están tratando de ocupar nuevos nichos ecológicos que antes estaban libres. La curva de distribución se amplía hacia la derecha, mientras que no puede ensancharse para la izquierda porque existe un mínimo de complejidad más abajo del cual un organismo biológico no puede reproducirse (barrera de mínima complejidad). Esta barrera mínima posiblemente esté representada por algunas bacterias o micoplasmas, que son las células más simples que se conocen.

Un nicho ecológico que todavía está libre, es el de un organismo tolerante que sepa desarrollarse de manera sustentable y no se autodestruya. La especie humana puede acabar extinguiéndose como los dinosaurios o transformarse para seguir ocupando un nicho ecológico privilegiado, en nuestro limitado planeta.

2.6. Filosofía científica del mejoramiento genético

2.6.1. Mejoramiento: Evolución en acción

La meta de un mejorador es obtener plantas o cultivos diferentes. El mejorador quiere lograr un cambio, de preferencia, benéfico. Por lo regular se trata de obtener una mayor producción en términos de cantidad y calidad. La evolución biológica es un cambio a lo largo de millones de años, sin embargo, es un cambio que se da al azar y no tiene dirección. A diferencia, el mejorador quiere lograr un cambio independientemente del azar, y además, sí quiere darle una dirección específica, y por si fuera poco, no quiere esperar miles de años para lograrlo. Es por eso que podemos decir, que el mejoramiento genético es un proceso de **evolución dirigida**, en donde la eficiencia y la rapidez son parámetros importantes.

Es interesante recalcar que el mismo Darwin cuando exhibió su famoso libro sobre el "Origen de las Especies", basó gran parte de sus argumentos en la

selección artificial de los mejoradores genéticos de la época. En sus libros dedicó varios capítulos a la selección artificial y la domesticación de animales y plantas por la mano del hombre.

Hay que recordar que la idea novedosa que postuló Darwin, es que también existía un proceso de selección en la naturaleza, y que ese proceso -al que llamó selección natural- era responsable de que los organismos silvestres pudieran cambiar tan drásticamente a lo largo del tiempo.

El mejoramiento genético es un **proceso iterativo**, con ciclos repetitivos de mutación, recombinación y selección. Es por eso que en realidad el mejoramiento no es otra cosa más que **evolución en acción**. La evidencia histórica indica que se ha logrado un gran avance, ya que a partir de la **revolución verde** en los años 50, la producción de trigo y de maíz en países desarrollados se ha incrementado a un ritmo promedio de alrededor del 10% por década (ver Figura 2-32).

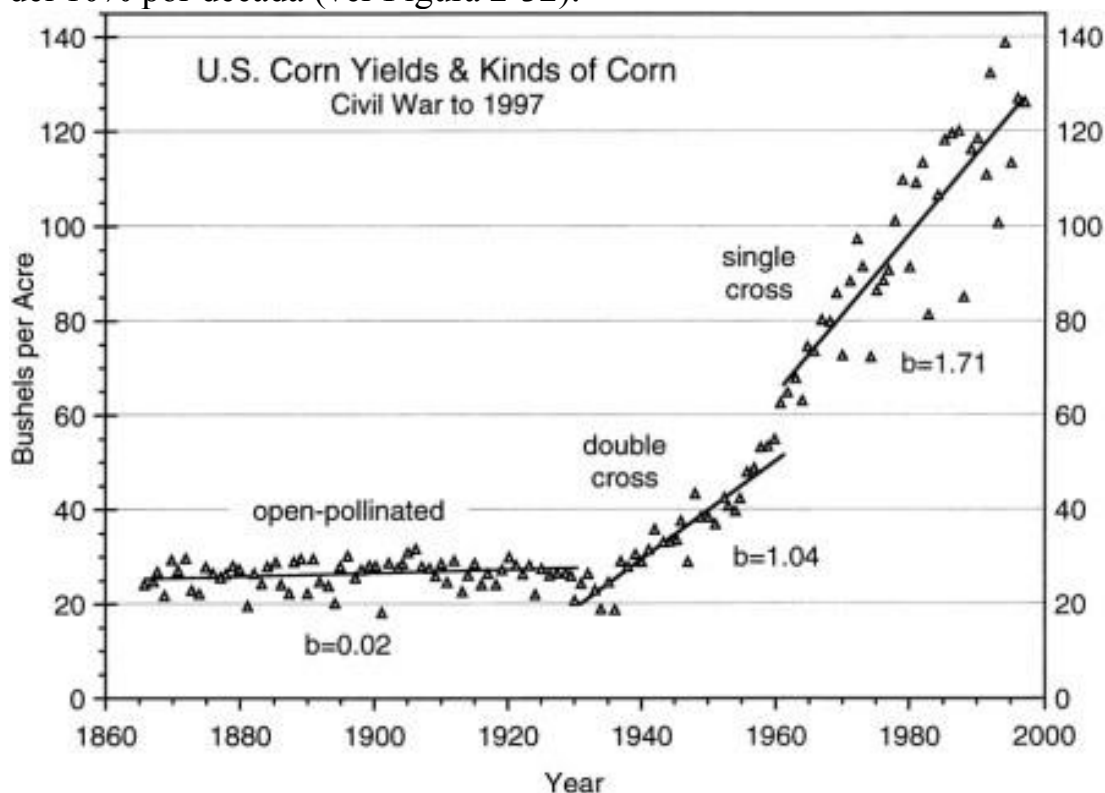


Figura 2-32. Rendimiento de maíz en Estados Unidos. Se indican los periodos dominados por variedades de polinización libre (OPVs), híbridos dobles e híbridos simples junto con los coeficientes de regresión. Crow, J. F. (1998) Genetics: 148 p923-928. Existen diferentes tipos de bushels en Estados Unidos (los hay de 60, 56 y de 48 libras), pero posiblemente en esta figura se refieren a los bushels de 60 libras. 1 Bushel por Acre equivale aproximadamente a 0.067 toneladas por hectárea.

En 2008 el rendimiento promedio de maíz en México fue de 2.4 ton/ha lo que equivale a 36 bushels per acre. Estos bajos rendimientos se deben a las pobres prácticas agronómicas y también a que en México actualmente predominan los materiales criollos más que los materiales híbridos. Para

incrementar el rendimiento en México, es necesario hacer más trabajos científicos y desarrollos tecnológicos.

Sin embargo es importante resaltar, que cuando una tecnología se importa, sin adecuarla a las condiciones del lugar, puede traer graves consecuencias desfavorables, como ocurrió con la revolución verde en la mayoría de los países de África. Es por ello que la transferencia de tecnología se debe hacer con criterios científicos, ecológicos, económicos y sociales.

En términos globales, se puede decir que el mejoramiento ha sido todo un éxito por el incremento en la productividad. Durante los últimos 50 años ha habido un aumento continuo y constante de los rendimientos por hectárea. Los datos los podemos consultar en las bases de datos de la FAO (ver siguientes figuras). El impacto que esto ha generado en la alimentación mundial es innegable.

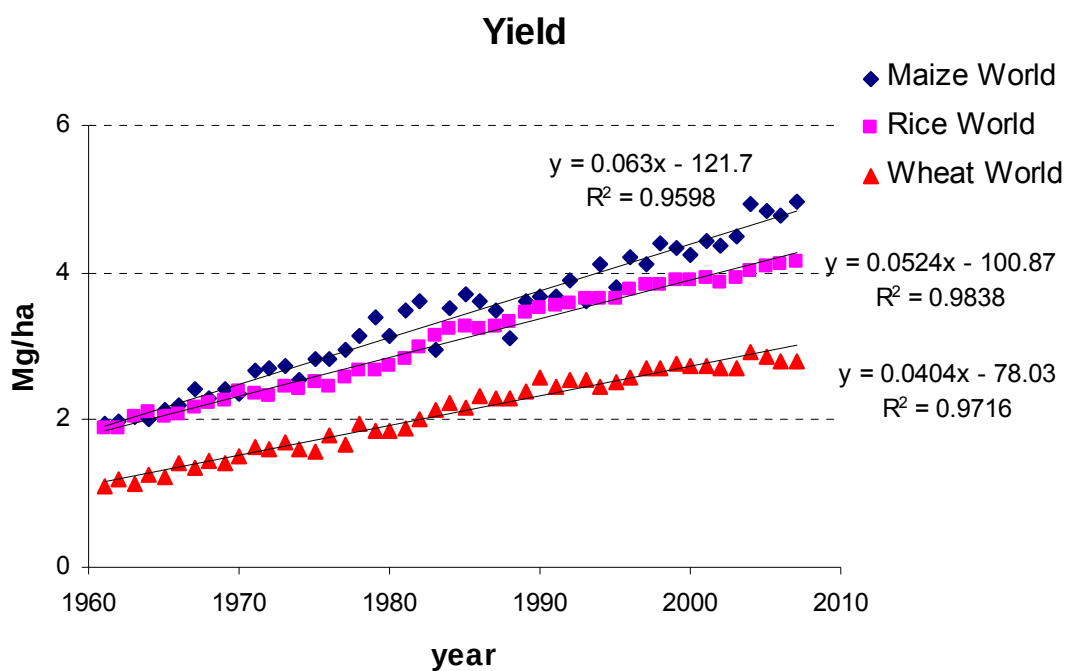
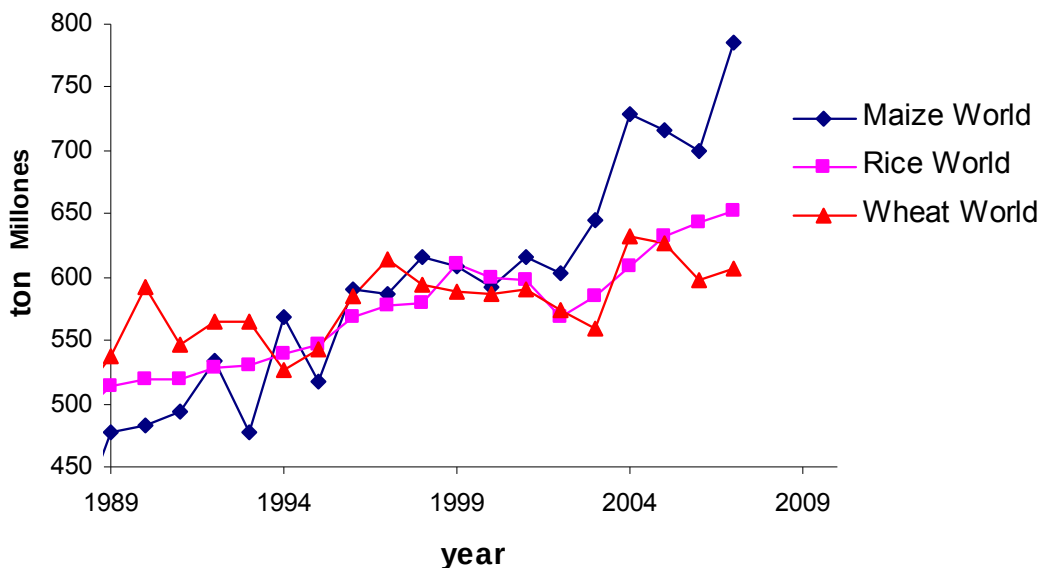


Figura 2-33. Rendimiento promedio global de los 3 cereales más importantes del mundo: maíz, arroz y trigo. Se puede ver que la productividad ha aumentado de forma lineal desde hace más de 40 años. Fuente de datos: Faostat 2009.

Este logro se debe al trabajo combinado de los mejoradores y los agrónomos. La curva de regresión muestra que los incrementos para cada cereal han sido diferentes. El maíz tiene la tasa más alta con una ganancia promedio de 63 kg/ha*año, seguido por el arroz con un incremento de 52 kg/ha*año y el trigo con 40 kg/ha*año. El mayor rendimiento de grano se

obtiene con maíz (>4 ton/ha), debido a su fotosíntesis más eficiente. El maíz es C4, mientras que el arroz y el trigo tienen fotosíntesis tipo C3.

World Cereal Production



FAOSTAT 2009

Figura 2-34. Producción global de los 3 cereales más importantes del mundo: maíz, arroz y trigo. Se puede ver que en 1989 el cereal más producido en el planeta era el trigo, seguido por el arroz y por último el maíz. A partir del año 2000 hubo una inversión en la jerarquía. Hoy en día (2003-2009), el cereal más importante del mundo es el maíz, seguido por el arroz y en tercer lugar por el trigo.

El hecho de que el maíz haya rebasado a los 2 otros cereales en los últimos 20 años se debe al éxito en el mejoramiento genético del maíz, principalmente por el aprovechamiento del fenómeno de la heterosis en esta especie. Sin embargo, los incrementos en la productividad han sido variables en diferentes países. En algunos países, el rendimiento ha incrementado más rápido que en otros. Los valores los podemos consultar en la base de datos de la FAO (ver siguientes figuras).

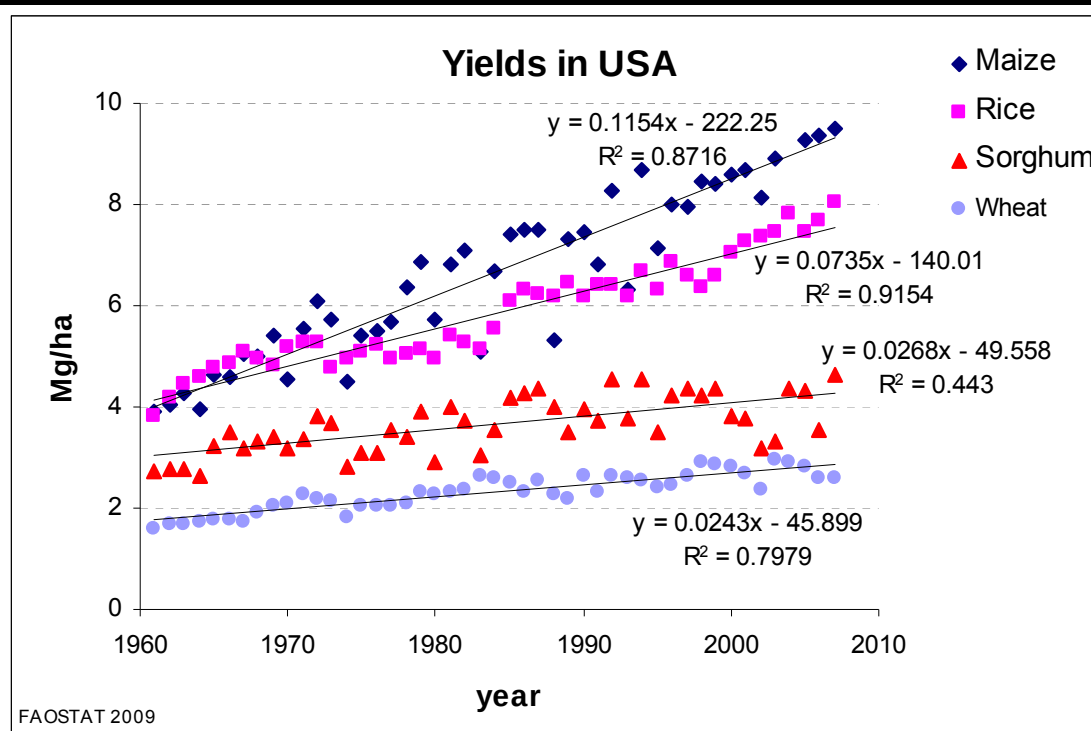


Figura 2-35. Rendimiento de cereales en Estados Unidos. La pendiente indica la tasa de incremento en los rendimientos por hectárea. Se puede ver que en Estados Unidos, la pendiente para maíz es mucho mas pronunciada que la de trigo.

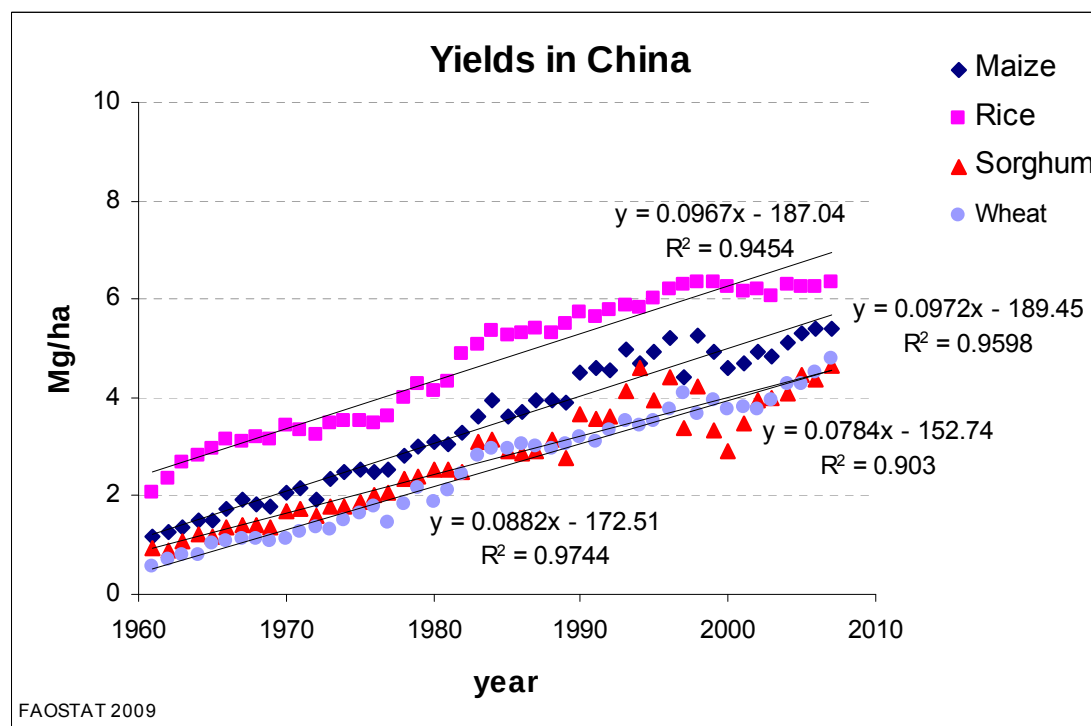


Figura 2-36. Rendimiento de cereales en China. Se puede ver que en China, la pendiente es muy similar para todos los cereales, que refleja el enorme esfuerzo de mejoramiento genético en todos ellos.

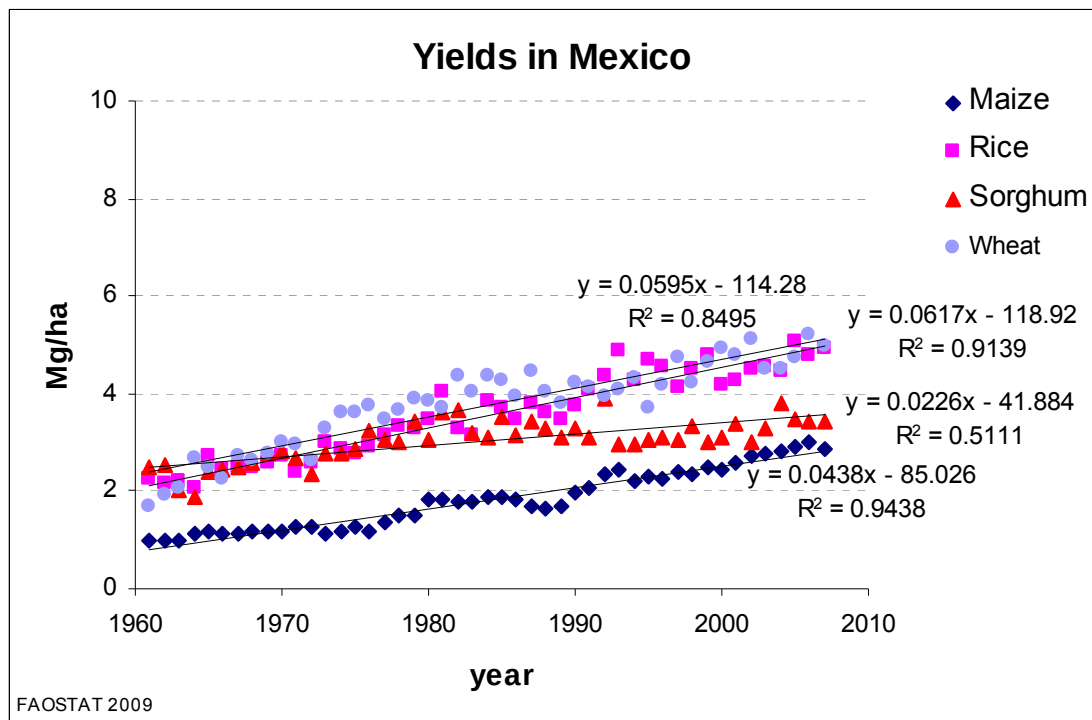


Figura 2-37. Rendimiento de cereales en México. Se puede ver que en México los rendimientos y las tasas de incremento son mas bajas que en los otros países, sobre todo, para maíz y sorgo. A diferencia de Estados Unidos y China, en México el maíz no es el primer cereal en rendimiento, sino el último.

Por lo regular el maíz es de los cereales más productivos del mundo. Sin embargo, en México se da un fenómeno muy particular. México es relativamente competitivo en la producción de trigo, sorgo y arroz. Por ejemplo, los rendimientos por hectárea de trigo que se obtienen en México (>4 t/ha), están por arriba de lo que se obtiene en promedio en Estados Unidos (<3 t/ha). En sorgo y arroz, México tampoco está tan mal.

En el caso de maíz, México si está muy por debajo de muchos países del mundo. En Estados Unidos se cosechan en promedio 9 t/ha, en China se obtienen 5 ton/ha, mientras que en México se producen menos de 3 t/ha. ¿Por qué? Para entenderlo mejor, debemos verlo desde una perspectiva evolutiva, considerando algunas tendencias históricas. A continuación se muestra una tabla que indica las tasas de incremento de los rendimientos.

kg/ha*año	México	Estados Unidos	China
Maíz	43.8	115.4	97.2
Arroz	61.7	73.5	94.5
Sorgo	22.6	26.8	78.4
Trigo	59.5	24.3	88.2

Tabla 2-1. Incremento por año de los rendimientos en kilogramos por hectárea. Los datos corresponden a la pendiente de las líneas de regresión de las figuras anteriores. El valor corresponde a la pendiente promedio (en kg/ha*año) durante los últimos 40 años.

En Estados Unidos se le ha dado mucho énfasis al mejoramiento de maíz. Durante las últimas 4 décadas se ha logrado un incremento promedio de 115 kg por año. China es líder en arroz, pero también ha logrado incrementar los rendimientos de todos los cereales a un ritmo acelerado. Sin embargo, en México, no sólo el rendimiento promedio es bajo (< 3 t/ha), sino que además el incremento en la productividad del maíz por año es mas bajo que el promedio mundial. El progreso que se obtiene en México (43.8 kg/ha*año) es menos de la mitad de lo que se obtiene en China ¿Por qué? ¿Cuál es la causa de que en México se de una velocidad de incremento menor para el maíz, a pesar de que es el país con el mayor recurso genético para esta especie? Le podemos echar la culpa a los políticos, a los mejoradores o a los agricultores. Pero tal vez haya razones aun más enigmáticas que los propios diputados.

2.6.2. Balance entre creatividad y eficiencia

Algunos científicos han observado que los centros de origen biológicos están positivamente correlacionados con la diversidad de alelos, genes y variedades. Sin embargo, también han notado que esos centros están negativamente correlacionados con la productividad.

Los rendimientos más altos se obtienen fuera de los centros de origen. Por ejemplo, la máxima productividad del maíz se da en otros países fuera de México. Una baja productividad de maíz se obtiene en el centro y sur de México, donde hay una gran diversidad de razas y variedades de maíz. En trigo, la máxima productividad se da fuera de Medio Oriente. En Irak e Irán – el centro de origen– los rendimientos de trigo son relativamente bajos comparados con el promedio mundial. En papa, la máxima productividad se da fuera de Perú. Se cosechan más tubérculos por hectárea en los valles de los Alpes (Europa) que en los valles de los Andes (América), de donde se originó la papa y donde todavía hay mucha diversidad de solanáceas. Y como esos, hay muchos otros ejemplos que sugieren que el origen de la diversidad, y la máxima productividad, son como polos opuestos. ¿Podría ser un fenómeno universal? Tal vez sea solo una coincidencia que en la especie humana, que

tiene sus raíces en el continente africano, el máximo desarrollo social, cultural, científico y económico también se ha dado fuera del centro de origen.

Se podría postular la hipótesis de que la diversidad genética es como un lastre que impide que se alcancen los más altos niveles de rendimiento. Es como si la diversidad y la productividad fueran fenómenos opuestos. ¿A qué se puede deber esto? Por un lado, las mutaciones por el azar generan diversidad, y por el otro, la selección limita las variantes que sobreviven. Es un equilibrio dinámico que siempre está en transición y movimiento. Sabemos que los procesos biológicos no son fenómenos estáticos, ya que la evolución se da por un balance de fuerzas opuestas.

El mejoramiento consiste en la selección de los mejores alelos, genotipos y variedades. Para ello se deben descartar variedades que no son tan productivas bajo las mismas condiciones. Para obtener el máximo rendimiento sólo se debe sembrar la variedad que es más rendidora. La productividad se “compra” a costa de la pérdida de un cierto porcentaje de diversidad. La disminución de la diversidad (por la selección) sólo se puede compensar por medio de mas cruza y recombinaciones que generan nuevas variantes de alelos. Sin embargo, un mejorador que no está dispuesto a descartar los genotipos menos favorables, nunca logrará obtener los más altos rendimientos. Es el precio que se debe pagar para hacer mejoramiento.

Los centros de origen se pueden considerar como **incubadoras de diversidad genética**. En esas regiones, la presión selectiva debe de ser relativamente baja, para que la tasa de mutación y recombinación puedan ser altas, lo que permite un alto grado de innovación y creatividad genética. Una vez incubados en los centros de diversidad, los genotipos pueden ser exportados a otras regiones del mundo donde son expuestos a condiciones selectivas más exigentes. Sólo los genotipos más rendidores tendrán éxito en un ambiente competitivo. En esos ambientes exigentes habrá una baja diversidad (muy pocos genotipos exitosos), pero con muy alta productividad.

El origen de los rendimientos tan altos en China y Estados Unidos se fundamenta en cierta medida en aquellos genes que alguna vez fueron preincubados en el centro de diversidad del maíz, en México. Cuando los españoles conquistaron y colonizaron América, nunca hubieran imaginado que las pocas semillas de maíz que se llevaron en sus barcos (el potencial genético de las semillas), iba a ser mucho más valioso que todos los lingotes de oro y plata que se llevaron a Europa. El valor comercial del maíz que se produce fuera de México es tan cuantioso que alcanzaría a financiar el sistema educativo y científico de todo un país.

El trabajo de incubación genética por los indígenas mesoamericanos durante cientos de años dio lugar a un sinnúmero de razas y variedades de maíz, que hoy en día representan un acervo biológico para toda la

humanidad. Esa diversidad genética es la razón por la que el maíz se ha convertido en el cereal más importante de todo el planeta (ver Figura 2-34).

De ahí radica la importancia y la perdurabilidad del mejoramiento genético:

*He who can make one ear of grain grow where none grew before
renders an endureable service to whole humankind*

Las anteriores evidencias evolutivas nos inspiran a hacer una reflexión más extensa en torno a la diversidad y la productividad. Para ello usaremos algunas analogías. Una región de origen biológico puede ser considerada como un **centro de incubación genética**. Es como un vivero. Se debe fomentar la creatividad y la imaginación para permitir así la generación de nuevas variantes y alelos.

Los centros de incubación son como las escuelas a las que enviamos a nuestros hijos para que aprendan y puedan ser creativos. Debemos permitir que descubran y que tengan sus propias experiencias. Debe ser un ambiente óptimo de baja presión. Debemos ser tolerantes con los errores de los jóvenes, ya que solo a través de la experimentación aprenderán a desarrollar sus habilidades. Una vez que fueron incubados, los jóvenes deberán salir al mundo exterior para demostrar sus aptitudes. Así se pondrán a prueba. Una persona que nunca sale al exterior, difícilmente alcanzará su máxima productividad. Por otro lado, un joven que nunca se preincubó en ese ambiente creativo y libre de presiones, nunca logrará lo mismo que una persona que haya disfrutado de ese privilegio.

El fomento a la diversidad si tiene un costo para los demás. ¿Quién lo paga? En las familias, los padres son los que trabajan para pagar la formación académica de sus hijos. En las empresas, el departamento de ventas es el que financia las operaciones del departamento de innovación y desarrollo. Digamos que los vendedores con su función generadora de ingreso tienen que financiar a los demás trabajadores de la empresa. En los gobiernos, las agencias federales y estatales dan becas para que los jóvenes estudien. ¿Los científicos somos un lastre para la sociedad? ¿Todos deben dedicarse a vender, o también a estudiar y a inventar? Más que un costo, la investigación es una inversión a largo plazo. Sin innovación no habrá nuevos productos, y sin ellos, tarde o temprano no habrá más ventas. ¿Quién recibirá los beneficios?

La historia de la evolución nos enseña el valor que tiene la diversidad. Si no hubiera mecanismos para la generación continua de diversidad genética, ya nos hubiéramos extinguido hace mucho. La innovación y el desarrollo tecnológico deben ser un esfuerzo de toda la sociedad. Los empresarios que ganan más deben apoyar (financiar con sus impuestos) a los investigadores que quieren hacer ciencia de alto nivel. ¿Por qué? El costo de la ignorancia es trivial comparado con el gasto en educación e investigación. La diversidad del conocimiento es la única forma de reducir el peligro de fracasar a largo plazo.

Lo que debemos preguntarnos en México es: ¿Qué valor le damos a la diversidad? ¿Qué valor tiene la educación? ¿Qué tan creativos queremos ser? ¿Cómo podemos incrementar la productividad agrícola para ser autosuficientes en la producción de maíz, con el mínimo costo posible de diversidad genética? ¿Quién debe pagar las investigaciones que se requieren en maíz? ¿Cómo podemos evolucionar para progresar? ¿Qué es lo que se requiere hacer?

Lo que más necesita México, y Latinoamérica, son mejores políticos y mejores estudiantes. Este libro sobre **Fundamentos de Mejoramiento Genético** sólo pretende remediar una de muchas necesidades. Un buen gobierno es aquel que tiene la suficiente visión a largo plazo para estar convencido de que la ciencia y la investigación no es un gasto sino una **inversión**. Un buen estudiante es aquel que le dedica suficiente tiempo y esmero al aprendizaje, porque sabe que el conocimiento es una buena inversión para su futuro. Es una inversión altamente rentable. Un buen libro es aquel que transmite conocimientos, pero también aquel que permite incrementar la creatividad de los jóvenes para fomentar la diversidad de sus propios pensamientos.

Por un lado, el razonamiento lógico debe ayudarles a analizar críticamente los datos. Deben descartar muchas ideas que no fueron confirmadas por la evidencia científica. La ciencia avanza por la falsificación de hipótesis. Esto significa que hay que descartar ciertos pensamientos. Esto disminuirá la diversidad. Pero por otro lado, la creatividad y la imaginación nos ayuda a unir pensamientos aislados, para generar conclusiones y desarrollar soluciones novedosas que permitan incrementar la eficiencia de nuestra tecnología y economía. La combinación de dos ideas seleccionadas puede dar lugar a decenas de otras ideas recombinadas. El mejoramiento genético, además de ser evolución en acción, también es un conjunto de diversas disciplinas que requieren el desarrollo de habilidades de muy diversa índole. Los mejoradores no sólo deben seleccionar y recombinar genes, sino también deben recombinar y seleccionar ideas y teorías, al mismo tiempo que deben aplicar y desarrollar nuevas metodologías y estrategias.

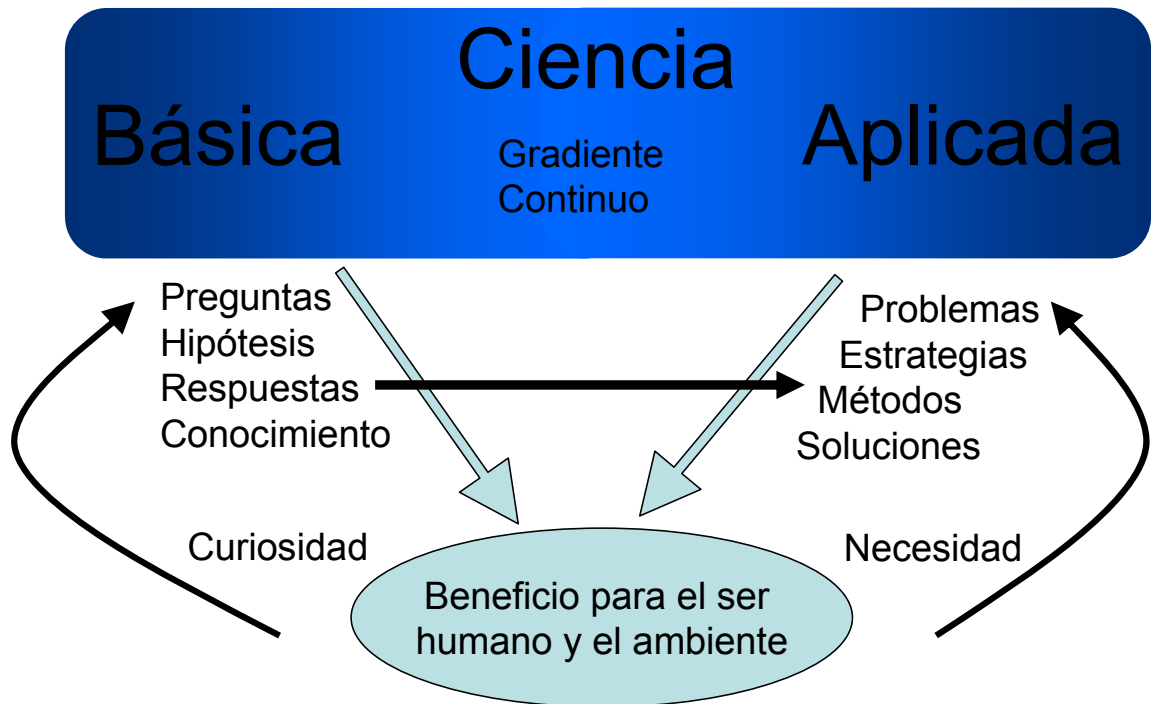
2.6.3. Ciencia básica y aplicada

La ciencia es un gradiente continuo ya que es difícil demarcar un límite entre un área y otra. Sin embargo, se puede definir enfoques para iniciar los experimentos de un proyecto de posgrado:

- Ciencia básica: contestar preguntas (tesis investigación)
- Ciencia aplicada: resolver problemas (tesis desarrollo tecnológico)

Es decir, para uno se plantean hipótesis y se obtienen respuestas, mientras que para el otro se identifican problemas y se generan soluciones (ver

figura abajo). Muchas veces, se necesita contestar muchas preguntas antes de resolver un solo problema. Es por ello que la ciencia básica es necesaria y fundamental para sustentar la ciencia aplicada. Se pueden hacer tesis de ciencia básica sin requerir de una aplicación, pero es imposible concebir tesis de ciencia aplicada sin los mínimos fundamentos de ciencia básica.



Posgrado



2.6.4. Combinación de ciencia, ingeniería y filosofía

Un mejorador es como un artista. Sin embargo, no pinta sobre papel ni esculpe piedras. Su materia prima es mucho más compleja y sutil: es la propia molécula del DNA. Los mejoradores moldean el material genético de una especie. Son los artistas del DNA, la molécula de la herencia. Son creadores de nueva biodiversidad, que que generan nuevas variedades que antes no existían. Para ser exitoso, el mejoramiento debe combinar conocimientos de diversas disciplinas. Se debe dominar la ciencia, la ingeniería y la filosofía. ¿Como delimitar las diferencias entre ellas?

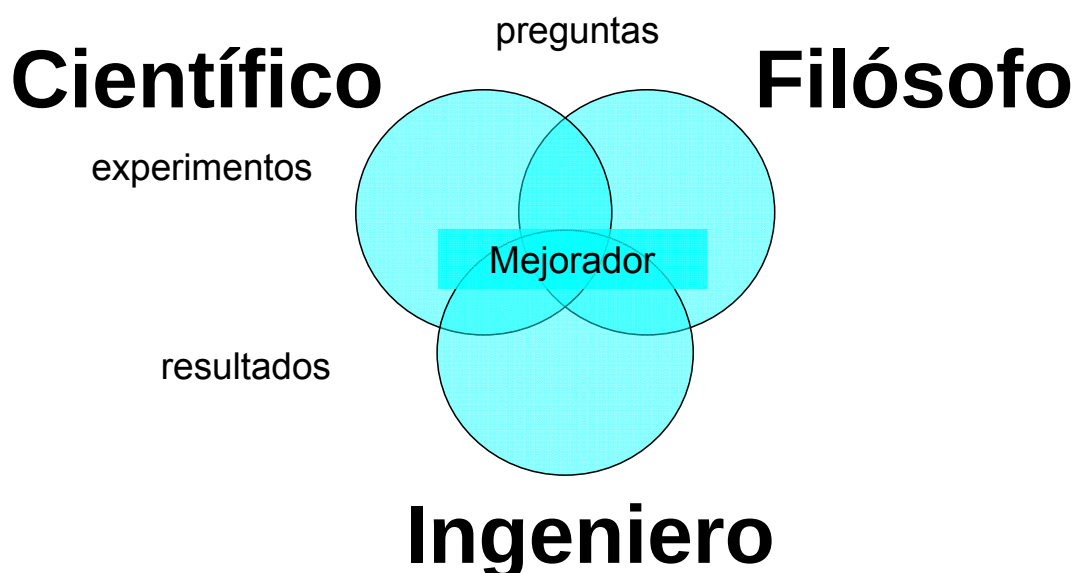


Figura 2-38. Áreas de interés intelectual para los fitomejoradores. El mejorador tiene que ser científico, ingeniero y filósofo al mismo tiempo.

La ciencia se distingue de la filosofía porque la primera esta basada en los **experimentos**. Un filósofo reflexiona, analiza las premisas y usa la lógica. Sin embargo, no hace experimentos. El científico usa su intelecto y sus sentidos para observar el mundo que lo rodea, aprovecha su creatividad para realizar experimentos y utiliza herramientas y equipos para recabar datos. Después analiza los datos (herramientas bioinformáticas) y usa su razón y conocimientos previos para llegar sus conclusiones. Un científico realiza experimentos, obtiene datos cuantitativos y trata de explicarlos por medio de teorías y leyes. Es muy difícil refutar un punto de vista filosófico, ya que existen muchas y muy diversas opiniones, y cada una de ellas puede tener su propia validez en un determinado contexto. La ciencia no es una cuestión de opinión. Es una cuestión de hechos. Una teoría científica puede ser falsificada con los datos de un experimento. Si la teoría científica no puede explicar los resultados, o peor aun, esta en contradicción con los resultados, entonces se dice que ese teoría ha sido falsificada, y por consiguiente se descarta. Se tiene que encontrar otra teoría alternativa que sea congruente con los resultados experimentales. La ciencia avanza poco a poco por medio de la falsificación. Es un método iterativo para el mejoramiento de las ideas y teorías científicas.

La técnica se distingue de la ciencia ya que no esta interesada en las preguntas de fondo, sino más bien enfocada a la solución de problemas prácticos. Un técnico agrícola o un ingeniero agrónomo, no pierde el tiempo en hacerse demasiadas preguntas sobre el porque de los fenómenos naturales. No se interesa en los procesos moleculares. La ingeniería civil no se rompe la cabeza con las leyes teóricas de la cuántica gravitacional, sino más bien quiere construir un puente que resista la carga de un tractor. La ingeniería pone sus prioridades en el costo, la rapidez y la eficiencia. El

mejoramiento tradicional no necesita tener conocimiento de las 4 letras químicas del DNA que determinan la secuencia de un gen. Les basta saber que hay características fenotípicas que se heredan de padres a hijos, y que por medio de la selección se puede enriquecer la frecuencia de aquellas variantes que más le convengan. Es por ello que solemos cuestionarnos: ¿Para qué me sirve saber todo eso? Tal vez lo captemos hasta mucho mas tarde.

La ciencia pone el énfasis en la veracidad, la repetibilidad y la precisión de los datos. Debe de haber congruencia con los paradigmas de la lógica y la razón. Para que un experimento sea científicamente válido y pueda ser reconocido como tal (por ejemplo en una publicación en una revista internacional), debe tener **repeticiones** y **controles**. Un resultado único no tiene validez científica si no puede ser reproducido de forma independiente. Es decir, es como un milagro que sucede una sola vez y que nunca más se podrá repetir de igual forma. La ciencia se basa en los experimentos, y estos solo tienen validez si son repetibles.

La ciencia no se interesa en lo que no se puede refutar ni tampoco se puede comprobar. Para lograr sus objetivos se ha definido un método científico, que establece el protocolo a seguir (hipótesis -> experimento -> resultados). Se reconocen ciertos mecanismos de crítica racional, análisis objetivo y escepticismo científico por medio de revisión externa e independiente de otros colegas investigadores, que en su conjunto se llama "comunidad científica". En la ciencia no existe ningún dogma que no pueda ser cuestionado. La ciencia no da una respuesta final, estática y definitiva. Todo podría ser interpretado de una forma muy diferente. Siempre es una aproximación, que puede ser perfectible. La ciencia nunca debe cerrar su mente ni aferrarse a una sola premisa. Lo que tiene mas validez son los resultados de los experimentos, no las interpretaciones, ni las explicaciones.

Tarea 2-39

Escriba un ensayo contestando alguna de estas preguntas filosóficas: ¿Cuál es la diferencia entre los dogmas de la ciencia y de la religión? ¿Cuál es la diferencia entre una hipótesis, una teoria y una ley?

Es imprescindible que el fitomejorador moderno sepa escoger las herramientas para incrementar la diversidad. Debe tener la capacidad de medir los parámetros que usará para seleccionar. Debe ser eficiente para hacer mucho con pocos recursos. Un buen fitomejorador tiene una mano prodigiosa para dirigir la evolución, en la dirección adecuada y por el

camino más viable. Ese es un talento que, muy pocos tienen de manera innata, pero que muchos pueden adquirir por medio del aprendizaje y la dedicación. El mejoramiento moderno es tanto un arte como una ciencia, es ingeniería y práctica, pero también incluye teoría y filosofía (Figura 2-38).

2.7. Los ácidos nucleicos

2.7.1. Nociones básicas del DNA

2.7.1.1. Analogía de la biblioteca

Empezaremos con una analogía para explicar el concepto de **información genética** y la relación que existe entre cromosomas, genes y proteínas. Podemos imaginar la forma en que trabajamos en la cocina. Analicemos el flujo de información y los procesos que utilizamos para preparar un pastel. Ahora comparémoslo con los procesos que usa la célula para obtener una proteína. La Figura 2-39 muestra este ejercicio al que llamaremos: la **analogía de la biblioteca y el pastel**:

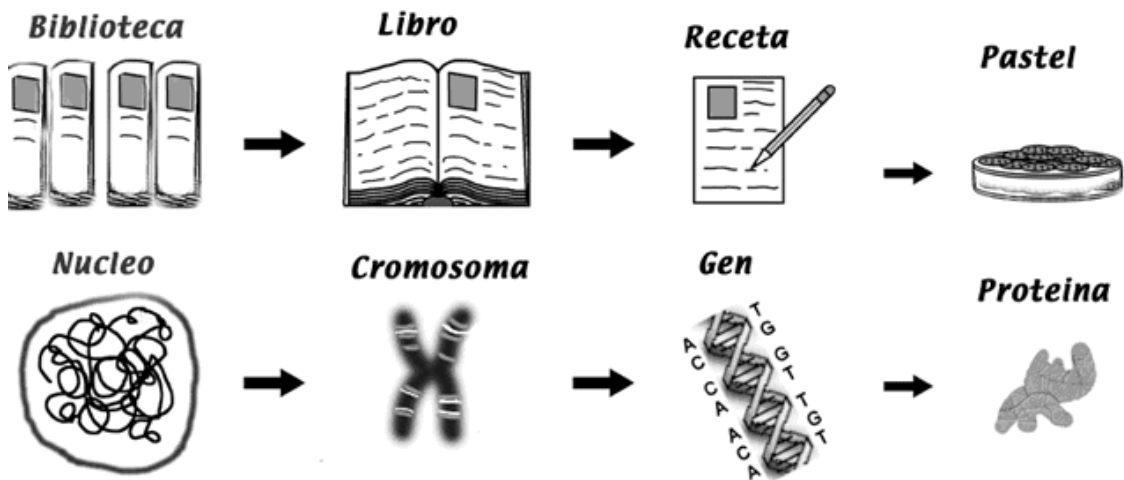


Figura 2-39. Analogía de la biblioteca y el pastel. Se muestra la correspondencia de la información que está en un núcleo, cromosoma, gen o proteína.

Esta forma de imaginarlo, es un excelente inicio para adentrarse en el mundo de los ácidos nucleicos y así, entender mejor las bases del mejoramiento genético.

Tarea 2-40

Explique con esta analogía lo que sucedería si usted arrancara una página de la biblioteca de un restaurante chino y se la agregara a un libro de un taller mecánico alemán. ¿Qué tan importante es el hecho de que los

mecánicos y los cocineros hablen el mismo idioma? Si el texto estuviera en chino y usted solo hablara alemán, ¿de qué sabor le saldría el pastel? Escriba un cuento sobre ello. Muéstrole el cuento a un familiar y después discútalo con sus demás compañeros de clases. ¿Qué tanto sirvió su cuento para que una persona común entienda mejor la ingeniería genética?

Tarea 2-41

Explique la relevancia científica y tecnológica de que el código genético sea universal. ¿Podría haber ingeniería genética si el lenguaje genético fuera diferente para cada especie? Postule una hipótesis de por qué el lenguaje humano ha cambiado tanto en pocos miles de años, mientras que el lenguaje genético no ha cambiado nada en miles de millones de años.

2.7.1.2. Nucleótidos

Existen cuatro nucleótidos básicos del **Ácido Desoxirribonucleico (ADN, DNA)**: la adenina (A), timina (T), citosina (C) y guanina (G). La diferencia entre los **ribonucleótidos** y los **desoxirribonucleótidos** es que la pentosa puede tener o no un grupo hidroxilo en el carbono 2. Una diferencia entre el DNA y el **Ácido RiboNucleico (ARN, RNA)** es que por lo regular el DNA casi siempre se presenta como una doble cadena, mientras que el RNA por lo regular es de cadena sencilla. Otra diferencia importante entre los nucleótidos que forman al DNA y el RNA es que la timina (T) se substituye por uracilo (U) (químicamente, la timina es un uracilo metilado en el carbono 5). A la base nitrogenada con la pentosa se le llama nucleósido, y dependiendo del número de grupos fosfatos esterificados en el carbono 5 del azúcar se le puede llamar mono-, di- o tri- fosfatos. La regla de apareamiento por cuestiones de tamaño es que las **purinas** (la A y la G con dos anillos cíclicos) se pueden aparear con las **pirimidinas** (la C, T o U con un solo anillo) (ver Figura 2-40.).

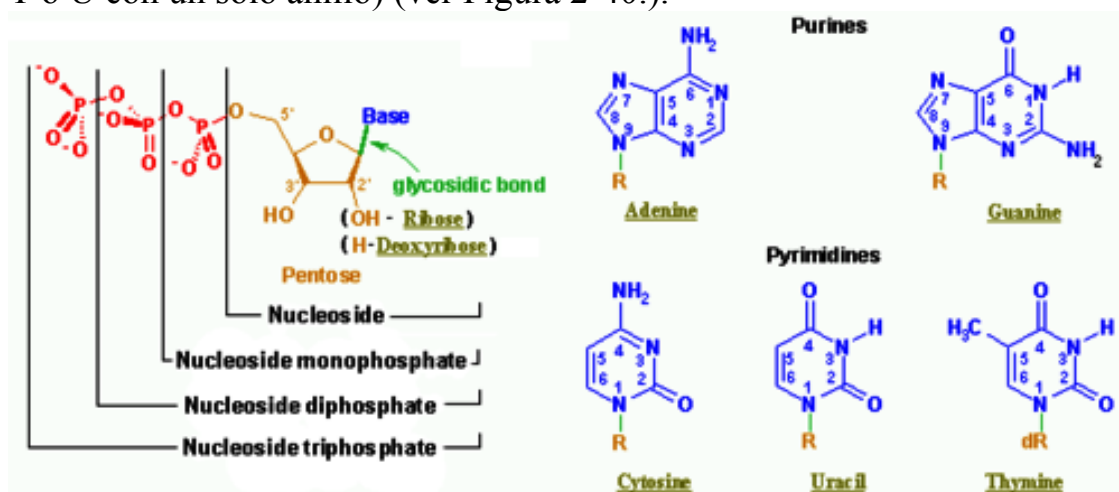


Figura 2-40. Nucleótidos

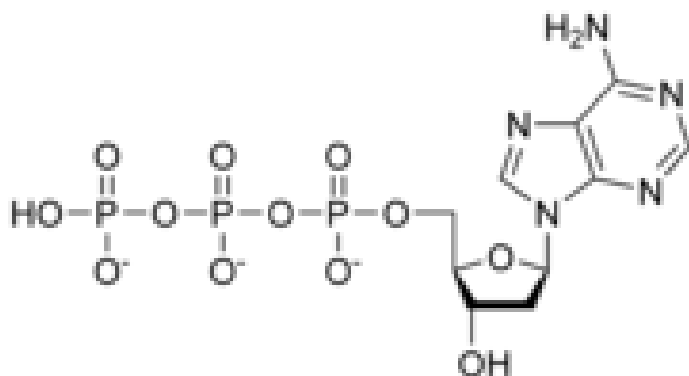


Figura 2-41. Estructura de dATP

Tarea 2-42

En la figura anterior se muestra una molécula de dATP, dibuje ahora una molécula de ATP. Use también las fórmulas que se muestran en esa figura para dibujar una molécula completa de GTP. ¿Puede aprender a dibujar las cuatro bases de memoria? Explique las diferencias de reactividad química entre el ATP y el dATP. ¿Qué consecuencias tiene la funcionalidad hidroxilo para la estructura y estabilidad de las moléculas de RNA y de DNA? Dé ejemplos de los usos que tienen esos dos nucleótidos dentro de la célula.

Si no hubiera CTP, pero si hubiera dCTP, ¿Se podría replicar el DNA? ¿Se podría transcribir el RNA? Justifique sus respuestas.

2.7.1.3. Doble hélice

El DNA es una cadena lineal que tiene dos hebras enrolladas en forma de doble hélice. Nos podríamos preguntar: ¿cada una de las hebras tiene la misma información? En realidad no es exactamente la misma, sino información complementaria. La complementariedad está dada por la forma en que los cuatro nucleótidos se aparean, **la A con la T, y la G con la C**. Entre las bases se forman puentes de hidrógeno que estabilizan el apareamiento complementario. El par de AT está formado de dos puentes de hidrógeno, mientras que el par GC forma tres (Figura 2-42). El número de puentes de hidrógeno tiene repercusiones sobre la temperatura a la que la doble hebra se separa. Por ejemplo la temperatura de apareamiento (T_m) para la PCR (Reacción en Cadena de la Polimerasa, por sus siglas en inglés: *Polymerase Chain Reaction*) depende del porcentaje de GCs. Entre mayor proporción de GCs mayor será la T_m .

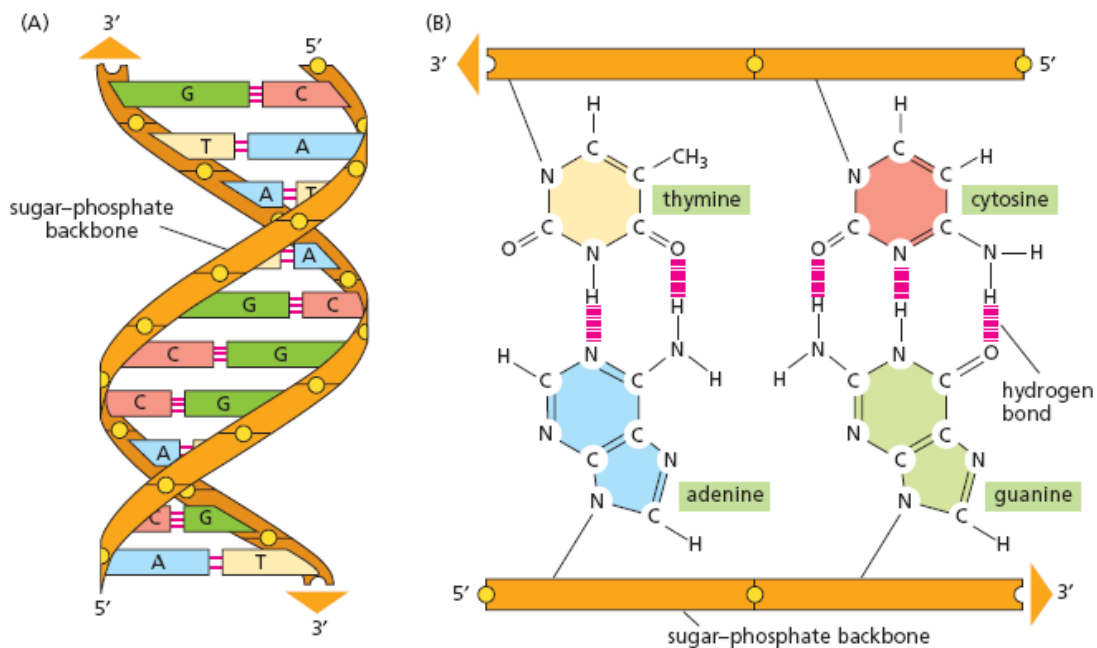


Figura 2-42. Complementariedad de las bases del DNA.

La convención establece que la secuencia de DNA siempre se debe de escribir **del 5 prima al 3 prima** (5' y 3', respectivamente). De cualquier forma, es una buena práctica siempre incluir el número prima cuando se escribe una secuencia dada. Sobre todo cuando uno la escribe en orden inverso.

Tarea 2-43

Si en un genoma bacteriano tenemos 15% de timina, ¿qué porcentaje de citosina tendremos? Diga si existe alguna regla que sea universal. Averigüe si en todos los virus esta regla de proporcionalidad entre diferentes bases también aplica. Averigüe que familias de virus tienen DNA de cadena sencilla. Explique las excepciones a esas reglas.

Respuesta: si hay 15% de timina en una cadena de DNA doble, entonces habrá 15% de adenina, y por consiguiente habrá 35% de citosina de 35% de guanina.

Tarea 2-44

Para discutir entre profesor y alumnos

¿Que hubiera pasado si Watson y Crick hubieran considerado la proporción variable de A, T, G y C en ciertos organismos? Considere esta hipótesis: La postulación de la doble hélice del DNA sólo fue posible al ignorar los resultados sobre la proporción variable de A, T, G y C en los virus de una sola cadena. Haga una tabla en defensa y en contra de la práctica científica de darle más peso a ciertos resultados e ignorar otros resultados.

Tarea 2-45

Una secuencia escrita en la nomenclatura estándar: ATGTCTGCGAT

Indique ahora cuales moléculas de las siguientes siete opciones si corresponden o no a esa misma doble cadena de DNA:

Secuencia	corresponde:	Si	No
3' -ATGTCTGCGAT-5'		☐	☐
5' -ATGTCTGCGAT-3'		☐	☐
3' -ATCGCAGACAT-5'		☐	☐
5' -ATCGCAGACAT-3'		☐	☐
5' -TAGCGTCTGTA-3'		☐	☐
3' -TAGCGTCTGTA-5'		☐	☐
3' -TACAGACGCTA-5'		☐	☐

Respuesta: Esta pregunta parece fácil, pero nuestra experiencia, es que menos del 20% de los estudiantes pueden contestar a todas las opciones correctamente. Tómese el tiempo necesario para responderla y entenderla bien. Para resolver este problema, lo mejor es escribir la doble cadena de la secuencia en cuestión, y también de cada una de las 7 opciones. Hecho esto ahora vea cuales son iguales o diferentes considerando la orientación y la hebra. Lea las secuencias siempre del cinco prima al tres prima. La segunda, la cuarta, la penúltima y la última si corresponden a la misma doble cadena que la secuencia en cuestión. Las demás no. Si no pudo responder esta pregunta o se equivocó en alguna opción, debe estudiar más sobre el DNA y sobre todo, entender los conceptos de la orientación y también de lo que es *reverse complement* (complemento inverso).

2.7.1.4.Replicación

La información que está contenida en el DNA se copia por medio de un mecanismo que se llama **replicación semiconservativa**. Esto se refiere a que el DNA se duplica formándose dos copias, en las cuales cada una contiene una hebra molde (cadena templado) y una hebra nueva (Figura 2-43).

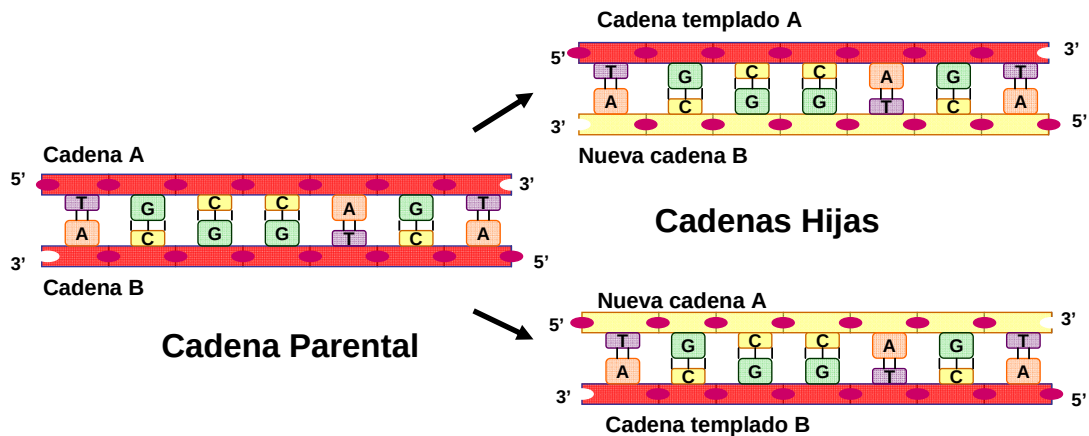


Figura 2-43. Replicación semiconservativa de DNA. La hebra de color rojo representa la hebra parental, mientras que la de color amarillo es una nueva hebra que se forma durante la replicación.

Cuando usamos una fotocopidora para copiar un documento, lo que hacemos es transferir la información de una hoja vieja a una hoja nueva. Al final obtenemos dos hojas con la misma información, pero una de las hojas es nueva, y la otra es la “vieja original”. A esta forma de copiar se le llama “conservativa”. Para generar dos hojas de papel de manera “semiconservativa” lo que tendríamos que hacer es partir la hoja original en dos, después fotocopiar cada una de ellas en una mitad de hoja y después volver a pegarlas

de forma que obtengamos dos hojas con información idéntica, pero que cada una de las copias tengan tanto papel nuevo, como papel viejo.

Tarea 2-46

Supongamos que quisiera averiguar si el DNA se copia de manera conservativa (como una fotocopidora normal) o de manera semiconservativa (mitad nueva y mitad original). Reflexione sobre esto y diseñe el experimento que debería hacer para demostrar la forma en que se replica el DNA. Como herramientas puede usar nucleótidos con una masa molecular mayor a lo normal y de esta forma distinguir las hebras de DNA viejas y nuevas que se generan *in vitro* con una polimerasa. Diseñe el protocolo del experimento y dibuje el resultado que esperaría. ¿Qué gráficas usaría para convencer a una audiencia científica de sus resultados? Haga las gráficas o figuras y presente sus conclusiones a su maestro de clases. Realice una búsqueda bibliográfica y averigüe la forma en que los científicos de los años 50 hicieron los experimentos para contestar esta pregunta (Experimento de Meselson y Stahl)

2.7.1.5. Síntesis direccional

Un concepto importante del DNA es que la síntesis de los ácidos nucleicos se lleva a cabo de manera direccional, del 5' prima al 3' prima. Esto tiene como consecuencia, que una de las hebras se puede copiar de manera continua, pero la otra se copia de manera discontinua en forma de pequeños tramos. Estos pequeños tramos se llaman fragmentos **Okazaki** (Figura 2-44). La replicación inicia en un zona del DNA con secuencias regulatorias, que en el caso de los plásmidos de bacterias se llama **ori** (del inglés *origin of replication*). A partir del ori se forma una **burbuja de replicación** que se va extendiendo hasta que se ha copiado la totalidad del DNA.

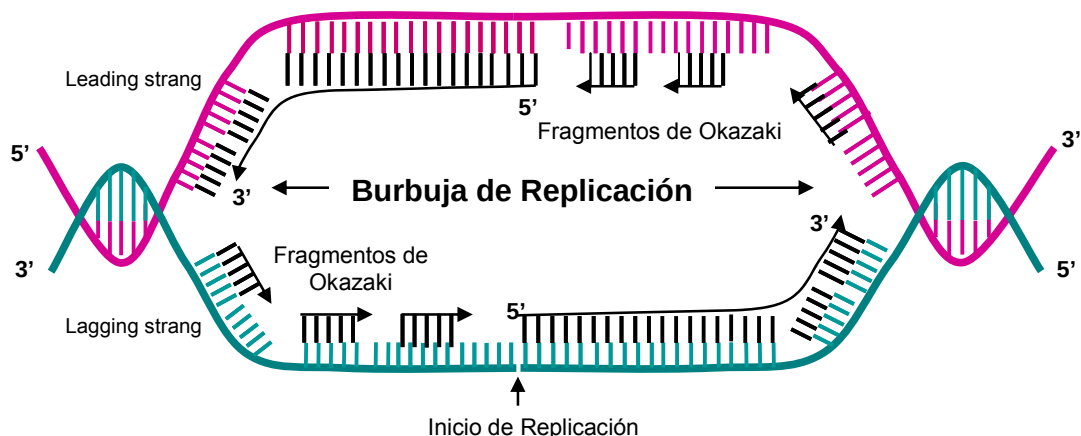


Figura 2-44. Burbuja de replicación con el *leading* y el *lagging strand*.

2.7.1.6.El gen

El concepto de gen surgió durante la época de Gregor Mendel, mucho antes de que se supiera del DNA. Hoy en día, decimos que un gen es un segmento lineal de DNA que puede contener información funcionalmente revelante, codificada por la secuencia de los nucleótidos. Se dice que el gen es la unidad básica de la herencia, ya que es responsable de un cierto carácter fenotípico unitario. Sin embargo, así como el átomo resultó ser divisible, el gen también está formado de elementos de menor tamaño. La Figura 2-45 muestra un gen típico de una bacteria.

Gen bacteriano

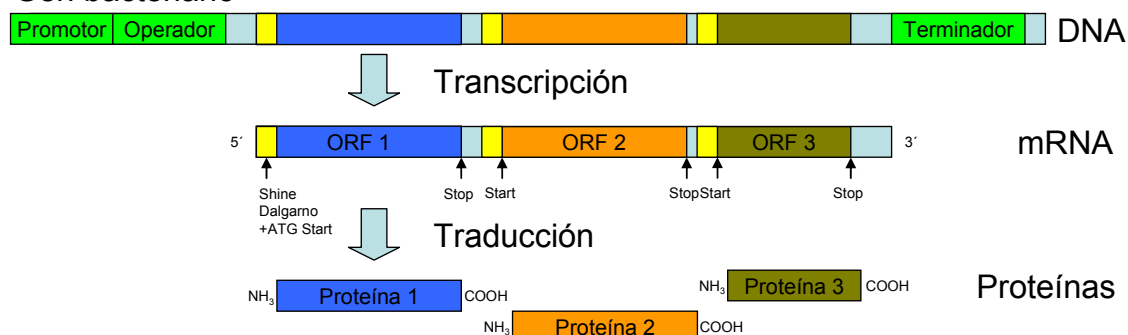


Figura 2-45. Estructura y expresión de un gen bacteriano.

Por lo regular, se dice que un gen es un segmento de DNA que codifica para una proteína. En ese caso, algunos genes bacterianos pueden codificar para varias proteínas, ya que el RNA tiene varios marcos de lectura abiertos (ORF por sus siglas en inglés: *open reading frame*). A diferencia de los genes bacterianos, los genes de las células eucarióticas son un poco más grandes y complejos. Además, se requiere de una edición del RNA mediante un proceso de **splicing y maduración** que es necesario para que el RNA mensajero (**mRNA**) pueda ser exportado del núcleo al citoplasma, donde se da la traducción (ver Figura 2-46).

Gen eucariótico

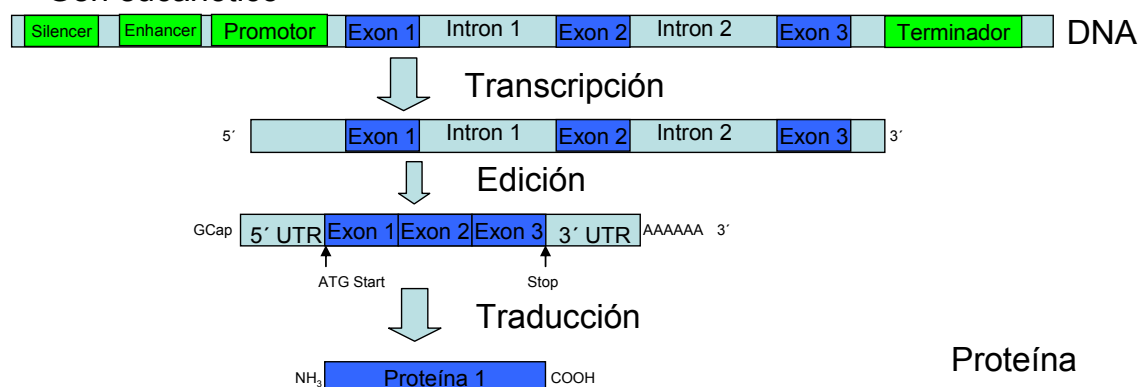


Figura 2-46. Estructura y expresión de un gen eucariótico.

Tarea 2-47

Realice un dispositivo o una fórmula que relacione el tamaño del mRNA con el tamaño del gen, el número de intrones, el tamaño medio de intrones y el tamaño de la región reguladora.

2.7.1.7. Dogma central

Uno de los primeros conceptos que se postularon en la biología molecular, fue el hecho de que el flujo de información era en una sola dirección a partir del DNA, pasando al RNA y después a las proteínas. Digamos que, aunque en la replicación sí puede haber un flujo de información bidireccional, la transcripción y la traducción se entendían como un flujo unidireccional (ver Figura 2-47.).

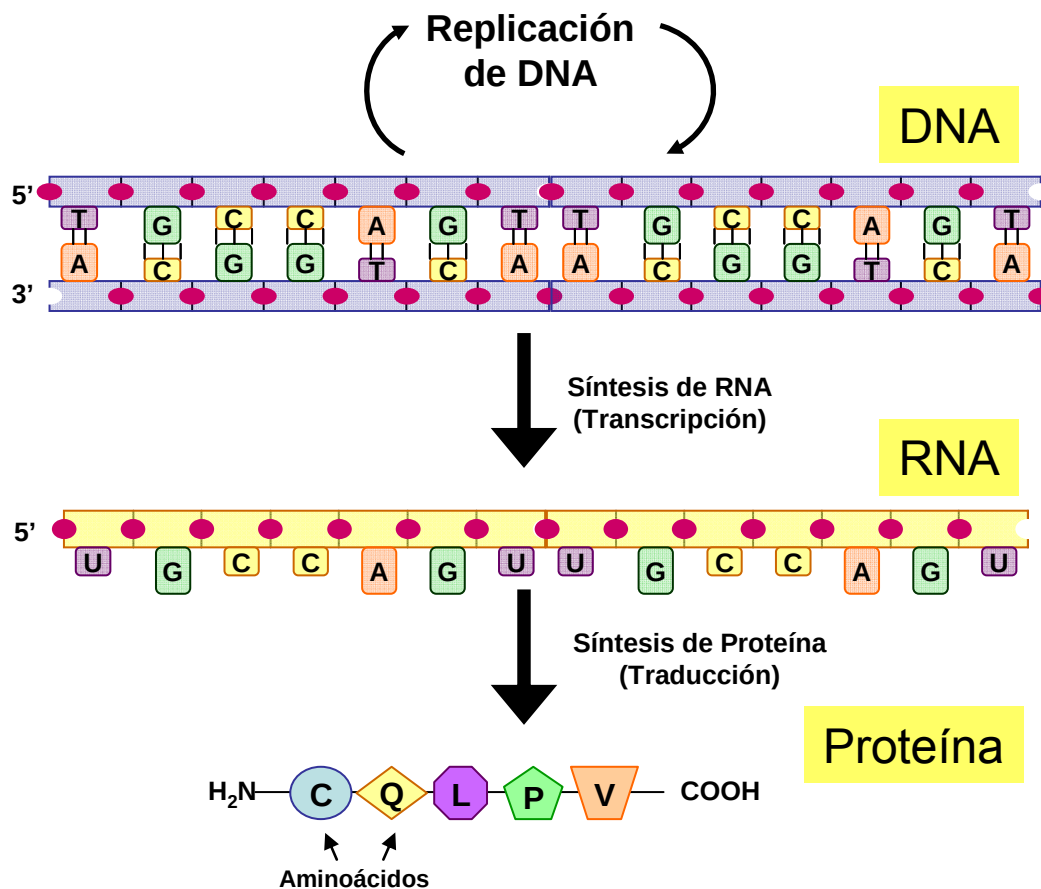


Figura 2-47. Flujo unidireccional de información: Dogma central de la biología molecular.

Para explicar eso a manera de una analogía, diríamos que el DNA es una molécula dictadora que no escucha ni tiene en cuenta la información que puedan adquirir sus moléculas súbditas como el RNA y las proteínas. Otra forma de verlo, es el debate entre la visión Lamarkista y la visión Darwinista de la evolución. Lamark decía que la experiencia y las características aprendidas durante la vida de un organismo eran heredables. Hoy en día, sabemos que la educación que tuvieron nuestros padres no la heredamos en forma de genes, es por eso que todos los niños y jóvenes tenemos que aprender muchas cosas desde cero.

Tarea 2-48**Para discutir entre profesor y alumnos.**

Trate de reflexionar un poco más sobre las consecuencias más amplias del dogma central de la biología molecular. Tratemos de hacerlo con una analogía sobre el aprendizaje. Por ejemplo el canto de los pájaros, ¿que tanto depende de los genes que heredan y de lo que aprenden durante jóvenes? Elabore una tabla en donde mencione las ventajas y desventajas de aprender algunos conocimientos desde el inicio. ¿Para qué tipo de habilidades es ventajoso heredarlas de manera innata y que otras es mejor adquirirlas durante nuestro desarrollo? ¿Por qué a veces no escuchamos a nuestros padres o maestros? Algunas cosas tenemos que aprenderlas por nuestras propias experiencias, ya que no solemos aprender mucho de los errores de nuestros antecesores. Discuta esto con sus compañeros de clases. Presenten sus conclusiones a su maestro de curso.

El postulado central de la biología molecular se hizo cuando todavía no se tenía información experimental suficiente que pudiera sustentar o refutar esta hipótesis de manera universal, y es por eso que se le llamó dogma. Actualmente sabemos que el dogma no es tan estricto como se postuló al inicio. Existe una enzima llamada **transcriptasa reversa** que a partir de RNA sí puede formar DNA. Entonces, a ese DNA se le llama **cdNA** (del inglés *copied DNA*). De hecho, esta enzima es la base de muchísimas aplicaciones de la ingeniería genética. Cada vez se conocen instancias en donde el flujo de información entre el DNA y el medio ambiente no es en un solo sentido.

Hoy en día todavía nos falta aprender mucho sobre las proteínas, pero ya hay indicios de que algunas proteínas pueden hacer cambios substanciales a la información del DNA y del RNA, ya sea en forma de edición de RNA, splicing diferencial, *exon shuffling*, metilación, mutación inducida, etc.

Tarea 2-49

En su cuaderno de trabajo prepare un glosario en donde defina lo que es: hipótesis, objetivo, teoría, ley, dogma, fe, evidencia, resultados, datos, análisis, síntesis, información. Ponga especial énfasis en las similitudes y diferencias entre estas palabras y los conceptos a los que se refieren. Use ejemplos de las ciencias biológicas para explicar de manera concreta estos términos.

2.7.1.8. Síntesis de proteínas usando el código genético universal

La síntesis de las proteínas a partir de un RNA mensajero se llama traducción. El RNA se lee del 5' al 3' y la proteína naciente se va formando del término amino al término carboxilo (ver Figura 2-48.).

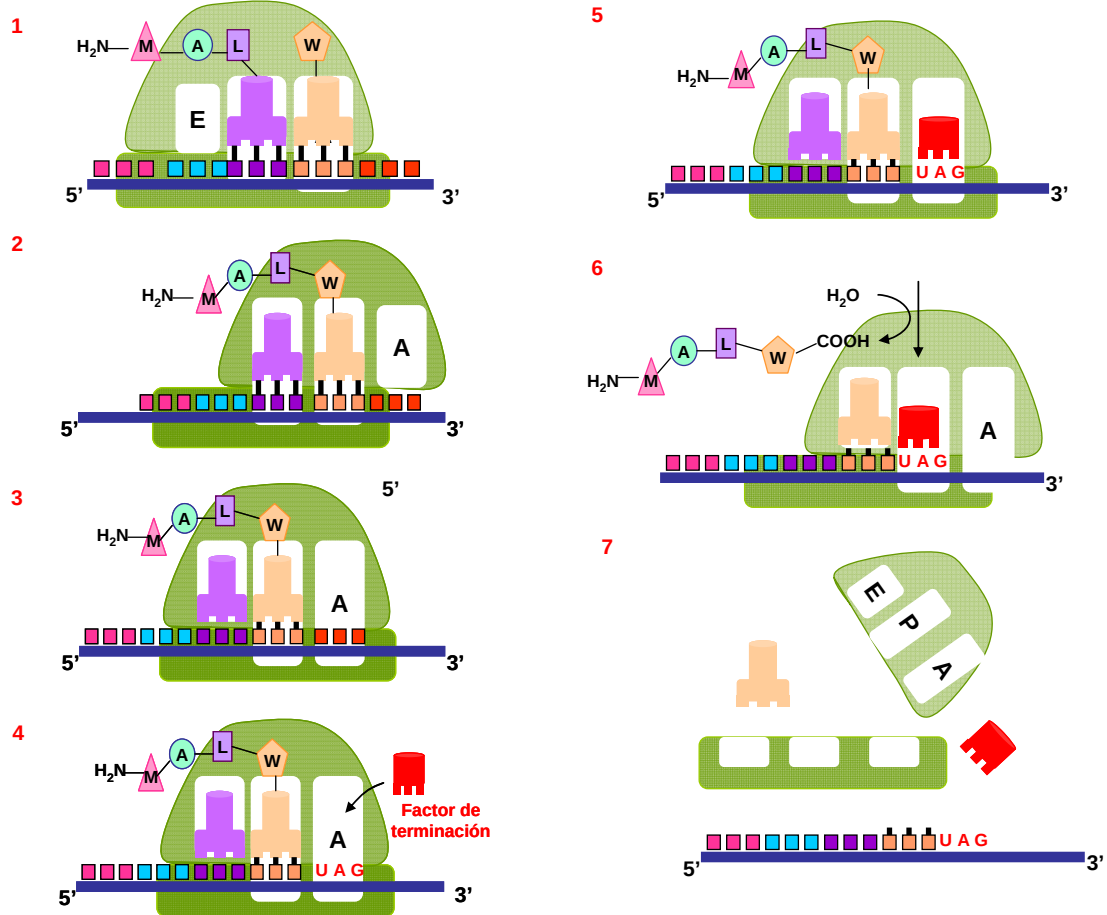


Figura 2-48. Ribosoma.

Este proceso se lleva a cabo en el ribosoma con ayuda de los RNA de transferencia (tRNA). Los tRNA van entrando al ribosoma según la información del mRNA y de manera simultánea se va formando la proteína. Cuando se llega al codon de paro (UAG), en lugar de un tRNA, entra un factor de terminación, con el cual se libera la cadena peptídica y todo el complejo de traducción (subunidad ribosomal pequeña y grande, y mRNA). Los tRNA asignan un aminoácido a cada una de las 64 combinaciones posibles de tres bases. La información de los ácidos nucleicos se lee en forma de **tripletes** (grupo de tres bases contiguas). Ese código de tripletes se ha encontrado casi inalterado en todos los organismos vivos y es por eso que se le ha llamado **código genético universal** (ver Figura 2-494).

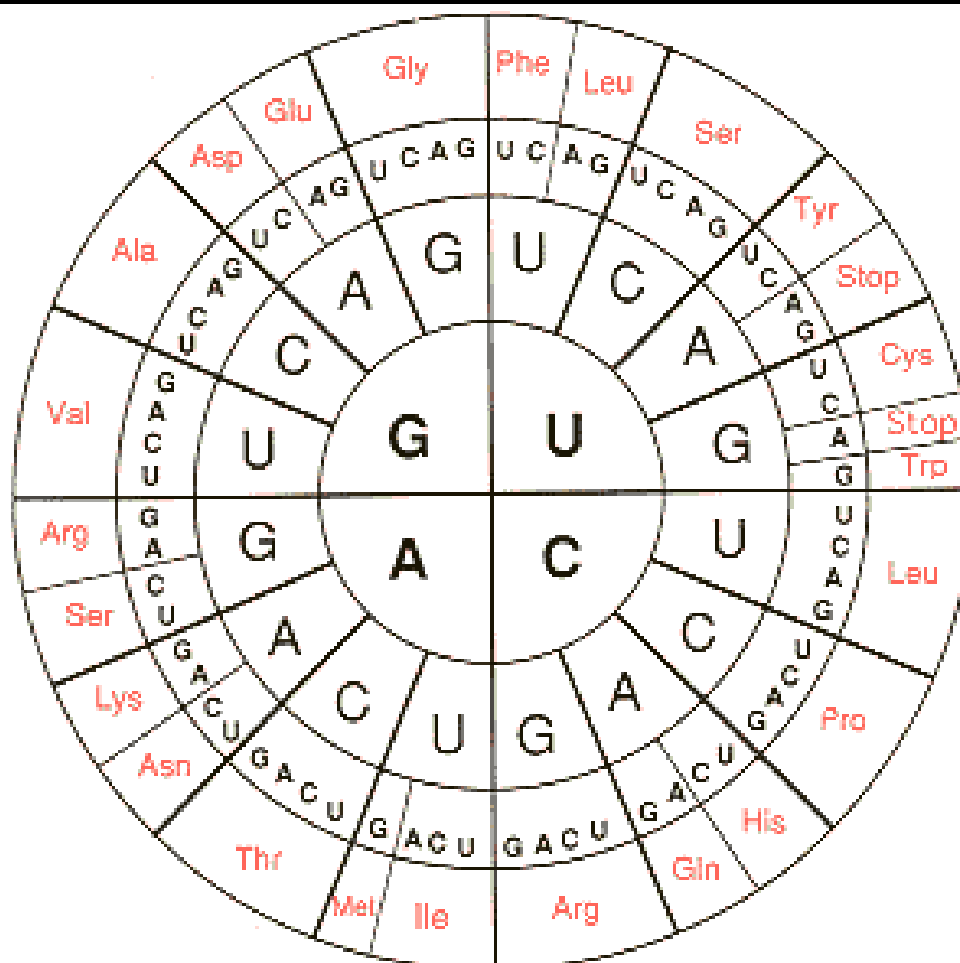


Figura 2-49. Código genético universal.

Una de las evidencias más contundentes de que todos los seres vivos actuales compartimos a un solo ancestro común, se deriva de la forma en que se lee el DNA. Ese código es el idioma genético que habló la célula primogénita de la cual se derivaron todos los organismos y especies que sobrevivieron después de miles de millones de años de selección y evolución. Una de las premisas más importantes para la ingeniería genética es precisamente esa, la universalidad del código genético. Si el DNA no se leyera de la misma forma en diferentes organismos, entonces no tendría ningún sentido aislar un gen de una bacteria y tratar de meterlo en el genoma de una planta.

Si usamos una analogía, diríamos que es como si todos los seres humanos hablaran el mismo idioma y de esta forma fuera más fácil intercambiar información entre personas de diferentes culturas. Por ejemplo, podríamos usar los mismos libros que los griegos, estadounidenses y los chinos y no tendríamos tanto trabajo de hacer traducciones entre los diferentes libros de texto.

Tarea 2-50

Elabore una presentación explicando que son los tripletes y como se lee el código genético universal.

2.7.1.9. Abreviaturas de los aminoácidos

Los estudiantes que pretendan trabajar con genes y proteínas, tarde o temprano tendrán que aprender las abreviaturas de una letra de los diferentes aminoácidos que conforman las proteínas. Los que deseen ser expertos, adicionalmente tendrán que tener una buena noción de los grupos químicos funcionales que están presentes en cada uno de los aminoácidos.

Abreviación		Nombre completo	Triplete típico	Molécula
1 letra	3 letras			
D	Asp	Ácido aspártico	GAT	$\text{HOOC-CH}_2\text{-CH(NH}_2\text{)-COOH}$
E	Glu	Ácido glutámico	GAG	$\text{HOOC-(CH}_2\text{)}_2\text{-CH(NH}_2\text{)-COOH}$
A	Ala	Alanina	GCT	$\text{CH}_3\text{-CH(NH}_2\text{)-COOH}$
R	Arg	Arginina	CGT	$\text{HN=C(NH}_2\text{)-NH-(CH}_2\text{)}_3\text{-CH(NH}_2\text{)-COOH}$
N	Asn	Asparagina	AAC	$\text{H}_2\text{N-CO-CH}_2\text{-CH(NH}_2\text{)-COOH}$
C	Cys	Cisteina	TGC	$\text{HS-CH}_2\text{-CH(NH}_2\text{)-COOH}$
F	Phe	Fenilalanina	TTT	$\text{Ph-CH}_2\text{-CH(NH}_2\text{)-COOH}$
G	Gly	Glicina	GGG	$\text{NH}_2\text{-CH}_2\text{-COOH}$
Q	Gln	Glutamina	GAG	$\text{H}_2\text{N-CO-(CH}_2\text{)}_2\text{-CH(NH}_2\text{)-COOH}$
H	His	Histidina	CAT	$\text{NH-CH=N-CH=C-CH}_2\text{-CH(NH}_2\text{)-}$

				COOH
I	Ile	Isoleucina	ATA	CH ₃ -CH ₂ -CH(CH ₃)- CH(NH ₂)-COOH
L	Leu	Leucina	CTC	(CH ₃) ₂ -CH-CH ₂ - CH(NH ₂)-COOH
K	Lys	Lisina	AAG	H ₂ N-(CH ₂) ₄ - CH(NH ₂)-COOH
M	Met	Metionina	ATG	CH ₃ -S-(CH ₂) ₂ - CH(NH ₂)-COOH
P	Pro	Prolina	CCC	NH-(CH ₂) ₃ -CH- COOH
S	Ser	Serina	AGT	HO-CH ₂ -CH(NH ₂)- COOH
Y	Tyr	Tirosina	TAC	HO-p-Ph-CH ₂ - CH(NH ₂)-COOH
T	Thr	Treonina	ACG	CH ₃ -CH(OH)- CH(NH ₂)-COOH
W	Trp	Triptófano	TGG	Ph-NH-CH=C-CH ₂ - CH(NH ₂)-COOH
V	Val	Valina	GTT	(CH ₃) ₂ -CH-CH(NH ₂)- COOH

Tarea 2-51

Traduzca al idioma de las proteínas la siguiente secuencia nucleotídica. Use la abreviatura de una letra de los aminoácidos.

atggtgaaggcctggtggtgggcgcggcg
gtcgtggcgtgcgcggcgggtgggggtggcg

gcggtgctggtgcaccgcggcgcaggcgc
gacgcgcgcgtcctgggggtccgcggaggcg

gagaggagggcggcgcgcggccgcgctgatac
gaggaggtggagcgcagcctcgccacgccc

acggcgctgctgcggggcatcgcggaacgcc
atggtcgcccagatggagcgcggcctgcgc

ggggacatccacgcgcagctcaagatgctc
attagctacgtcgacaacctccccaccgga

gatgaacatgggctgttttatgcactggat
cttggagggaccaacttccgtgtgctgcga

Tarea 2-52

¿Qué significa que el código genético esté degenerado? Usando el código genético (Figura 2-49) traduzca la siguiente secuencia proteínica a su respectiva secuencia nucleotídica:

M D F E S L N P G E Q I Y E K M I
S G M

Y L G E I V R R I L L K L A H D A
S L F

G D V V P T K L E Q P F I L R T P
D M S

A M H H D S S H D L K T L G S K L
K D I

V G V A D T S L E V R Y I T R H I
C D L

V	A	E	R	G	A	R	L	A	A	A	G	I	Y	S	I	L
K	K	I														

Tarea 2-53

Considere el siguiente fragmento de DNA:

5' GCTTCCCAAGGT 3'
3' CGAAGGGTTCCA 5'

Si la cadena superior es la cadena de la plantilla usada por la polimerasa de RNA.

- Dibuje el RNA transcrito.
- Etiquete sus terminos 5' and 3'.
- Escriba las tres posibles cadenas correspondientes de aminoácidos.
- Etiquete sus extremos amino y carboxilos.
- Repita los pasos a-d asumiendo que la cadena inferior es la cadena de la plantilla.

En caso que tenga dificultades o quiera profundizar más, responda los problemas del capítulo 2 y 3 de Griffiths *et al*, Modern Genetic Analysis.

Tarea 2-54

Aprenda más sobre genética. Consulte la página:

<http://www.ucm.es/info/genetica/grupod/>

Tarea 2-55

Aprenda más sobre genética. Consulte la página:

<http://www.ncbi.nlm.nih.gov/books/bv.fcgi?call=bv.View..ShowTOC&rid=mga.TOC&depth=10>

Introduzca una palabra clave en la ventana de búsqueda de esa página. Despues saldrán los links a los capítulos en donde viene explicado ese término.

Tarea 2-56

Consulte los diversos libros electrónicos que estan disponibles en la página del NCBI:

<http://www.ncbi.nlm.nih.gov/sites/entrez?db=books>

Libros recomendados para consulta:

- Griffiths AJF, Gelbart WM, Miller JH, Lewontin RC. Modern Genetic Analysis. WH Freeman & Company.
- Watson, J.D., Caudy, A. A., Myers, R.M. and Witkowski, J.A.; Recombinant DNA: Genes and Genomes — A Short Course; Third Edition; Cold Spring Harbor Lab. Press; CSH New York; 2007.
- Baum, S. J., Scott-Ennis, R. J., and Hill, J W; Chemistry and Life: An Introduction to General, Organic and Biological Chemistry; Prentice Hall; 1999.
- Crothers, D. M., Kill, J., Tinoco, I., Hearst, J. E., Wemmer, D. E., and Bloomfield, V. A.; Nucleic Acids: Structures, Properties, and Functions; University Science Books; 2000.
- Gait, M. G., and Blackburn, G. M.; Nucleic Acids in Chemistry and Biology; Oxford University Press Inc.; 1995.
- Van Holde, K.E., Johnson, W.C., and Ho, P.S.; Principles of Physical Biochemistry; Prentice Hall Inc.; 1998
- Creighton, T. E.; Proteins: Structures and Molecular Properties; W.H. Freeman & Co., 1992.

2.7.2. Nociones avanzadas del DNA

2.7.2.1. Conformaciones de la doble hélice

La doble hélice del DNA puede estar en diferentes conformaciones. La que más se conoce es la conformación B, que es la que primero se descubrió por Watson, Crick y Wilkins. Sin embargo, existen otras conformaciones como la Z y la A, que también se presentan *in vivo* dentro de las células. Todavía no se sabe la relevancia de este fenómeno, pero se intuye que tiene importancia para la forma en que las proteínas se pueden unir al DNA, y así, influyen en la forma en que se expresan los genes en una determinada región. Aunque todavía no se sabe si la conformación es heredable, la conformación A, B, Z del DNA puede ser considerado un factor epigenético, ya que puede afectar la expresión de los genes, pero no es directamente atribuible a la secuencia de los nucleótidos *per se*. (Figura 2-50).

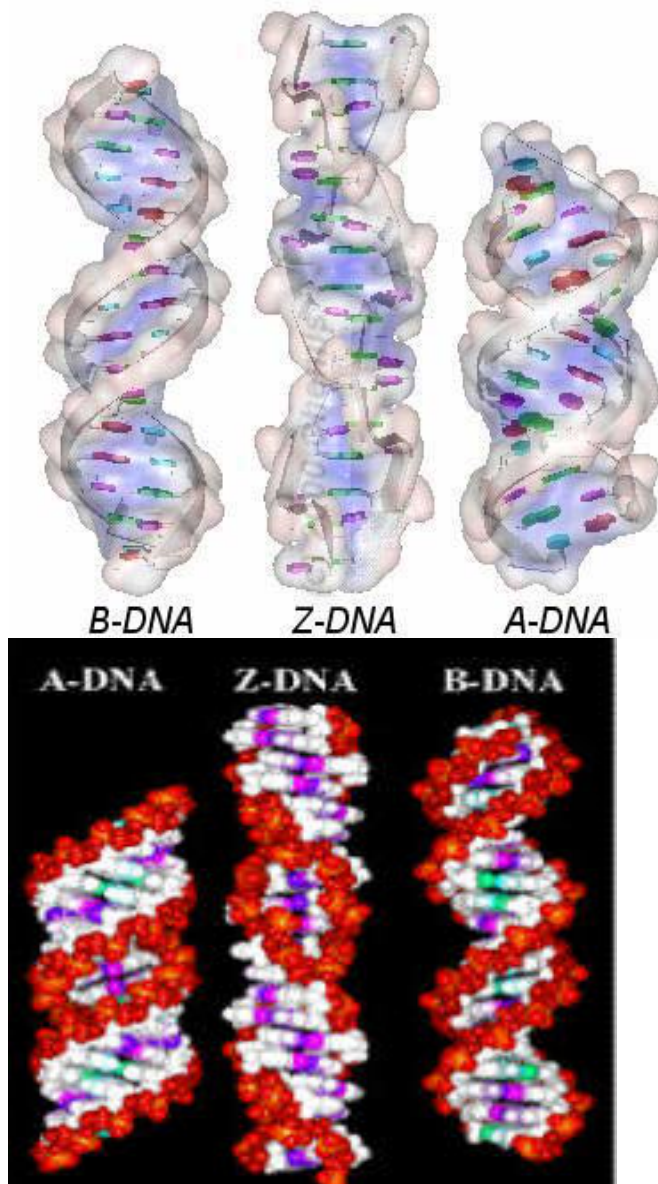


Figura 2-50 Diferentes Conformaciones de DNA.

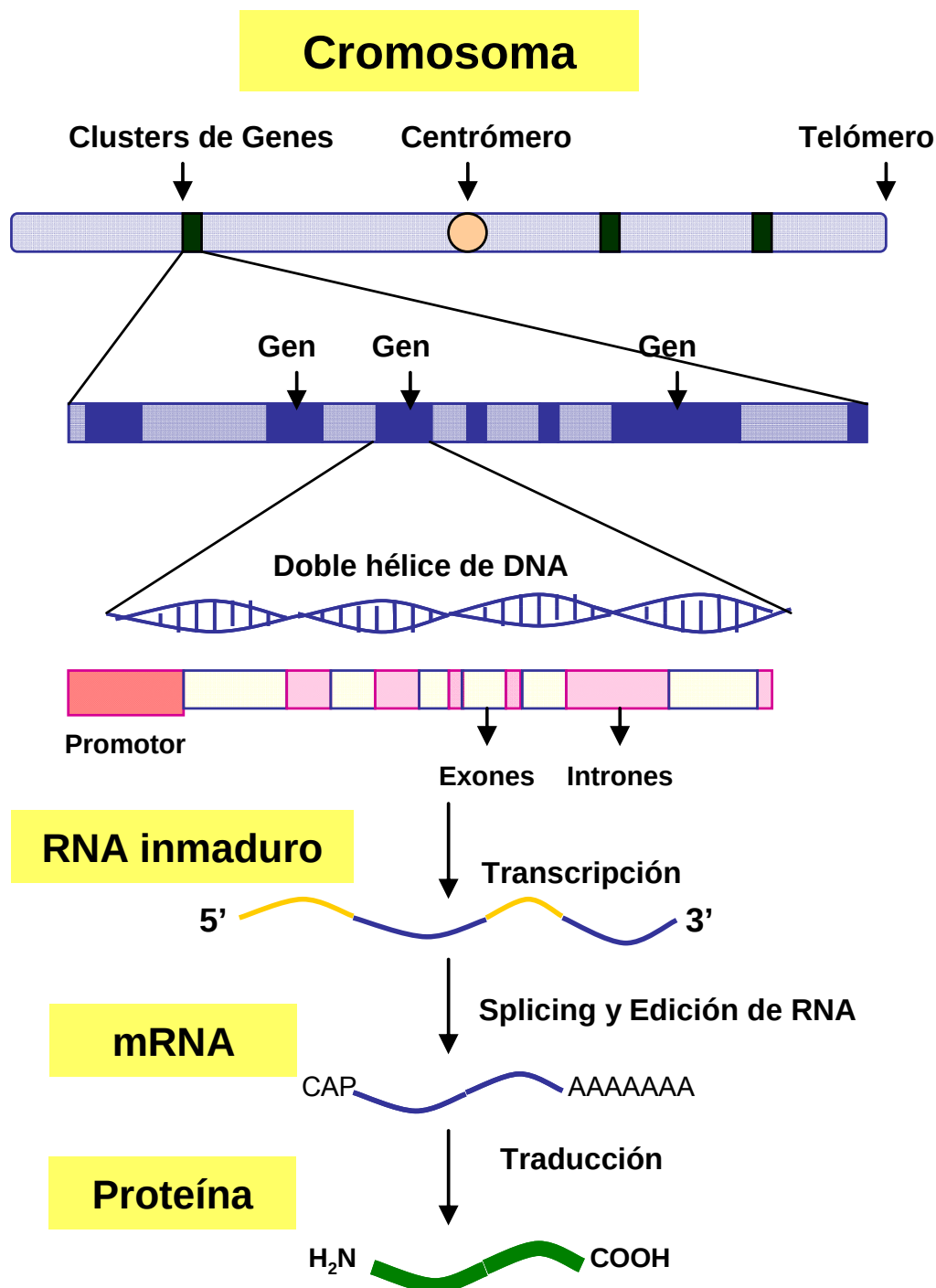


Figura 2-51 Estructura de cromosomas y genes. Se muestra también el proceso de maduración del RNA y la conversión final en una proteína.

Tarea 2-57

¿Qué es un gen? ¿Cómo se expresa una proteína? Elabore un ensayo explicando con detalle todos los pasos moleculares que se llevan a cabo, desde el DNA con su estructura de promotor y exones, pasando del RNA inmaduro, al RNA maduro, hasta la proteína. Use la Figura 2-46 y la Figura 2-51 como guía y explique todos los procesos que están involucrados.

Indique también en que compartimento celular se lleva a cabo cada proceso, ya sea núcleo, citoplasma o retículo endoplasmático.

2.7.2.2. Complejo de replicación

Complejo Replisoma

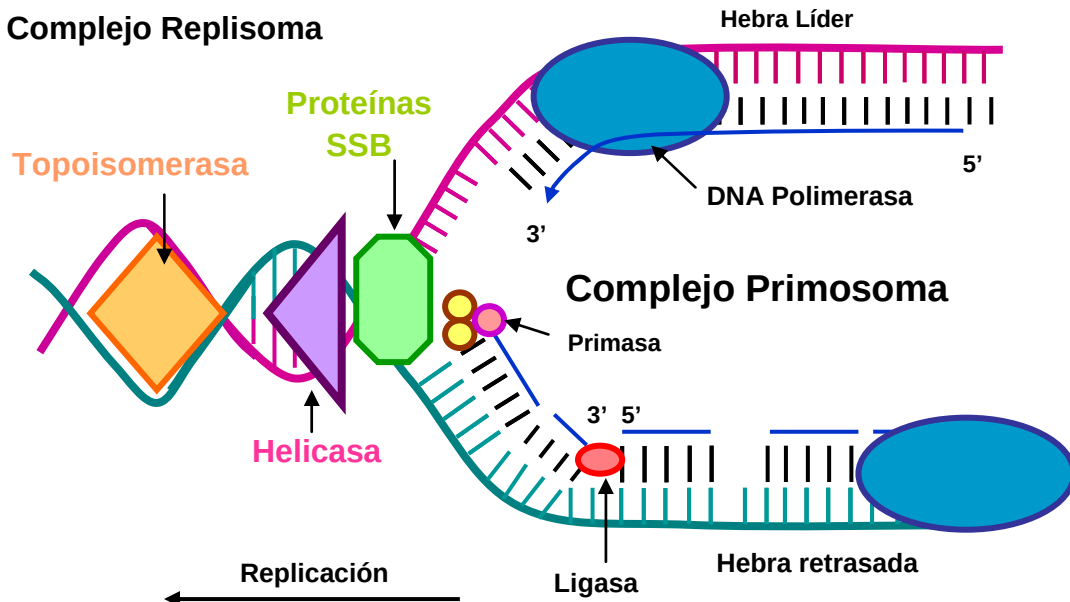


Figura 2-52 Complejo de replicación. Una serie de diversas proteínas y enzimas se agrupan en grandes complejos proteínicos que permiten así la replicación del DNA.

Tarea 2-58

Explique la función de las diferentes enzimas que participan en la replicación del DNA. ¿Qué hace la polimerasa? ¿Cuál es la función de la helicasa? ¿Qué hace la topoisomerasa? ¿Para qué sirven las proteínas SSB? ¿En qué momento actúa la ligasa? ¿Por qué la replicación es diferente para cada hebra del DNA? En la Figura 2-52 ¿cuál es el *leading strand* y cuál es el *lagging strand*?

Tarea 2-59

Aprenda más sobre la replicación. Consulte la página:

<http://www.ucm.es/info/genetica/grupod/Replicacion/Replicacion.htm>

2.7.2.3. Intrones y exones

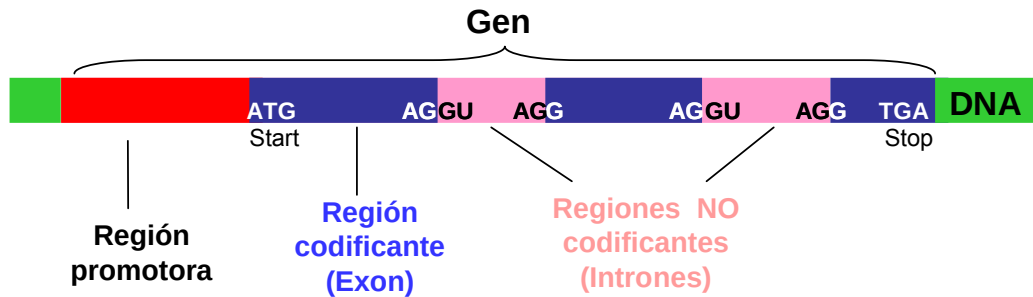


Figura 2-53. Promotor, intrones y exones.

Tarea 2-60

¿Cuál es la diferencia entre un exon y un intron? ¿Qué repercusiones genéticas, moleculares y evolutivas puede haber?

Tarea 2-61

Averigüe más sobre las secuencias cortas para las señales de: start, stop, sílice-start, splice-end. ¿De qué sirve esto para las predicciones de genes por métodos bioinformáticos?

2.7.2.4. Mecanismo de splicing

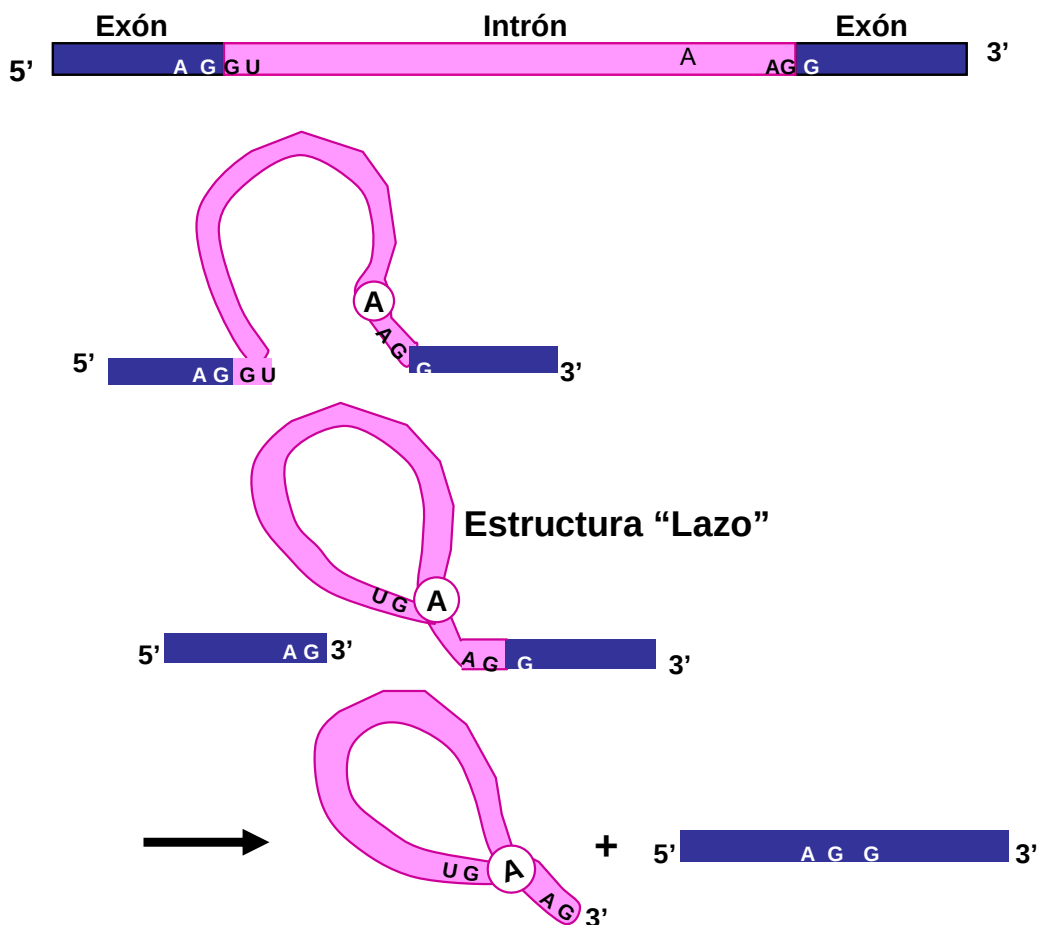


Figura 2-54. Formación de un lazo de splicing.

2.7.2.5. Splicing diferencial

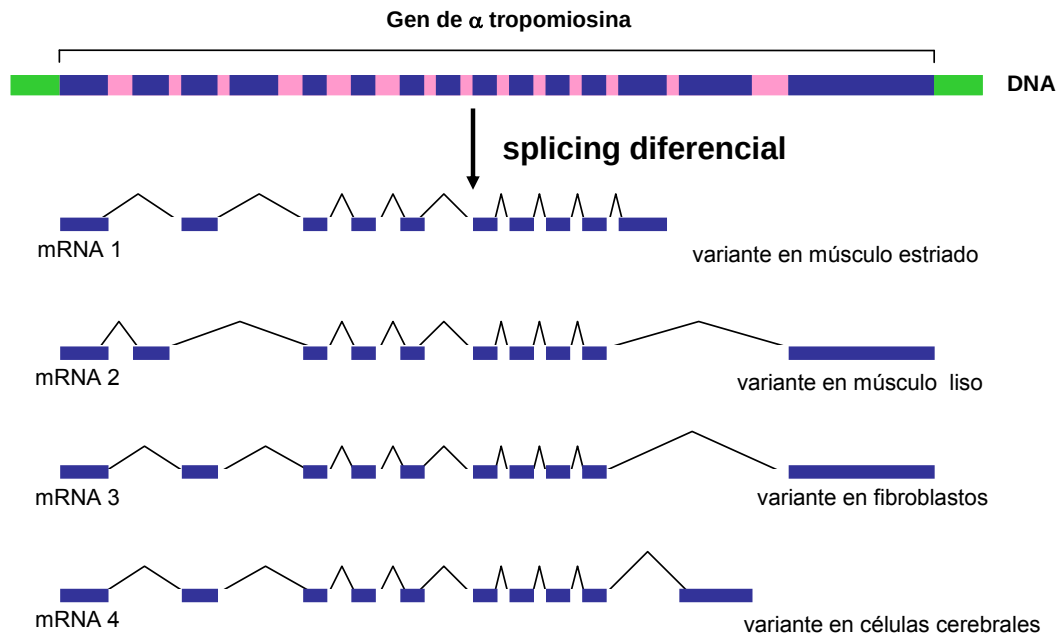


Figura 2-55. Ejemplo de splicing diferencial. Gen humano de tropomiosina. Denz et al., (2004) Biochem. Biophys. Res. Commun. 320(4):1291-7

2.7.2.6. Mecanismo de recombinación

La recombinación comienza cuando dos moléculas de DNA se unen para formar una sinapsis de recombinación. Las hebras duplex del DNA se unen a la enzima recombinasa, el extremo 3' terminal de cada una de las hebras desenrolladas está vinculado a un residuo de tirosina (Y) de la enzima recombinasa. Cuando un extremo 5' es cortado de otra hebra en el complejo, se forman nuevos enlaces fosfodiester. En seguida el complejo se isomeriza y se repiten los cortes y formación de enlaces fosfodiester, logrando de esta manera un producto de recombinación (figura 2-50).

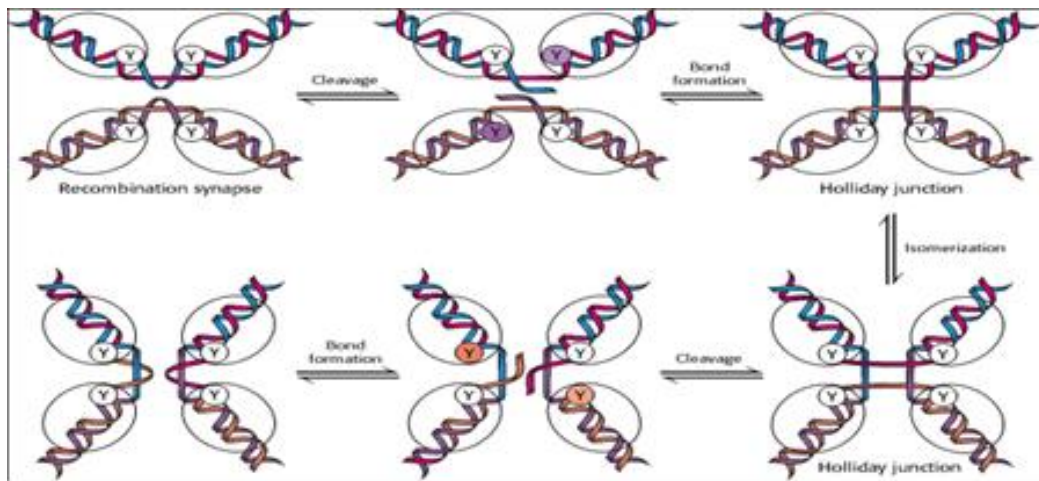


Figura 2-56. Mecanismo de recombinación.

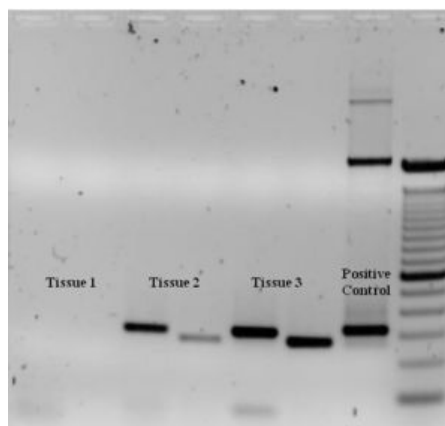
3. Herramientas moleculares

3.1. PCR

3.1.1. Historia y principios

La reacción en cadena de la polimerasa, conocida como PCR por sus siglas en inglés (**P**olymerase **C**hain **R**eaction), es una técnica para obtener un gran número de copias de un fragmento de DNA en particular. Fue descrita en 1986 por Kary Mullis. Una de las primeras referencias fue:

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, *et al.* (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239: 487–491



Para realizar la técnica de PCR se necesita:

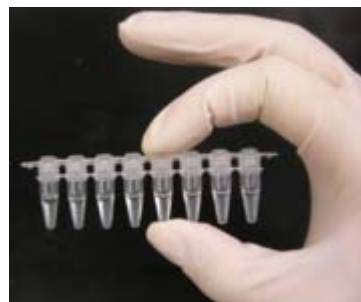
- DNA polimerasa que sea estable a temperaturas altas. La más común es la polimerasa de *Thermus aquaticus* (Taq) que resiste los 100°C.
- Mezcla de cuatro desoxinucleótidos trifosfato (dNTPs), que son los sustratos para polimerizar el nuevo DNA. (dATP, dTTP, dCTP y dGTP)
- DNA molde (DNA *template*) con la región que se va a amplificar.
- Dos *primers* (oligos, iniciadores o cebadores) complementarios al DNA molde. Son secuencias cortas, normalmente de 20 a 25 nucleótidos, que son reconocidos por la polimerasa para iniciar la reacción de elongación. Deben estar orientados correctamente (antiparalelos mostrándose uno hacia el otro) para que puedan amplificar exponencialmente.
- Iones divalentes como el magnesio (Mg^{2+}). El magnesio se une a los nucleótidos estabilizando los enlaces diésteres de fosfato. Actúa como cofactor de la polimerasa.
- Iones monovalentes, como el potasio.
- Una solución amortiguadora (*buffer*) que mantiene el pH 8-9.
- Termociclador, el aparato que va a mantener la temperatura necesaria en cada una de las etapas que conforman un ciclo. Hoy en día los nuevos termocicladores tienen todos una tapa caliente para evitar la evaporación-



condensación del líquido, pero de todos modos en algunos casos se suele agregar una gota de aceite para evitar la evaporación.

La PCR normalmente se realiza con un volumen de reacción de 20-50 μl en pequeños microtubos de polipropileno estériles que se colocan en el termociclador. Por lo regular se hace una mezcla de reacción (*master mix*) de la reacción de PCR (agua, amortiguador, dNTPs y Taq agregados en ese orden). Ese *master mix* se reparte en varios tubos a los que posteriormente se les agrega ya sea 1 μl de templado de DNA y el par de *primers* específicos.

Por lo regular se usa Tris como buffer y cloruro de magnesio (MgCl_2) y cloruro de potasio (KCl) como sales.



Ciclo de amplificación

El proceso de PCR por lo general consiste en una serie de 20 a 35 ciclos; cada ciclo suele consistir de dos a tres pasos de temperaturas diferentes. Las temperaturas usadas y el tiempo aplicado en cada ciclo dependen de una gran variedad de parámetros. Estos incluyen la enzima usada para la síntesis de DNA, la concentración de iones divalentes, la cantidad y proporción de dNTPs en la reacción y la temperatura de unión de los *primers*.

Inicialización

Este paso consiste en llevar la reacción hasta una temperatura de 95°C, que se mantiene durante 1-9 minutos. Esto es necesario para las DNA polimerasas (hot start) que requieren activación por calor. Este paso también es útil cuando se está usando DNA genómico que requiere de más tiempo para desnaturalizarse completamente. Sin embargo, solamente es necesario agregar este paso de inicialización antes del primer ciclo.

Desnaturalización

Este paso se realiza también por medio de calentamiento a 95°C. La temperatura elevada (energía cinética molecular) rompe los enlaces de hidrógeno y hace que las dos hebras se separen (el DNA se desnaturaliza).

Unión del primer

A continuación se producirá la hibridación del *primer*. Este se une a su secuencia complementaria en el DNA molde. Para ello es necesario bajar la temperatura a 50-65°C (según el valor de la temperatura de alineación T_m de los *primers*) durante 20-40 segundos permitiendo así el alineamiento. Los puentes de hidrógeno estables entre las cadenas de DNA (unión DNA-DNA) sólo se forman cuando la secuencia del *primer* es muy similar a la

secuencia del DNA molde. La polimerasa une al pedazo de doble cadena formado por el *primer* y el DNA molde y empieza a sintetizar DNA. Los *primers* definirán los límites de la región de la molécula que va a ser amplificada.

Extensión/elongación de la cadena

En esta fase, la DNA polimerasa se une al fragmento de doble cadena que forma el DNA molde y el *primer* e inicia la síntesis. La polimerasa sintetiza una nueva hebra de DNA complementaria a la hebra molde añadiendo los dNTP's complementarios en dirección 5'→ 3', uniendo el grupo 5'- fosfato de los dNTPs con el grupo 3'- hidroxilo del final de la hebra de DNA creciente (la cual se extiende). La temperatura para este paso depende de la DNA polimerasa que usemos. Para la Taq polimerasa, la temperatura de máxima actividad es de 72°C. Como regla general se le da 1 min por cada mil bases de fragmento que se piensa amplificar.

Elongación final

Etapa única que se lleva a cabo a una temperatura de 72°C entre 5-15 minutos tras el último ciclo de PCR. Con ella se asegura que cualquier DNA de cadena simple restante es totalmente ampliado. También se le da tiempo a la Taq-polimerasa para que agregue un nucleótido terminal de A que puede servir después para clonar el fragmento en vectores con un *overhang* T (Vectores tipo TOPO como pGEM)

Para verificar que la PCR ha generado el fragmento de DNA previsto, se emplea electroforesis para separar los fragmentos de PCR de acuerdo a su tamaño. Típicamente se emplean geles de agarosa horizontales para fragmentos grandes; y geles de acrilamida verticales para los más pequeños. El tamaño de los productos de la PCR se compara con marcador de peso molecular conocido, que se corren en una de las líneas del gel (1) (Figura 3-1). Es una buena práctica poner un control positivo y un control negativo para todas las PCRs que se realizan en el laboratorio. Se deben anotar las condiciones exactas del ciclo de temperaturas, así como los volúmenes de reactivos y las muestras que se usaron para cada experimento.

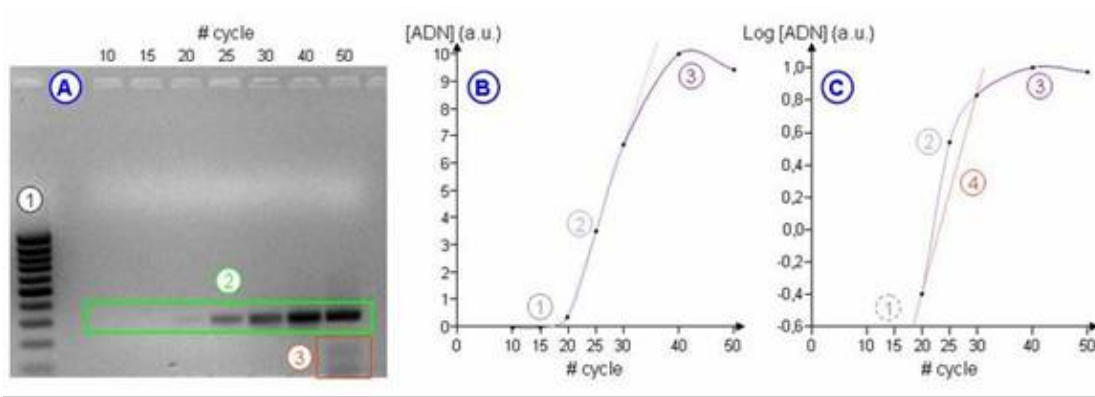


Figura 3-1. Productos de PCR después de diferente número de ciclos

Tarea 3-1

Un estudiante pipeteó los siguientes componentes:

10 μ L 10x PCR buffer solution (500 mM KCl, 100 mM Tris-Cl, 30 mM DTT, 15 mM MgCl₂, 1 mg/ml BSA)

10 μ L 10x dNTP stock (2mM of each dNTP in TE buffer)

1 μ L Oligo primer 5' stock (50 μ M)

1 μ L Oligo primer 3' stock (50 μ M)

76 μ L sterile H₂O

1 μ L Template DNA

Cuando corrió la PCR se dio cuenta que no hubo amplificación. ¿Podría decir por qué no funcionó? ¿Qué le hizo falta?

Tarea 3-2

Otro estudiante programó el termociclador de la siguiente manera:

Step	Temperature	Time (seconds)
1	95 °C	300 s
2	55°C	30 s
3	72°C	60 s
4	Go to step 1	for 30 cycles
5	4°C	Hold

Cuando dejó corriendo su reacción de PCR se dio cuenta de que la máquina tardó mucho en terminar todo el programa. Sin embargo, el equipo era un poco viejo y por eso pensó que era lento. Cuando corrió su gel, la amplificación no fue muy buena. ¿Podría usted postular una hipótesis de por qué la máquina tardó tanto, y al final se dio tan mala amplificación?

Respuesta a tarea 1: se le olvidó agregar 1 μ L Taq polymerase (2.5 units/mL)

Respuesta a tarea 2: el paso 1 de desnaturalización a 95°C por 300s sólo se debe hacer la primera vez por tanto tiempo. Ya después en los ciclos repetitivos es suficiente hacerlo por solo 30 s. La Taq polimerasa pierde actividad a altas temperaturas prolongadas y además se corre el riesgo de que los nucleótidos se degraden y la muestra se evapore. Mejor sería realizar un programa como:

Step	Temperature	Time (seconds)

1	95°C	270 s
2	95°C	30 s
3	55°C	30 s
4	72°C	60 s
5	Go to step 2	for 30 cycles
6	4°C	Hold

3.1.2. Técnicas y aplicaciones

La PCR es hoy en día una de las técnicas más importantes para un laboratorio de biología molecular. Se usa tanto para las amplificaciones de genes y las clonaciones, como para la detección de los niveles de expresión (PCR en tiempo real por sus siglas en inglés RT-PCR), como para la secuenciación de genes (método Sanger). La PCR es la técnica que primero debe dominar cualquier estudiante que pretenda hacer algo de biología molecular. Existen muchos libros y manuales que explican los trucos y los secretos del PCR. También existen muchos sitios web en donde podrá encontrar mayor información. Sin embargo, únicamente con la práctica en el mismo laboratorio se puede hacer al maestro. Hacer una PCR puede ser muy fácil, pero también puede ser un dolor de cabeza. No es fácil optimizar las variables para una buena amplificación (obtener suficiente DNA, pero sin que salgan otras bandas inespecíficas).

Tarea 3-3

Busque algunas aplicaciones de PCR. Haga un resumen de dos posibles aplicaciones para el mejoramiento genético. Para más información, consulte el siguiente link:

<http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=hmg.section.580>

Tarea 3-4

Haga una búsqueda en Genbank para encontrar el gen de la hexocinasa de maíz y diseñe unos *primers* para amplificar un fragmento de 100 pares de bases. La información de cómo usar esta base de datos se encuentra en el capítulo 7. Ponga los dos primers, el *forward* y el *reverse*, en su orientación estándar del 5' al 3'.

3.2. Modificadores del DNA

Para poder hacer cambios dirigidos en el material genético de una especie se necesitan herramientas que puedan modificar el DNA de una manera predecible. Por ejemplo: para cortar un gen de un cromosoma y después insertarlo en algún otro sitio en el genoma de otra especie, se necesita un cierto tipo de enzimas que cortan y pegan la cadena de ácidos nucleicos en sitios específicos. Podemos decir que para hacer ingeniería genética primero necesitamos tener, a nivel molecular, unas tijeras y también un

pegamento. A continuación se describen con más detalle algunas de esas enzimas que hacen ese trabajo.

3.2.1. Enzimas de restricción

Las enzimas de restricción cortan el DNA en regiones específicas. Por lo regular son secuencias correspondientes a palíndromos invertidos de 4 o 6 bases. El corte puede resultar en una doble cadena con terminación simple (**blunt**, o **extremos romos**) o resultar en una terminación de cadena sencilla complementaria (**sticky end**, o **extremos cohesivos**). Esto tiene enorme relevancia para la clonación de genes en sitios específicos y en una orientación determinada.

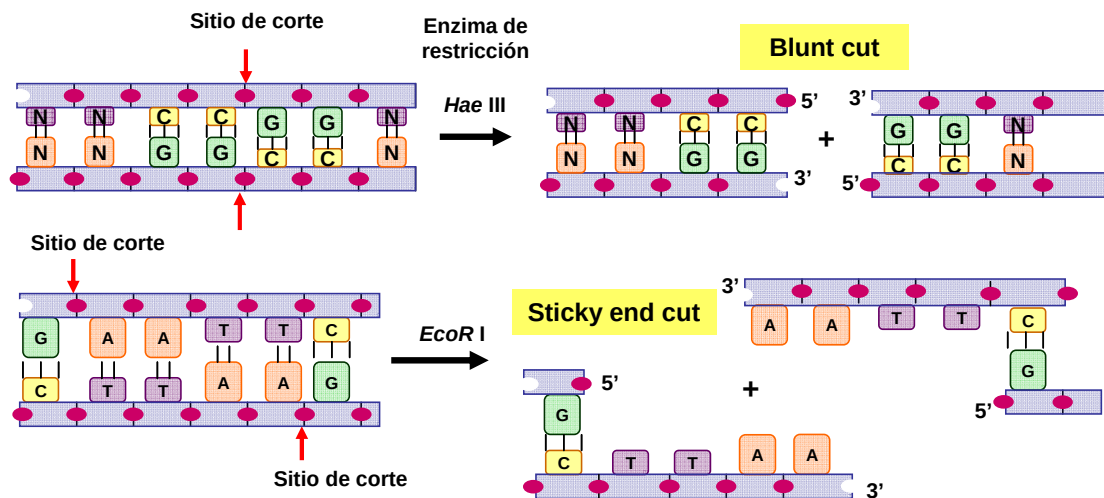


Figura 3-2. Ejemplos de cortes de restricción. Arriba se muestra la enzima *Hae*III que hace un corte tipo *blunt*, mientras que abajo la enzima *Eco*RI hace un corte tipo *sticky end*.

3.2.2. Ligasas

Si las enzimas de restricción son las tijeras microscópicas, entonces las ligasas son el pegamento molecular. El uso combinado de estos dos tipos de enzimas permite hacer el *cut & paste* (cortar y pegar) de la ingeniería genética. Esto se usa para reemplazar genes o para generar nuevas variantes quiméricas de ellos. Las ligasas usan ATP para fosforilar los extremos 5' de la cadena de DNA, y de esta forma se puede unir dos pedazos de DNA que antes estaban separados.

Las DNA ligasas catalizan la formación de un enlace fosfodiéster entre el 5' de un fosfato y el 3' de un nucleótido. Se usan para unir covalentemente cadenas de DNA. Por ejemplo, un fragmento dentro de otro. La DNA ligasa usada más habitualmente es la T4 DNA ligasa, que requiere ATP como cofactor. La ligación de moléculas de DNA puede realizarse de múltiples maneras (Figura 3-3).

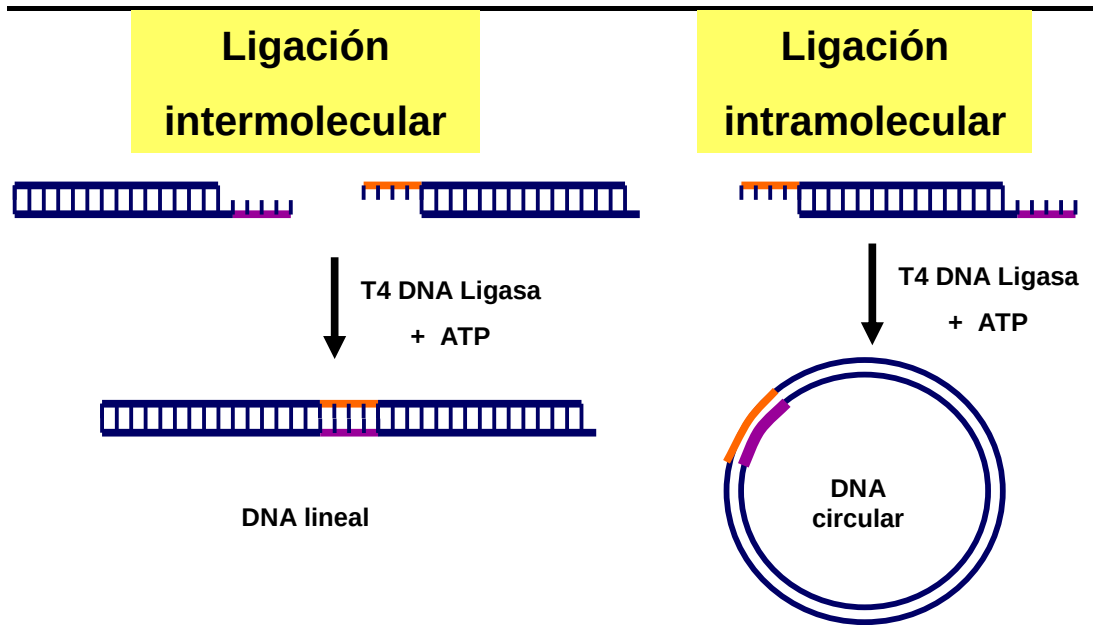


Figura 3-3. Ligación intermolecular entre dos cadenas distintas de DNA. Ligación intramolecular entre los dos extremos de la misma cadena de DNA.

La ligación en principio se producirá o bien entre extremos romos o bien entre extremos de DNAs digeridos con la misma enzima de restricción (*EcoRI*) y que, por tanto, pueden aparearse. Sin embargo, en ocasiones, también pueden producirse entre moléculas digeridas con enzimas diferentes si los extremos que generan son compatibles (*Bam*HI, *Bgl*II, *Bcl*I, *Xho*II).

3.2.3. Recombinasas

La enzima recombinasa cataliza la recombinación homóloga basada en unas secuencias de reconocimiento attB1 y attB2. En los laboratorios de biología molecular mas avanzados, la BP clonasa y la LR clonasa están sustituyendo cada vez más a las enzimas de restricción y a las ligasas, para la clonación de genes y construcción de vectores (Figura 3-4).

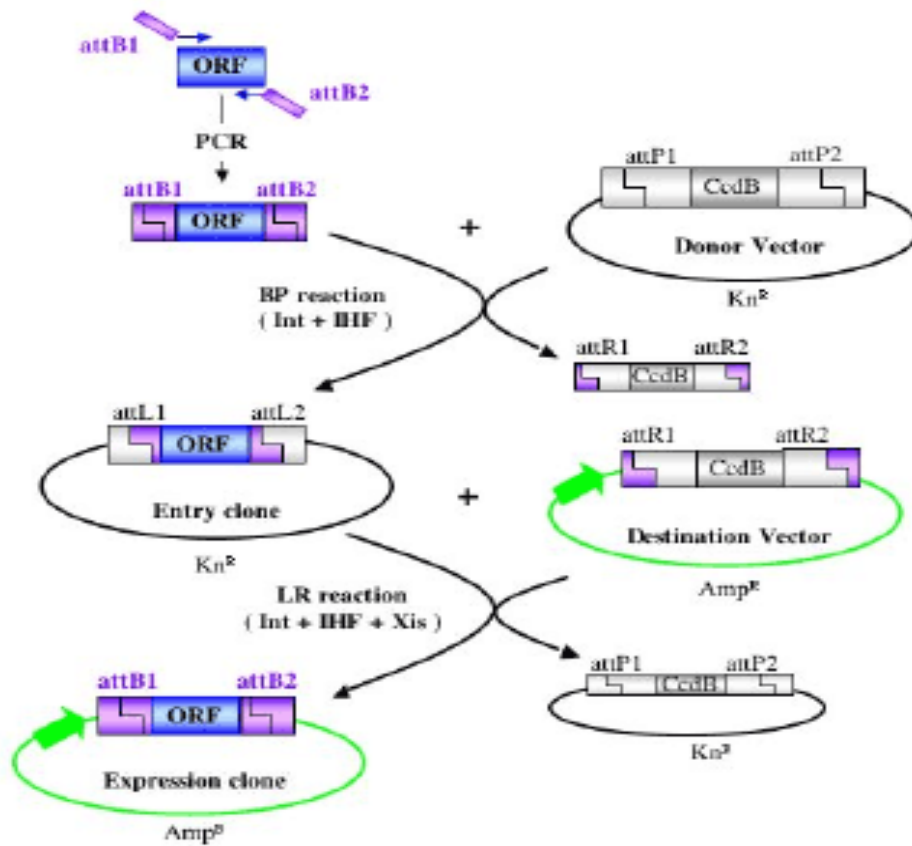


Figura 3-4. Estrategia de clonación por recombinación.

Tarea 3-5:

En internet haga una búsqueda sobre la Recombinasa y la Tecnología Gateway de Invitrogen. Averigüe que es un Vector Gateway y cuáles son los que más se usan hoy en día. Visite la página de Invitrogen para averiguarlo: <http://www.invitrogen.com>

Tarea 3-6:

Descargue el manual pdf de *CloneMiners* de Invitrogen. Averigüe ¿para qué sirve la técnica, cómo funciona, cuales son los pasos a seguir? Prepare una presentación para que lo explique en un seminario a sus compañeros de clases:

<http://tools.invitrogen.com/content/sfs/brochures/710-02224-Tb%20CloneMiner%20.pdf>

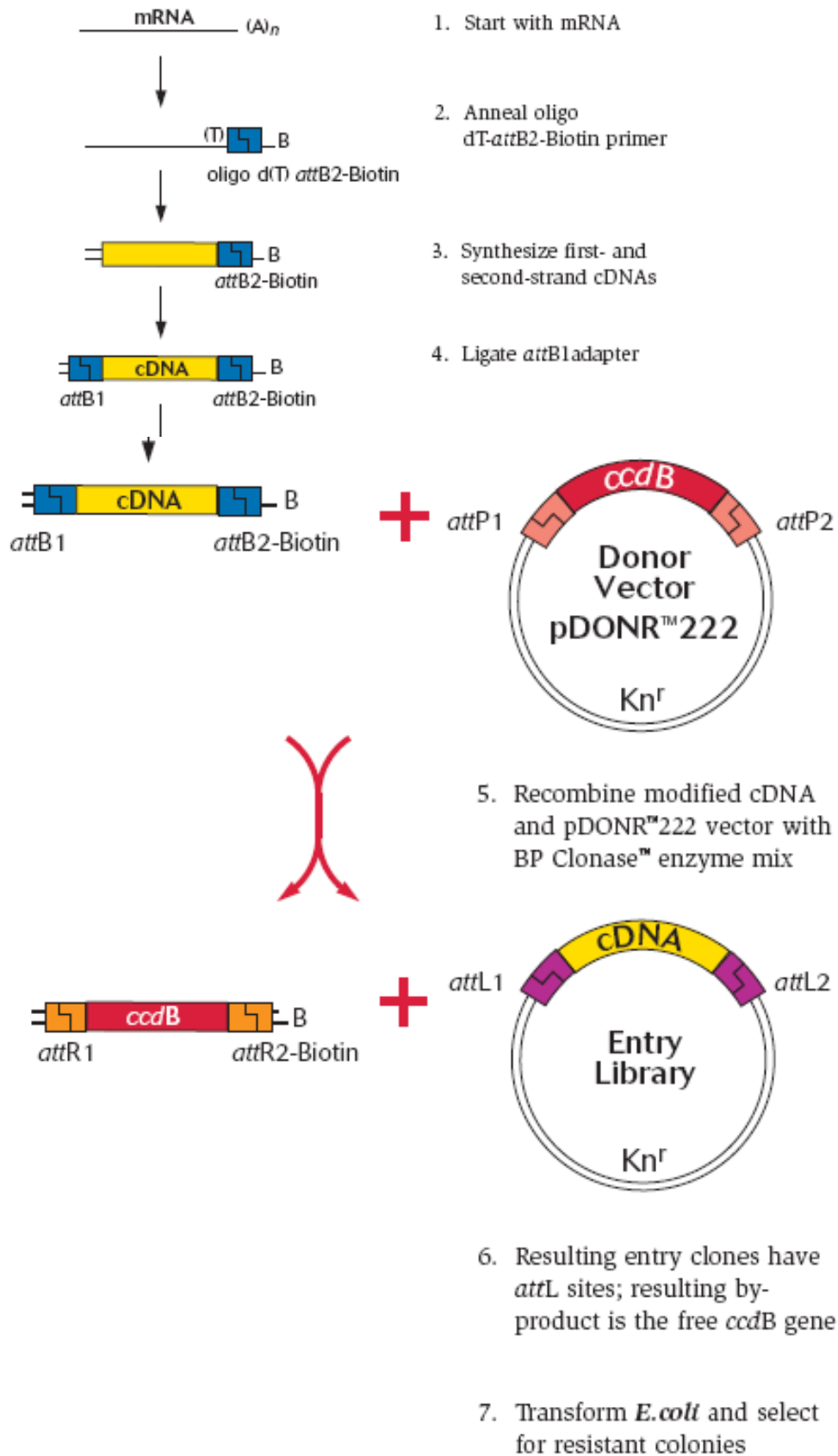


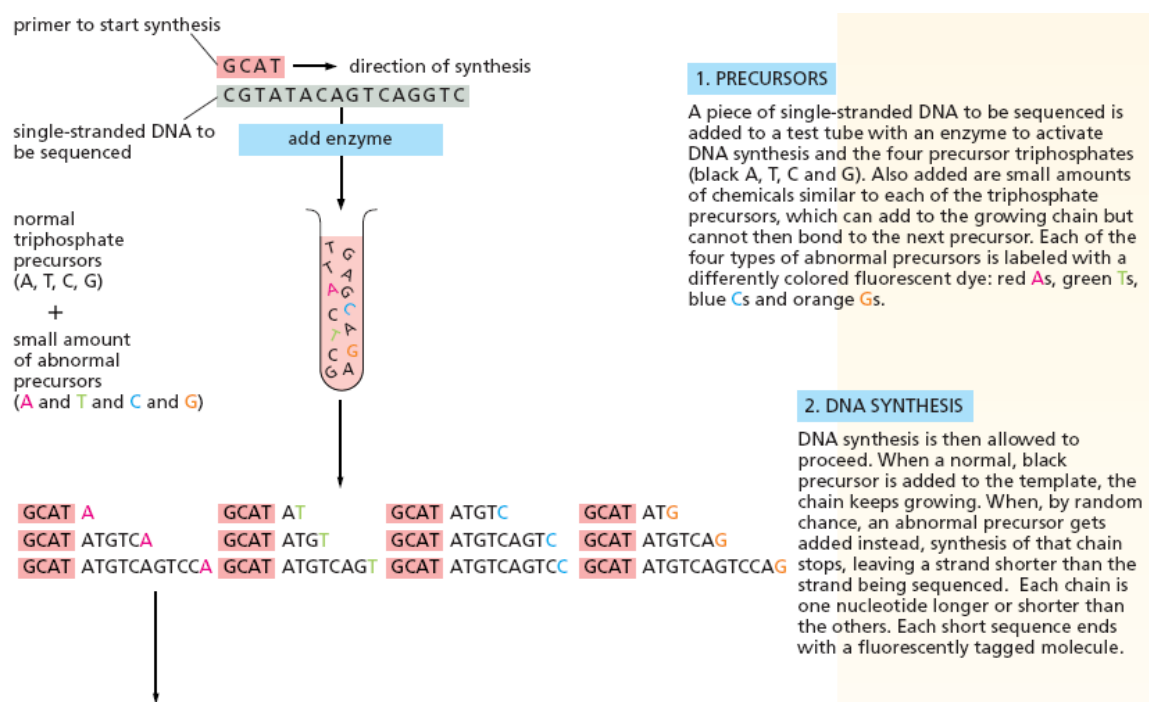
Figura 3-5. Tecnología *Gateway* para subclonación entre diferentes vectores.

3.3. Métodos de secuenciación

Hoy en día existe una gama muy amplia de métodos para determinar la secuencia nucleotídica de los genes. En los últimos cinco años ha habido una gran explosión de plataformas tecnológicas de secuenciación. Estas permiten secuenciar genomas completos en una sola corrida. A continuación presentamos algunos de los métodos clásicos y explicamos los principios en que se basan las tecnologías modernas.

3.3.1. Sanger

El método Sanger moderno se basa en la separación por electroforesis capilar de una sola reacción de PCR con nucleótidos dideox marcados con fluoróforos. Antes se usaba radioactividad, y además se usaban cuatro reacciones separadas en donde se le agregaba solo un nucleótido dideox, el cual terminaba la amplificación y no dejaba que se extendiera al final la cadena con ese nucleótido en específico. Cada una de estas cuatro reacciones se cargaba en un carril por separado y después se corría en geles de acrilamida. Con el tiempo se mejoró la técnica al usar nucleótidos marcados con cuatro fluoróforos diferentes que pueden ser distinguidos por el uso de laser con diferentes longitudes de excitación y la detección con varios filtros de emisión. De esta forma, se puede correr la reacción en un solo tubo capilar, mejorando mucho la resolución y la economía del método (Figura 3-6).



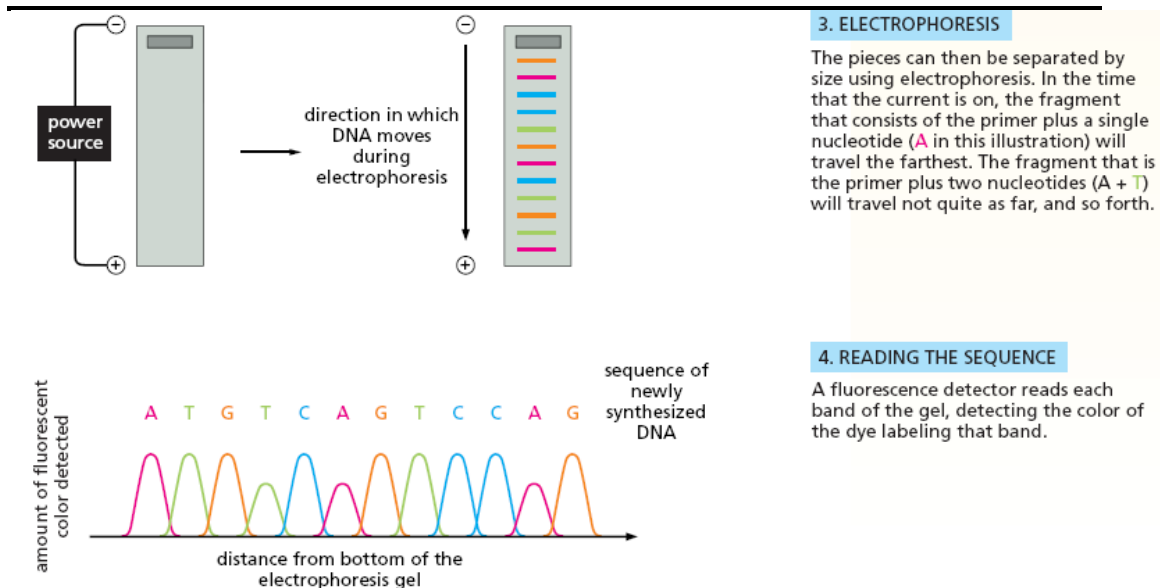
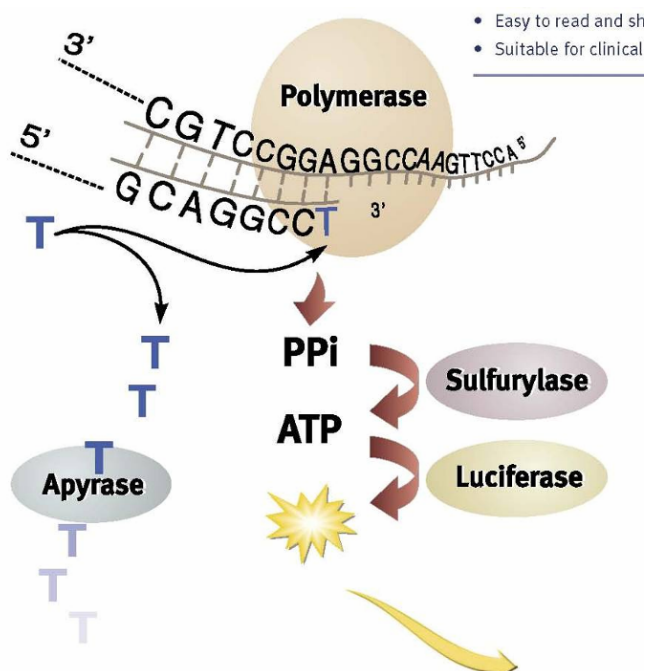


Figura 3-6. Secuenciación por Sanger con nucleótidos fluorescentes.

3.3.2. Pirosecuenciación

Secuenciación en tiempo real más rápido, eficiente y masivo que el método Sanger. El método de pirosecuenciación se basa en la liberación de pirofosfato (PPi) por cada nucleótido incorporado por la polimerasa durante el proceso normal de la síntesis de DNA. La enzima sulfurilasa usa el PPi y un análogo de ADP (adenosina 5'-fosfosulfato o APS) para formar ATP. Este ATP después es usado por la luciferasa para producir fotones. La luz producida es la señal que se detecta con una cámara CCD.



A la pirosecuenciación también se le llama método basado en síntesis en tiempo real, porque la síntesis de DNA y la detección de los nucleótidos se hacen al mismo tiempo (no como en Sanger, que los dos procesos están temporalmente separados). La especificidad en Sanger se basa en el marcaje fluorescente específico de cada nucleótido. Esto permite agregar todos los dNTPs al mismo tiempo durante el PCR. En la pirosecuenciación, el PPi liberado no permite distinguir los nucleótidos, y es por eso que cada dNTP se tiene que agregar por separado (sistema de microfluidos para agregar secuencialmente los nucleótidos. Ciclo repetitivo: dATP, lavado, dGTP, lavado, dCTP, lavado, dTTP, lavado, etc.) .

La pirosecuenciación se lleva a cabo en varios pasos que a continuación se describen en inglés, con el fin de practicar este idioma.

Step 1

A sequencing primer is hybridized to a single stranded, PCR amplified, DNA template, and incubated with the enzymes, DNA polymerase, ATP sulfurylase, luciferase and apyrase, and the substrates, adenosine 5' phosphosulfate (APS) and luciferin.

Step 2

The first of four deoxyribonucleotide triphosphates (dNTP) is added to the reaction. DNA polymerase catalyzes the incorporation of the deoxyribo-nucleotide triphosphate into the DNA strand, if it is complementary to the base in the template strand. Each incorporation event is accompanied by release of pyrophosphate (PPi) in a quantity equimolar to the amount of incorporated nucleotide.

Step 3

ATP sulfurylase quantitatively converts PPi to ATP in the presence of adenosine 5' phosphosulfate (APS). This ATP drives the luciferase mediated conversion of luciferin to oxyluciferin that generates visible light in amounts that are proportional to the amount of ATP. The light produced in the luciferase-catalyzed reaction is detected by a charge coupled device (CCD) camera and seen as a peak in a Pyrogram. The height of each peak (light signal) is proportional to the number of nucleotides incorporated.

Step 4

Apyrase, a nucleotide degrading enzyme, continuously degrades ATP and unincorporated dNTPs. This switches off the light and regenerates the reaction solution. The next dNTP is then added.

Step 5

Addition of dNTPs is performed one at a time. It should be noted that deoxyadenosine alfa-thio triphosphate (dATPaS) is used as a substitute for the natural deoxyadenosine triphosphate (dATP) since it is efficiently used by the DNA polymerase, but not recognized by the luciferase. As the process continues, the complementary DNA strand is built up and the nucleotide sequence is determined from the signal peaks in the Pyrogram.

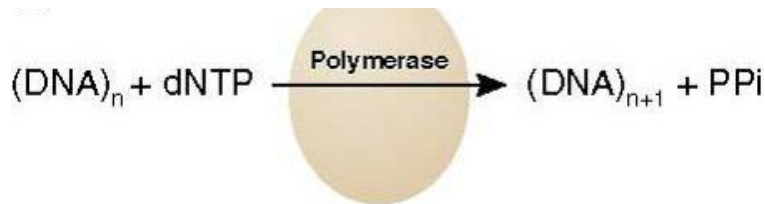


Fig A



Fig B

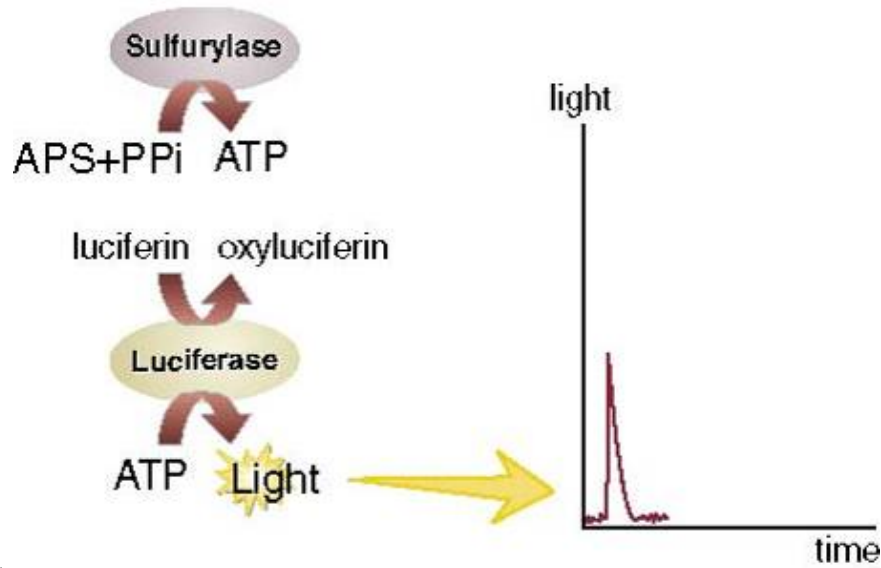


Fig C

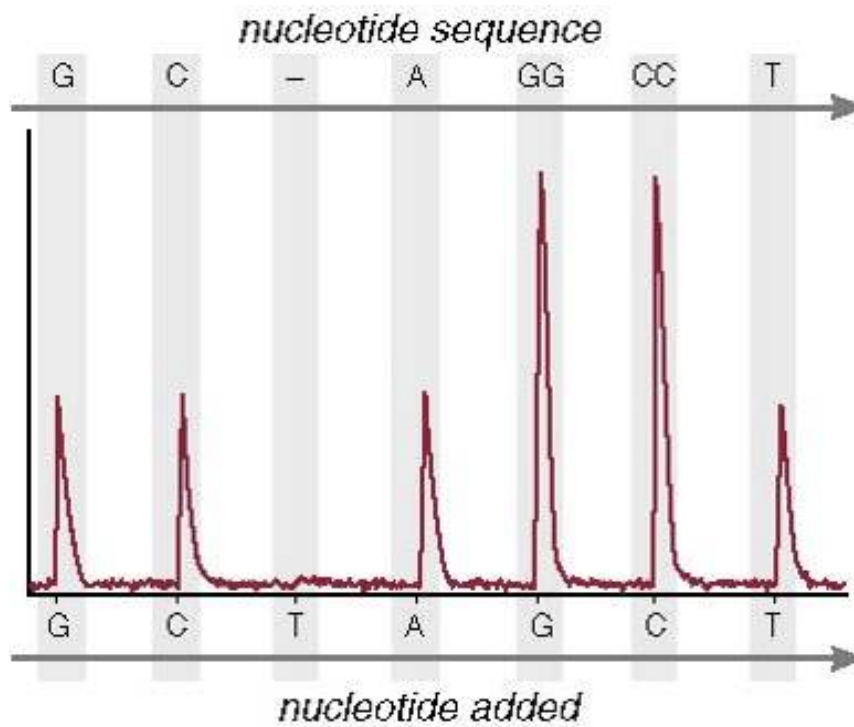


Fig D

Figura 3-7. Etapas de la Pirosecuenciación.**Tarea 3-7:**

Asigne las figuras a las explicaciones anteriores en inglés. De cada panel de figura, ¿a qué paso (step) corresponde?

Figura	Step
A	
B	
C	
D	

Preguntas:

- 1) ¿Para qué se usa la enzima apirasa?
- 2) ¿Qué reacción cataliza la sulfúrilasa?
- 3) ¿Qué reacción cataliza la luciferasa?
- 4) ¿Qué reactivos y enzimas se necesitan para la pirosecuenciación?
- 5) La luciferasa usa ATP como sustrato principal, pero también puede usar dATP como sustrato alternativo. Para secuenciar los nucleótidos G, T y C no hay problema ya que estos trinucleótidos no interfieren con la reacción de la luciferasa. Pero si se agrega dATP, la luciferasa generaría una señal de luz muy alta, que no sería proporcional a la cantidad de PPi que liberó la polimerasa. ¿Explique cómo se solucionó este problema de interferencia para que la técnica de secuenciación funcionara sin problemas?

Respuestas:

A2, B4, C3, D5

- 1) La apirasa degrada los deoxinucleótidos que no se han incorporado y acelera el proceso de lavado entre cada ciclo de diferentes nucleótidos.
- 2) La sulfúrilasa usa el PPi para convertir el APS en ATP.
- 3) La luciferasa usa el ATP para emitir luz. La cantidad de luz emitida refleja la cantidad de ATP presente, y esto es proporcional a la cantidad de PPi que había antes, que a su vez es proporcional a la cantidad de nucleótido que la polimerasa incorporó.
- 4) Necesitamos 4 enzimas: DNA Polimerasa, ATP Sulfúrilasa, Luciferasa y Apirasa. Se necesitan los sustratos APS y Luciferina. también se deben de agregar los dNTPs que a su vez liberaran el PPi que necesita la sulfúrilasa.
- 5) Se escogió entonces un sustrato análogo al dATP que la polimerasa si pudiera usar para liberar PPi, pero que la luciferasa no pudiera usar para generar luz. La luciferasa no puede usar desoxiadenosina alfa-tio trifosfato (dATPaS), pero la polimerasa afortunadamente si.

Allele 1: nnnnnnA C T G

Allele 2: nnnnnnG C T G

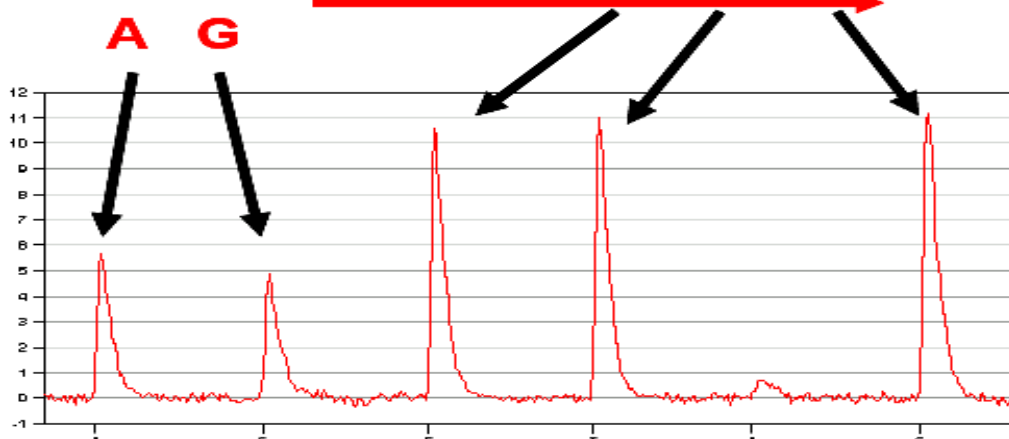


Figura 3-8. Pirograma mostrando el efecto de un SNP en la intensidad de los picos de fluorescencia en los ciclos secuenciales de la pirosecuenciación.

Pirosecuenciación masiva “Genome sequencer GS20”:

La técnica de pirosecuenciación aprovecha los avances de la nanotecnología, al utilizar esferas atrapadas en una placa que contiene 1,600,000 microceldas de 44 µm de diámetro, como se explica arriba, de tal suerte que al ingresar una esfera por cada celdilla, permite realizar una

reacción de secuenciación individual, basada en la síntesis complementaria de un DNA de cadena sencilla, el cual ha sido amplificado por una reacción de PCR y alineado a cada esfera. La incorporación de cada nucleótido durante la polimerización de DNA, libera una molécula de PPi, que al ser metabolizada por las enzimas presentes en el medio (luciferasa y APS entre otras) produce una señal quimioluminiscente que es capturada por una cámara de detección de fotones y que con el soporte de un poderoso programa de cómputo, es traducida como la adición de un nucleótido determinado, generando una secuencia individual por cada celda.

Es una tecnología patentada que ahora distribuye la compañía Roche®. En lugar de secuenciar hebras individuales, el método de luminiscencia permite detectar más de trescientas mil reacciones de secuenciación al mismo tiempo. Esto se hace en nanoesferas dentro de una nanorejilla hexagonal. El punto crítico es que se hace PCR de un trozo pequeño de DNA dentro de una pequeña gota. A esto se le llama PCR de emulsión o ePCR por sus siglas en inglés. Previamente se debe romper el DNA en trozos pequeños (técnica de nebulización), ligarlo a adaptadores, desnaturalizarlo y unirlo a nanoesferas, de forma que cada nanoesfera se encuentre con un fragmento.

¿Para qué sirve la PCR de emulsión? Las microgotas de emulsión capturan la nanoesfera junto con la mezcla de enzimas necesarias para hacer PCR. Cada nanoesfera tiene solo una secuencia de DNA. Es decir, en cada microgota hay una sola molécula molde, por lo que la amplificación creará muchas copias de solamente esa molécula. Se dice entonces que es una amplificación clonal.

Seguidamente se distribuyen las nanoesferas en los pozos de una nanoplaca. Dentro del tubo estarán los reactivos de pirosecuenciación y se añaden los nucleótidos. En cinco horas se pueden leer ~300 mil secuencias de 100-250 nucleótidos. En una corrida de pirosecuenciación obtendremos información de 20-50 millones de nucleótidos. Estas secuencias cortas se tendrán que ensamblar. Para ello se requiere de supercomputadoras con una capacidad muy grande de memoria (>50 GB de RAM). Por lo regular se usan clusters computacionales de varias docenas de nodos, cada uno con un procesador independiente. El cluster reparte el trabajo a los diferentes nodos de forma paralela, y así, se reduce el tiempo requerido para hacer los cálculos de ensamblaje.

La tecnología de pirosecuenciación fue desarrollada por la empresa *454 Life Sciences*, la cual comercializó el primer equipo de secuenciación masiva (Secuenciador GS 20), con una capacidad para descifrar 20 millones de bases por corrida (4 h), en secuencias individuales de 100 bases. Tan sólo tres años después la capacidad de secuenciación se ha ampliado para fragmentos de 400 bases.

En CINVESTAV Irapuato exploramos continuamente el potencial de las nuevas tecnologías de secuenciación. Actualmente se usan cuatro sistemas diferentes, el *Sanger*, el *454*, el *Solid*, y el *Ion Torrent*, ya que cada método tiene ciertas ventajas según el uso que se le quiera dar a los datos. La pirosecuenciación se ha utilizado con genomas bacterianos, pero también con genomas eucarióticos que son mucho más complejos. Se ha descifrado un genoma bacteriano, perteneciente al género *Bacillus*, el cual fue aislado del sistema acuático de las fosas de Cuatro Ciénegas. Con seis corridas (Secuenciador GS 20), se descifraron 129.6 Mb de secuencia. Esto corresponde a una cobertura mayor a 20x de su genoma. También se trabaja en la pirosecuenciación de una variedad de maíz criolla, el palomero conico toluqueño, usando bancos genómicos de DNA total y bancos genómicos de DNA enriquecido en regiones codificantes. A estas regiones se les llama MAGIs (*Maize Assembled Gene Islands*). Para ello, se utilizan estrategias para el enriquecimiento de regiones de DNA no metiladas. Los procesos de ensamblado y de anotación se tratan de realizar de manera paralela, pero estas representan el cuello de botella, ya que tardan mucho más tiempo que las propias corridas de secuenciación.

Tarea 3-8

Revise estas referencias bibliográficas y haga un resumen de su contenido.

Schaefer, B. 1995. Revolutions in Rapid Amplification of cDNA Ends: New Strategies for Polymerase Chain Reaction Cloning of Full-Length cDNA Ends. *Anal Biochem.* 227, 255-273.

Jordan, B.; DNA Microarrays: Gene Expression Applications; Springer-Verlag; 2001 y Schena, M.; Microarray Biochip Technology; Eaton Publishing Company/Bio Techniques Books Division; 2000.

3.3.3. Método de apareamiento y ligación

En esta sección explicaremos el sistema de secuenciación usando la plataforma SOLiD de *Applied Biosystems*. El sistema de SOLiD es un equipo de secuenciación masiva de última generación. Es aún más avanzado que el método de pirosecuenciación descrito en la sección anterior. La Figura 3-9 muestra el equipo que está instalado en CINVESTAV Irapuato. A diferencia de los demás métodos de secuenciación, el sistema SOLiD puede generar muchos millones de secuencias. La limitante es que son secuencias muy cortas, de sólo 50 bases.

Tabla comparativa de métodos de secuenciación:

Método	Longitud de lectura de cada secuencia	Número de secuencias por cada corrida	Cantidad de información nucleotídica por corrida
Sanger	800-1500 bases	1-384	576 kb
Pirosecuenciación 454	100-500 bases	250 mil	20 Mb
SOLiD	50 bases	30 millones	4 Gb



Figura 3-9. Equipo de SOLiD de Applied Biosystem.

Los procesos necesarios de la metodología SOLiD se dividen en varias etapas (ver Figura 3-10):

- La preparación de la muestra es dependiente de la aplicación. Se puede iniciar a partir de DNA o de RNA.
- La librería puede hacerse ya sea de fragmentos cortos (*fragment library*) o de fragmentos largos con extremos apareados (*mate-paired library*), (ver Figura 3-11).

- La librería se tiene que amplificar en microesferas usando PCR en emulsión.
- Las clonas de cada microesfera se depositan de manera aleatoria sobre la superficie de una laminilla de vidrio. Ahí se inmovilizan por medio de enlaces covalentes. La laminilla de vidrio se inserta en la máquina donde se procesa por un sistema robótico de microfluidos (ver Figura 3-9).
- Las clonas se secuencian por la técnica de ligación (ver Figura 3-13). Esto consiste en múltiples ciclos de apareamiento, ligación, lavado, detección, corte y lavado.
- Con eso se obtienen los datos primarios que tienen que ser interpretados para obtener la secuencia. Para ello se usa una codificación de cuatro colores para cada dos bases (dibases), (ver Figura 3-14.). El análisis de datos posterior es específico para cada tipo de aplicación. Existen diversos programas bioinformáticos propios de la empresa y también algunos de uso libre. Para mas detalles visite la página: <http://solid.appliedbiosystems.com>

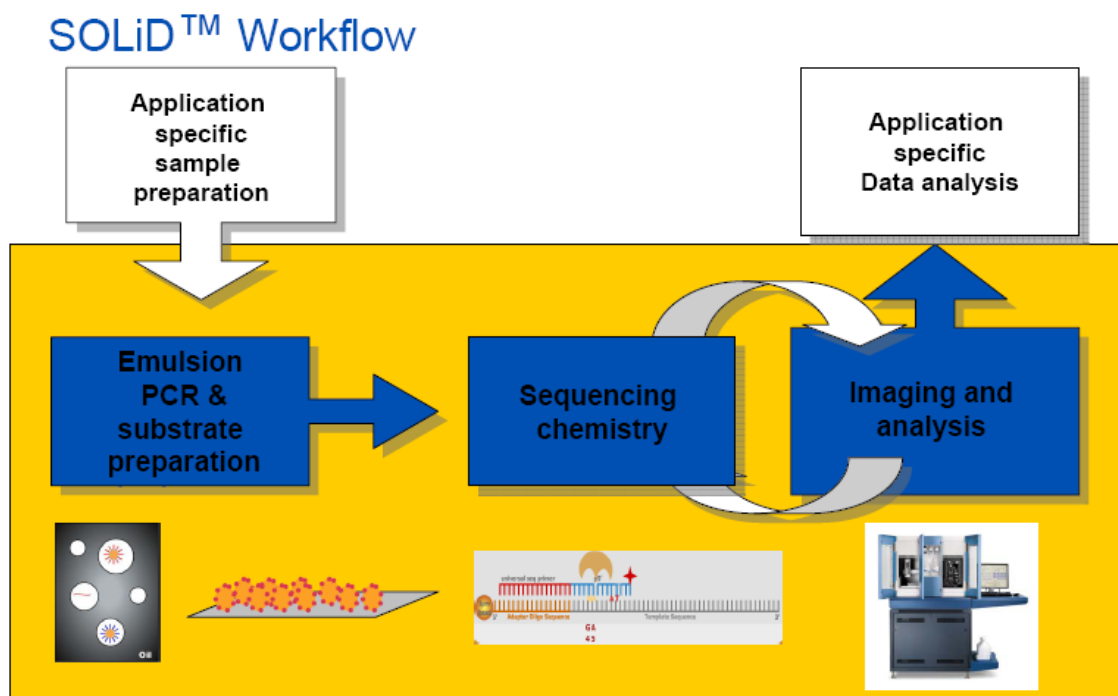


Figura 3-10. Etapas secuenciales de la plataforma SOLiD. Tanto la preparación inicial de la muestra como el análisis de datos son específicos para cada tipo de aplicación. En cambio, la parte central del proceso (enmarcada en amarillo) es común y universal para todas las aplicaciones.

Se pueden hacer dos tipos de librerías de fragmentos de DNA. Cada tipo de librería se puede usar de manera óptima para diversas aplicaciones.

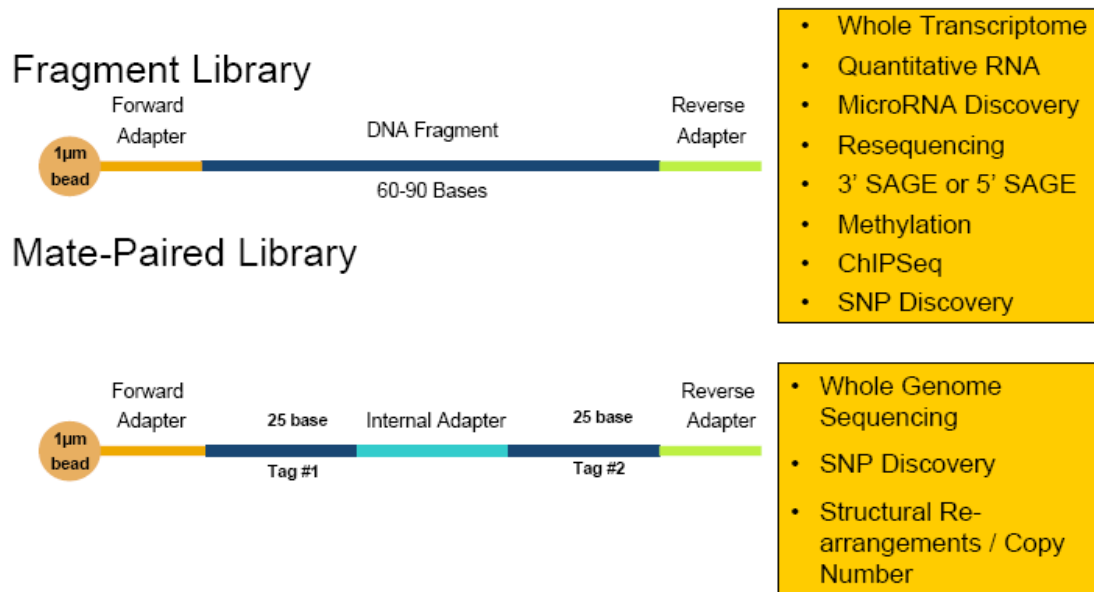


Figura 3-11. Tipos de librerías que pueden ser secuenciadas por el sistema SOLiD

Principio de secuenciación por ligación:

La secuenciación se lleva a cabo en ciclos repetitivos cada uno de varias etapas:

0. Adición de primer universal de secuenciación.
1. Adición de sondas fluorescentes y enzima ligasa. Apareamiento de sondas. Formación de unión covalente por ligación. Lavado de sondas no unidas.
2. Detección del fluoróforo con excitación láser.
3. Desfosforilación de hebras no extendidas por el *primer*.
4. Corte del fluoróforo. Se quitan tres nucleótidos unidos al fluorocromo dejando tres nucleótidos *n* con un fosfato libre en el 5' *prima*. Lavado de agente cortante.
5. Repetición de etapas 1-4 para extender la secuencia. La extensión se hace de cinco en cinco bases (dos bases detectoras y tres bases *n*).
6. Reseteo de primer de secuenciación. Desnaturalización y lavado.
7. Repetición de etapas 1-5 con el nuevo primer que está desplazado por una base.
8. Se hacen en total cinco rondas de secuenciación (etapas 0-6) con los respectivos *primers* de inicio universales *n*, *n-1*, *n-2*, *n-3* y *n-4*.

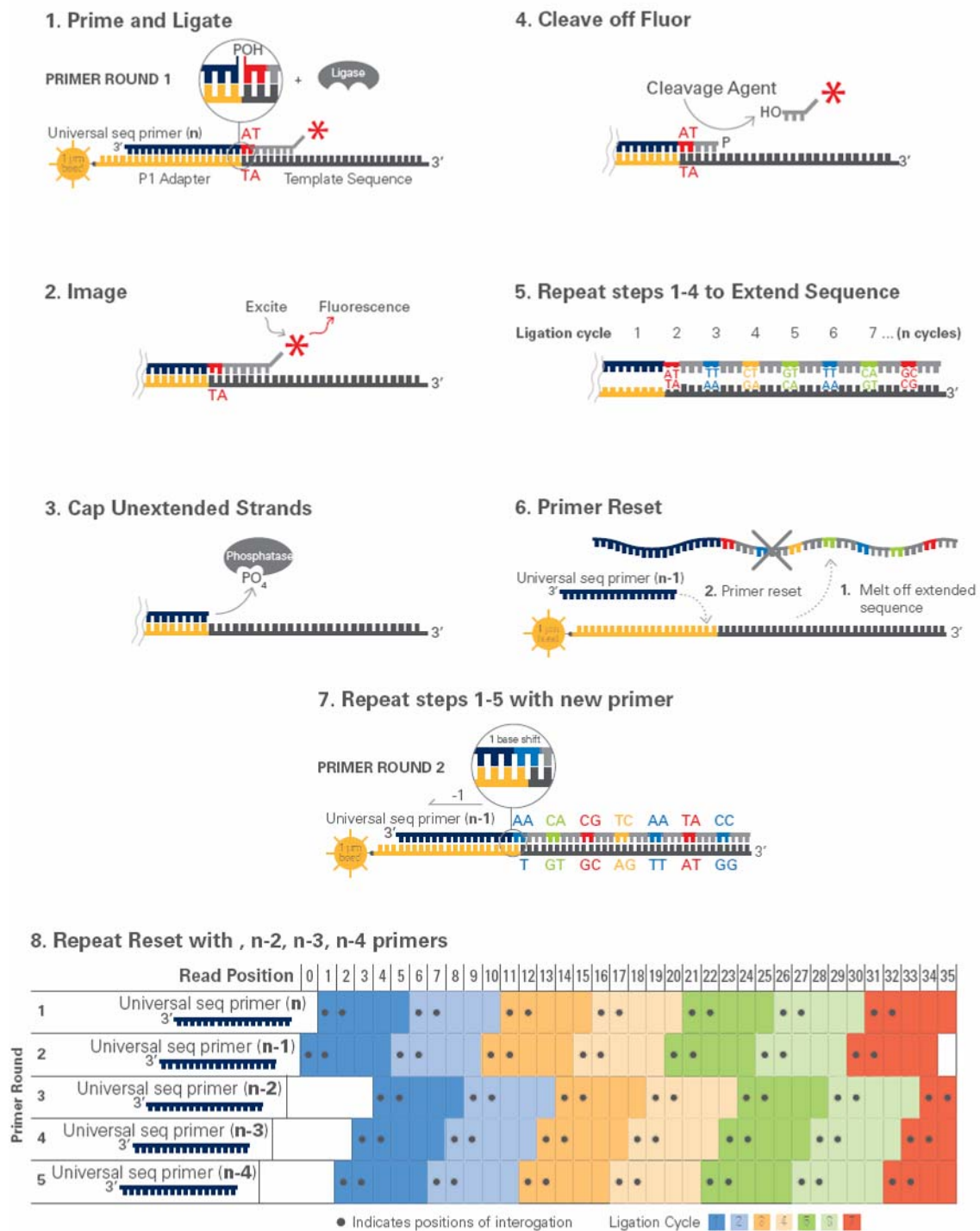


Figura 3-12. Principio de secuenciación por ligación. Etapas cíclicas dentro del sistema robótico.

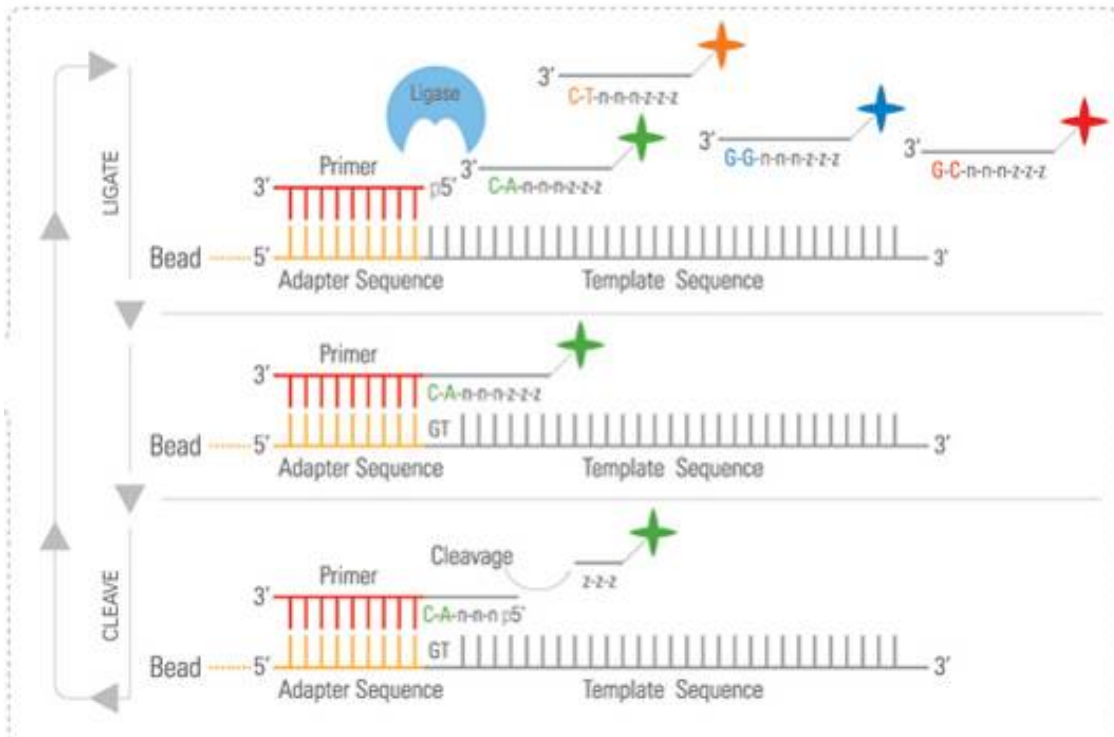


Figura 3-13. Principio de secuenciación usando la enzima ligasa y detección con el sistema SOLiD.

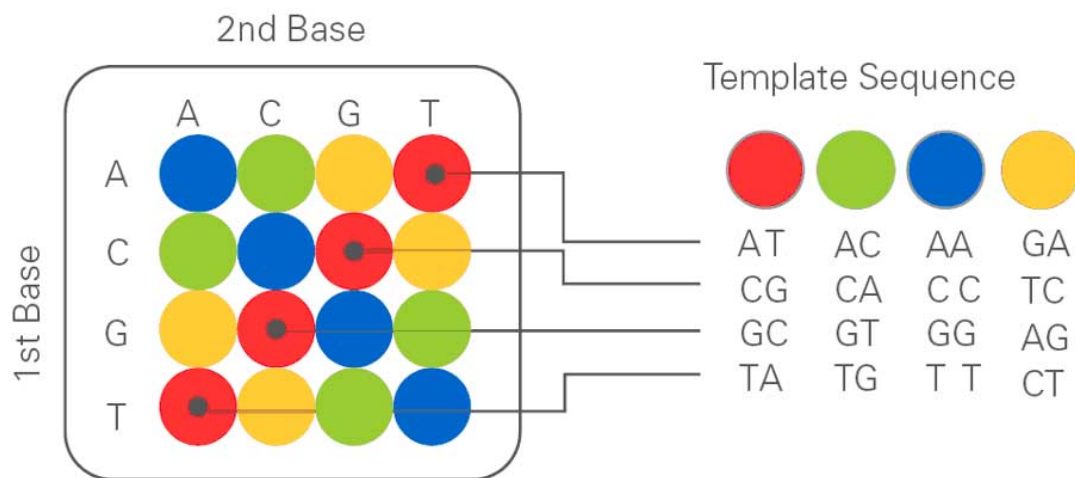


Figura 3-14. Principio de codificación de cuatro colores para cada dos bases (dibase).

Existen 16 posibles combinaciones de dibases nucleotídicas. Si sólo se usan cuatro fluoróforos diferentes, entonces sólo se pueden reconocer cuatro grupos de dibases. Por ejemplo, el color rojo puede ser para las dibases AT, CG, GC o TA. No podemos decir para cual de esas cuatro dibases corresponde el color rojo. Lo anterior indica, que una secuencia de cuatro colores no es suficiente para decidir cual es la secuencia exacta de nucleótidos. Habrá cuatro opciones de secuencias posibles con esos colores. Se necesita información adicional para distinguir entre las cuatro

versiones posibles. La siguiente figura muestra un ejemplo de una secuencia de color y sus cuatro posibles versiones nucleotídicas:

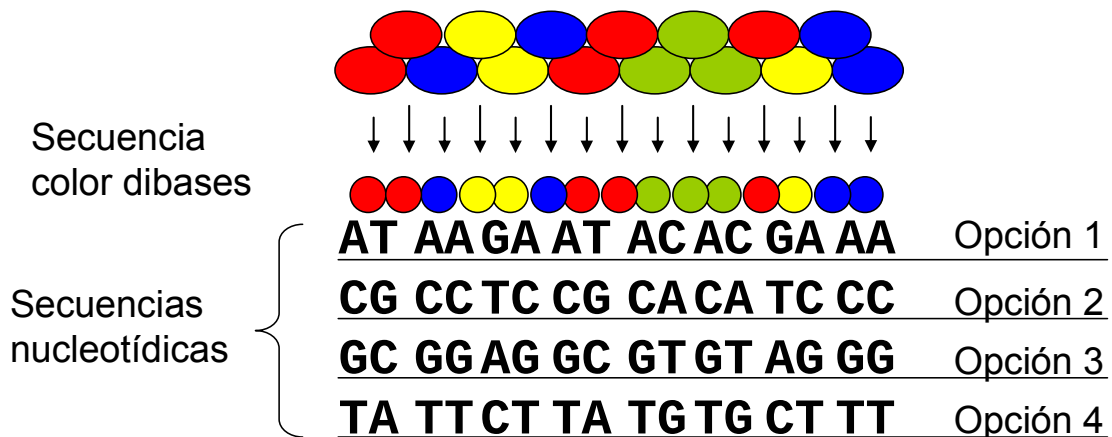


Figura 3-15. Secuencia en código de colores y sus posibles versiones nucleotídicas.

Por ejemplo: si de la secuencia de color mostrada en la Figura 3-15 se sabe que la primera base es A, entonces eso nos indica que el rojo es para la dibase AT y no para las otras opciones. Por consiguiente, la siguiente dibase que es azul, sólo puede ser para AA, y la siguiente que es amarilla, es para GA y así sucesivamente (opción 1). Es decir, con la secuencia de colores y el conocimiento de al menos una base de esa secuencia, podemos traducir el código de colores (sistema de dibase) al código nucleotídico (sistema de monobase). Así es como el sistema SOLiD convierte un código de 16 opciones a un código de cuatro opciones.

En la práctica, lo que se hace es alinear la secuencia de colores con una secuencia referencia, de forma que la secuencia referencia nos ayuda a definir las opciones nucleotídicas posibles. Una vez alineada con una referencia dada, se puede distinguir muy bien si hay algunas mutaciones como deleciones, inserciones o SNPs con respecto a la referencia. Esto se logra con una altísima precisión, ya que el sistema de dibase puede distinguir entre mutaciones reales y errores técnicos de secuenciación (ver Figura 3-16). Esta precisión se da por la característica de que cada base está representada por dos colores contiguos, de forma que para que un SNP sea real debe de haber un cambio en dos colores contiguos. Si en cambio, sólo se detecta un solo cambio de color, entonces indica que fue un error de secuenciación.

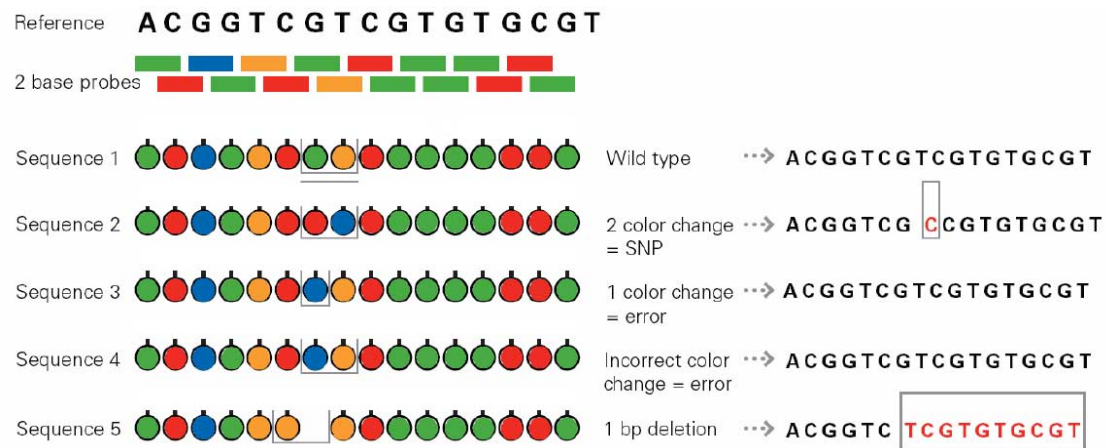


Figura 3-16. Opciones de detección de errores falsos o de confirmación de cambios reales, por ejemplo SNPs, delecciones o inserciones de una base.

Es por ello que el sistema de SOLiD es ideal para resecuenciar genomas ya conocidos en búsqueda de polimorfismos. Para lo que es menos indicado, es para secuenciar *de novo* cuando no existe suficiente información de referencia. La dificultad radica en que las secuencias cortas (<50 bases), que además tienen ambigüedad por sus colores dibases (para cada secuencia de color existen 4 secuencias nucleotídicas posibles). Todo esto hace mucho mas difícil el ensamblado *de novo* de un genoma eucariótico desconocido.

Toda esta traducción complicada entre diferentes códigos se podría evitar si se pudieran usar 16 fluoróforos diferentes para las 16 dibases posibles. Sin embargo, actualmente no existen tantas moléculas fluoróforas que tengan un espectro de absorción y de emisión lo suficientemente diferente para distinguir 16 colores diferentes. Tal vez, en un futuro próximo, si las haya y de esta forma, se pueda mejorar aún más los sistemas tipo SOLiD.

Tarea 3-9

Lea el siguiente texto en inglés y conteste las siguientes preguntas:

¿Qué combinación de características son las que permiten la alta confiabilidad de los datos generados con este sistema?

Ligase enzymology provides significantly higher selectivity than polymerase-based approaches and dramatically reduces errors associated with mis-incorporated bases. Two base encoding interrogates each base twice during sequencing allowing for rare SNP detection with extremely high confidence. The built-in primer reset function automatically caps system noise at given intervals to yield lower background and error rates compared to polymerase-based chemistries. A mate-paired technology that allows for a tight distribution across a range of insert sizes enables the accurate analysis and assembly of complex genomes. It is ideal to utilize larger insert sizes for detection and fine resolution of *de novo* genomic aberrations, such as inversions, translocations and duplications. Smaller insert sizes do not allow the same level of resolution as larger insert sizes.

Tarea 3-10

Lea el siguiente texto en inglés y conteste las siguientes preguntas:

¿Cuáles son las ventajas del sistema SOLiD? El sistema de SOLiD ¿para qué etapa de secuenciación es adecuado de manera única? ¿Cuánta información genera cada corrida? ¿Qué tan confiables son los datos generados con este sistema?

De novo sequence analysis has traditionally been conducted by shotgun approaches using automated capillary electrophoresis (CE) systems. When the initial sequence scaffold for a particular genome is available, it is then possible to perform deep sequencing to fill in the gaps or to identify polymorphisms, mutations, and structural variations between organisms. Large scale resequencing projects require a highly parallel system to provide the depth of coverage required for variation detection. A highly accurate system is also critical to reduce false positive rates and provide the sensitivity necessary to detect mutations in heterogeneous samples. With increased sensitivity, the SOLiD System enables pooling of samples, as well as the detection of minor alleles.

Automated CE sequencing remains the gold standard for de novo analysis in which long read lengths are desired, and for smaller targeted resequencing experiments. The Applied Biosystems SOLiD™ System for next generation sequencing is uniquely suited for finishing or large scale resequencing projects. The system generates up to 4 gigabases of mappable data in a single run (dual slide), providing the coverage and statistical resolving power required to correctly identify low frequency SNPs and structural variations across a large number of amplicons. The combination of 2-base encoding, high fidelity ligase, and primer reset functionality result in an overall accuracy of 99.94%, and a consensus accuracy at 15x coverage of 99.9999%.

Tarea 3-11

Lea el siguiente texto en inglés y después, describa con sus propias palabras en español para qué aplicaciones se puede usar el sistema SOLiD.

Targeted Resequencing

The high accuracy inherent in the SOLiD System is required for targeted resequencing.

The SOLiD System technology is amenable to region-specific pullout methods and barcoding, that when coupled with alignment against a known reference sequence, will yield a very flexible format for region-specific resequencing.

Gene Expression

The massive amount of sequence tags generated in a single run can be used to identify and quantify exon sequences associated with alternative splicing and promoter usage. Simple tag counting allows for digital readout of expression levels and the ability to detect lowly expressed genes with greater sensitivity.

MicroRNA Discovery

MicroRNAs (miRNAs) are highly conserved, small, RNA molecules encoded in the genomes of plants and animals. The SOLiD System technology will

enable the identification and quantification of large numbers of small RNAs from diverse species.

Chromatin Immuno-precipitation (ChIP)

The vast number of sequencing reads generated by the system coupled with high accuracy and matepaired technology enable a whole new way to study ChIP with unmatched specificity and accuracy.

Whole Genome Sequencing

The SOLiD System enables a new level of whole genome sequencing with massive throughput and mate-paired technology. The system's throughput allows for high coverage levels and the ability to sequence multiple genomes in a single run. These and other features dramatically reduce the time and cost required by other technologies.

Tarea 3-12

Lea las siguientes explicaciones en inglés, y elabore su propio diagrama de flujo con sus respectivas explicaciones en español.

A. Library Preparation

The SOLiD System can use two types of libraries—fragment or mate-paired. The type of library used depends on the application and information needed.

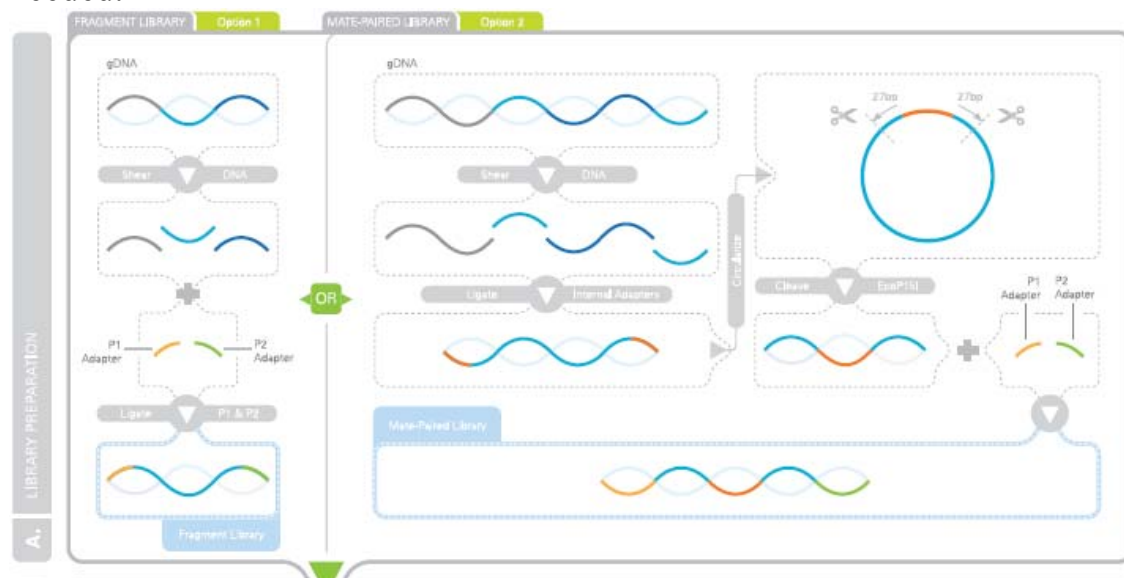


Figura 3-17. Preparación de librerías para el sistema SOLiD.

B. Emulsion PCR/Bead Enrichment

Clonal bead populations are prepared in microreactors containing template, PCR reaction components, beads and primers. After PCR, the templates are denatured and bead enrichment is performed to separate beads with extended templates from undesired beads. The template on the selected beads undergoes a 3' modification to allow covalent bonding to the slide.

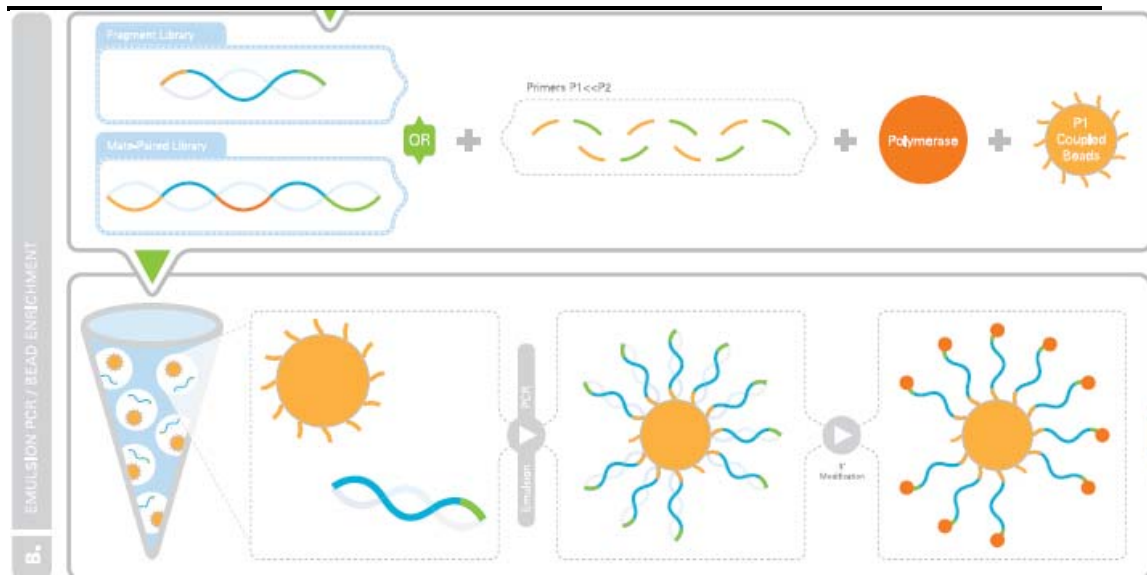


Figura 3-18. PCR de emulsión para el sistema SOLiD.

C. Bead deposition

3' modified beads are deposited onto a glass slide. Deposition chambers offer the ability to segment a slide into one, four or eight chambers during the bead loading process. A key advantage of the system is the ability to accommodate increasing densities of beads per slide, which will result in a higher level of throughput from the same system.

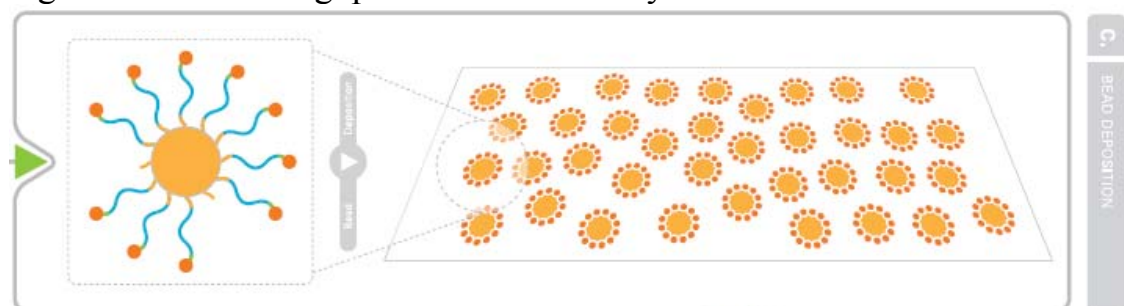


Figura 3-19. Fijación al azar y con alta densidad de bolitas con DNA amplificado en una laminilla de vidrio.

D. Sequencing by Ligation / Data Analysis

Primers hybridize to the P1 adapter sequence within the library template. A set of four color dye-labeled probes compete for ligation to the sequencing primer. Specificity of probe ligation is achieved by interrogating every 4th and 5th base during the ligation series. Five to seven rounds of ligation, detection and cleavage record the color at every 5th position with the number of rounds determined by the type of library used. Following each round of ligation, a new complementary primer offset by one base in the 5' direction is laid down for another series of ligations. Primer reset and ligation rounds (5–7 ligation cycles per round) are repeated sequentially five times to generate 25–35 base pairs of sequence for a single tag. With mate-paired sequencing, this process is repeated for the second tag. The use of color space for two base encoding makes it easy for the system to

quickly and accurately discriminate between system errors and true polymorphisms.

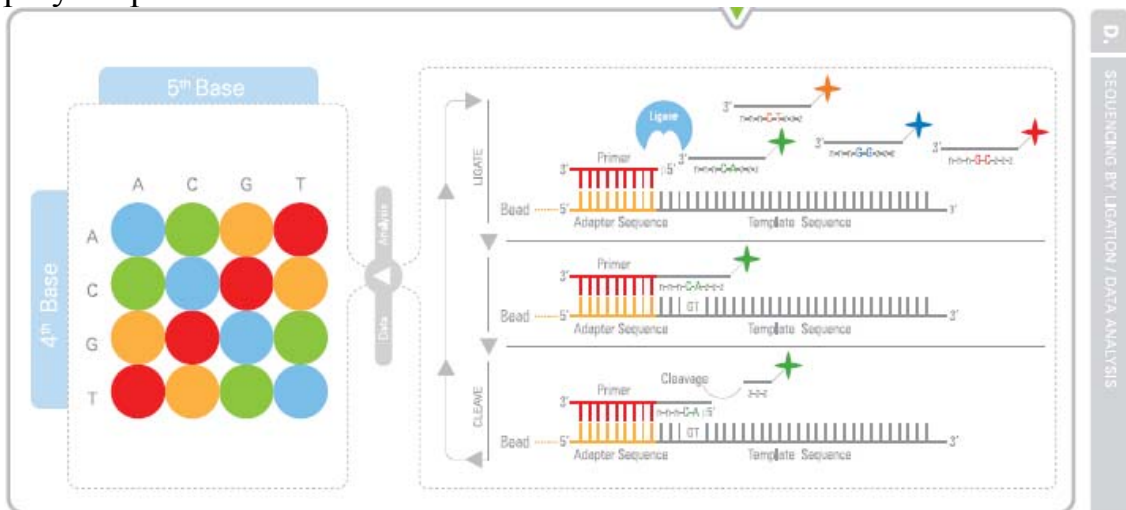


Figura 3-20. Sistema de codificación de colores para el SOLiD. Ciclos repetidos de unión, ligación y detección.

3.3.4. Método de síntesis con fluoróforos

Tarea 3-13.

Consulte las paginas web de **Illumina** para aprender más sobre la tecnología que implementa esa compañía. Trate de explicar el sistema de secuenciación *Solexa* con sus propias palabras.

<http://www.illumina.com/pages.ilmn?ID=203>

http://www.illumina.com/downloads/SS_DNAsequencing.pdf

4. Detección de polimorfismos genéticos

4.1. Marcadores

4.1.1. Características de los marcadores

A los polimorfismos genéticos también se les llama **marcadores moleculares** por la posibilidad de distinguir una variante de DNA de otras. En el caso ideal, un buen marcador tiene las siguientes características:

- **Inferencia no destructiva:** Permite analizar al individuo sin perder su integridad o descendencia. El DNA se puede extraer de una muestra o de un tejido sin matar al organismo.
- **Es frecuente:** tiene que ser abundante en el genoma para que aparezcan suficientes marcadores a lo largo de todos los cromosomas sin permitir muchos huecos.
- **Es polimórfico:** Debe tener como mínimo 2 variantes para que nos genere una información útil. De nada sirve analizar un marcador que es igual en toda la población.
- **Presenta herencia mendeliana:** Está basado en una secuencia de DNA específica y se hereda por medio de los cromosomas del núcleo. (no tiene influencia epigenética)
- **Es de carácter codominante:** Para distinguir entre el homocigoto y el heterocigoto. Los marcadores moleculares puramente dominantes no permiten identificar a los heterocigotos. Una de las desventajas de los marcadores fenotípicos es que no detectan a los heterocigotos.
- **Requiere poca cantidad de DNA:** Esto es importante, sobretodo cuando las muestras son pequeñas (por ejemplo a partir de semillas). La cantidad de DNA de buena calidad es limitante para los RFLPs. Los microsatélites requieren de muchos menos DNA para funcionar.
- **Fácilmente analizables, interpretables e intercambiables entre laboratorios:** Para evitar resultados diferentes. Una electroforesis con errores de 5 pb puede dar resultados poco reproducibles y erróneos. Además se pueden sumar errores de la PCR, sobretodo si se usan muchos ciclos. Al copiar fragmentos repetitivos la Taq polimerasa puede tener problemas. Es necesario hacer intercambios entre laboratorios y muestras control con el fin de validar los resultados de microsatélites.
- **Automatizables:** Los RFLPs, AFLPs y microsatélites requieren de nuestra interpretación manual, ya que todavía no existe un software que haga la evaluación computarizada. Los SNPs sí son

automatizables, sobre todo si se usan microarreglos (por ejemplo, la tecnología de Illumina).

4.1.2. Tipos de marcadores genéticos

Marcadores Tipo I (basados en caracteres expresados)

Antes se usaban sólo estos porque se pueden detectar fácilmente (color de semilla, forma, tamaño, etc.). Por lo general se trataba de marcadores monogénicos dominante-recesivos para características fenotípicas muy simples, por ejemplo: el color y forma de los chícharos de Gregor Méndel o el color de los ojos y la forma de las alas de las mosquitas de Thomas Morgan.

- Marcadores fenotípicos: forma de la hoja, color de la flor, semilla, presencia de lígulas, tricomas, entre otros.
- Marcadores bioquímicos: carotenos, antocianinas, clorofila, almidón tipo waxy, etc.

Marcadores Tipo II (basados en proteínas)

Patrón de Proteínas

Haciendo electroforesis SDS-PAGE de proteínas se pueden detectar diferencias entre genotipos (polimorfismos). No es tan eficaz ya que en los geles de una dimensión sólo se pueden diferenciar relativamente pocas proteínas por tamaño y abundancia. Los extractos totales de proteínas no permiten detectar muchos cambios. Solamente los cambios mayores del tipo ausencia-presencia o modificación sustancial en el peso molecular. Se ha propuesto el uso de MALDI-TOF para detectar cambios en patrones proteínas, pero en la práctica se usó poco para mejoramiento genético.

Isoenzimas

Las isoenzimas o isozimas son enzimas que difieren en la secuencia de aminoácidos, pero que catalizan la misma reacción química. Estas enzimas suelen mostrar diferentes parámetros cinéticos (i.e. diferentes valores de K_m), o propiedades de regulación diferentes. La existencia de las isoenzimas permite el ajuste del metabolismo para satisfacer las necesidades particulares de un determinado tejido o etapa del desarrollo.

Pueden encontrarse diferentes isoenzimas en los diversos compartimientos intracelulares. Muchas proteínas, y no sólo las que tienen actividad catalítica, pueden presentarse en formas múltiples en las células.

En bioquímica, las isoenzimas son isoformas (variantes estrechamente relacionadas) de las enzimas. En muchos casos, son codificadas por genes homólogos que han divergido con el tiempo. Aunque de forma estricta, las aloenzimas representan enzimas de diferentes alelos de un mismo gen y las isoenzimas representan enzimas de diferentes genes (parálogos) cuyos

productos catalizan la misma reacción, los dos términos se suelen usar indistintamente.

Las isoenzimas permiten una detección fácil y muy específica por medio de la actividad enzimática que se visualiza en un gel nativo. Al realizarse en condiciones no desnaturalizantes, es posible detectar variantes con ligeros cambios en la secuencia y la carga neta de la proteína. Muchas veces las isoenzimas no tienen cambios drásticos de la masa total. La ventaja es que es un método muy sensible. Las enzimas amplifican la señal que es detectada en geles de actividad. En algunos casos las enzimas también tienen algo que ver con alguna función molecular, por lo que pueden estar muy ligados con característica fenotípica deseada. Hoy en día, muy pocos laboratorios miden isoenzimas, aunque en el pasado se usaron para entender mejor la base molecular de la diferenciación celular y morfogénesis.

Marcadores Tipo III (basados en secuencias de DNA)

Tipo IIIa Marcadores no-neutrales (genes codificantes):

Son secuencias de DNA que sí codifican para alguna proteína o enzima, y por lo tanto, los polimorfismos que observamos están sujetos a una presión directa por selección natural o artificial.

Tipo IIIb Marcadores neutrales: (DNA no codificante):

Son polimorfismos de DNA que no están sujetos a una selección directa. Por lo regular son secuencias anónimas que se encuentran en regiones intergénicas. Estos marcadores son útiles cuando están ligados cercanamente a los genes de interés (menos de 5 cM).

En el transcurso de la historia, se ha pasado de marcadores polimórficos del tipo fenotípico, a los de tipo bioquímico y de isoenzimas, hasta los basados en DNA directamente, que son los que más se usan hoy en día.

Tarea 4-1.

Observe el video en la página web:

<http://www.slideshare.net/lautorresm/pcr-marcadores-moleculares>

Explíquelo con sus propias palabras.

Consulte los siguientes sitios para mayor información:

<http://www.angelfire.com/magic2/bioquimica/enzimas9.htm>

http://es.wikipedia.org/wiki/Marcador_gen

4.1.3. Ejemplos de marcadores

4.1.3.1. Single Nucleotide Polymorphism (SNP)

Los *Single Nucleotide Polymorphisms* (SNPs) se originan de las mutaciones puntuales. Son polimorfismos de nucleótido único, es decir,

son variaciones de secuencia del ADN que ocurren por un cambio en un solo nucleótido (A, T, C, o G) de la secuencia. La mayoría de SNP, (dos de cada tres SNP), se originan de la sustitución de la citosina (C) con timina (T). Esta mutación se da a raíz de la metilación de las citocinas. Los SNPs son evolutivamente estables ya que no cambian mucho de una generación a otra, lo que facilita algunos estudios de poblaciones. También tienen elevada facilidad de análisis porque admiten diferentes tipos de genotipaje. Existe un amplio abanico de técnicas para analizarlos. Algunas de estas pueden ser automatizadas y usadas en masa a gran escala. Un ejemplo es la plataforma de *Illumina* que se explica más adelante. Otra ventaja de los SNPs, es que son muy abundantes en el genoma. Las estimaciones actuales son que hay aproximadamente un SNP por cada 100-300 bases en el genoma humano. En algunas plantas el promedio es de uno cada mil bases. Lo anterior implica, por ejemplo, que en todo el genoma humano hay aproximadamente 10 a 30 millones de potenciales SNPs. Más de 4 millones de SNPs se han identificado y la información se ha puesto a disposición del público en internet.

Tarea 4-2

Revise las siguientes páginas web y haga un resumen de lo que tenga que ver con los marcadores moleculares:

www.intl-pag.org/10/abstracts/

www.terralia.com/revista12/pagina43.htm

<http://www.bioline.org.br/request?la05005>

<http://bcs.whfreeman.com/mga2e/bioinformatics/ch01/index.htm>

Problemas de los SNP's:

Los SNP's ofrecen menos información que los microsatélites. El punto débil es la capacidad de discriminación; un SNP puede tener sólo 4 alelos (4 bases diferentes), pero en la práctica sólo hay 2 (**bialélicos**). Los cambios más frecuentes de los SNPs son los cambios de C a T, o de G a A. Un microsatélite puede tener muchos más de 4 alelos, en algunos casos hay hasta 20 diferentes variantes. Es por ello que es necesario utilizar más marcadores SNPs en comparación con los microsatélites.

Ejercicio:

Algunos investigadores han determinado empíricamente que la cantidad de información que pueden dar 4 SNP's es similar a la de un microsatélite. ¿Cuántos SNP's necesitamos para una prueba de paternidad? Teniendo en cuenta que se requieren al menos ocho microsatélites para pruebas de paternidad, entonces si usamos SNPs necesitaríamos analizar al menos 32 SNPs para decir con la misma certeza si el individuo es o no es el padre.

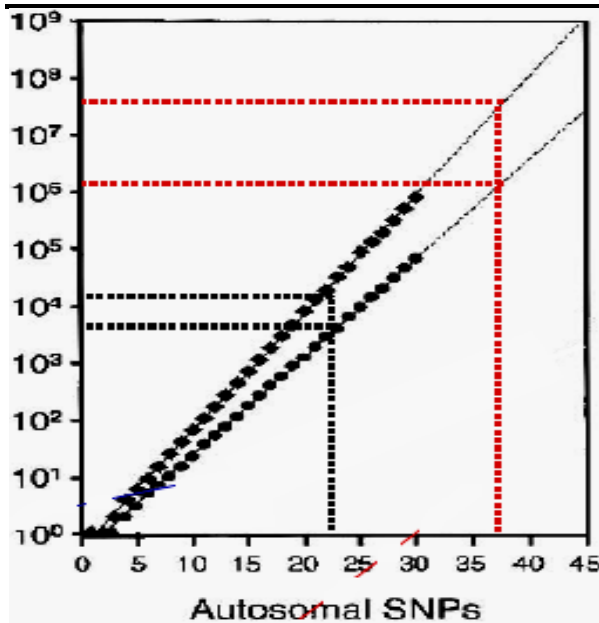


Figura 4-1. Gráfico-SNP's dada una N para que todos los individuos puedan ser discriminados.

Ejemplo:

La gráfica demuestra que hacen falta más SNP's para distinguir dos individuos consanguíneos que para los no consanguíneos. Si tuviéramos la certeza de que en nuestro conjunto de SNP's los alelos, como mínimo estuvieran en una frecuencia ($F \approx 0.2$), entonces con un conjunto de 30-40 SNPs sería suficiente y significativo. Pero, ¿cuál es el problema de esta aproximación? Garantizar que todos los SNP's se encuentran con una frecuencia (0.2). Esto no es tan fácil. Lo ideal sería diferenciar a estos dos individuos entre el intervalo de 10^6 a 10^9 individuos. Pero para obtener estos valores es preciso disponer de un rango entre 50-100 SNP's.

Se analizaron 40 SNP's para ver si cumplían el requisito de 0.2 usando como ejemplo cuatro poblaciones. Existen muchos SNP's donde no hay polimorfismo. Es decir, su frecuencia era $F = 1$. Por consiguiente, su poder de discriminación era cero. Otros SNP's se vio que en alguna de las poblaciones no cumplían $F > 0.2$. De los 40 SNP's, en el mejor de los casos, hay 20 que sirven.

Conclusión:

Para una población en particular es relativamente fácil conseguir un set de 40 SNP's. Pero si es necesario que estos se puedan utilizar para diversas poblaciones, entonces uno necesita analizar un mínimo de 90 SNP's, porque de estos, tenemos la esperanza de que 35-40 cumplan con el requisito de $F = 0.2$.

$A = 0,8$	$a = 0,2$
AA	Aa aa
$P^2 = 0,8^2$	$2PQ$ $Q^2 = 0,2^2$
$\Sigma = (P+Q)^2 = 1$	

¿Cuántos SNP's necesitamos si el alelo más raro tiene una frecuencia de 0.2? Lo ideal es que cada alelo fuera $0.5 = p = q$ dándonos así la máxima información. Pero en la práctica, para un mismo

SNP las frecuencias son diferentes entre diferentes poblaciones. Averigüe más sobre la función de Hardy-Weinberg para resolver la pregunta.

¿Cuánta más información puede dar un alelo con frecuencia 0.5 en comparación con un alelo de frecuencia 0.2?

Técnicas de genotipaje económico y fiable con SNP's:

En la década de los ochentas estos análisis se realizaban por *Southern Blot*, ya que no existían secuenciadores. Cuando se pudo secuenciar DNA, tras la aparición de la PCR (1989), surgen los *arrays* (arreglos) o chips que permiten realizar análisis masivos:

- **Métodos basados en hibridación:** Ensayos de genotipaje para discriminar y genotipar. Diagnóstico de variabilidad por polimerización: *Primer* + Polimerasa que hace crecer la cadena.
- **Discriminación por ligación:** De variantes alélicas en función de si 2 oligos son combinables (OLA, RCA, LCR, DOL).
- **Discriminación por corte:** Se discrimina en función de si se rompe o no el DNA. Se usa una enzima de restricción y después se tiene que analizar el tamaño del fragmento.
- **Secuenciación de DNA:** Al margen de estas cuatro técnicas hay mucha gente que prefiere más secuenciar el trocito de DNA; si ven dos picos será heterocigoto y si se ve un pico, será homocigoto. Esto antes era impensable porque los kits de secuenciación eran muy caros. La secuencia se hace con el método dideoxido en secuenciadores automáticos basados en electroforesis capilar. Podemos secuenciar DNA, pero además podemos genotipar.

Tarea 4-3.

Lea el siguiente texto en inglés y comente con sus compañeros lo que comprendió.

The discovery of single nucleotide polymorphisms (SNPs) and insertions/deletions, which are the basis of most differences between alleles, has been simplified by recent developments in sequencing technology. SNP discovery in many crop species, such as corn and soybean, is relatively straightforward because of their high level of intraspecific nucleotide diversity, and the availability of many gene and expressed sequence tag (EST) sequences. For these species, direct readout of SNP haplotypes is possible. Haplotype-based analysis is more informative than analysis based on individual SNPs, and has more power in analyzing association with phenotypes. The elite germplasm of some crops may have been subjected to bottlenecks relatively recently, increasing the amount of linkage disequilibrium (LD) present and facilitating the association of SNP haplotypes at candidate gene loci with phenotypes. Whole-genome scans may help identify genome regions that are associated with interesting phenotypes if sufficient LD is present. Technological improvements make the use of SNP

and indel markers attractive for high-throughput use in marker-assisted breeding, EST mapping and the integration of genetic and physical maps.

4.1.3.2. Microsatélites

Los satélites son regiones del genoma repetidas en tándem, que no transcriben y están ampliamente distribuidos en los diferentes cromosomas. Cuando se fragmenta y se desnaturaliza el DNA, la mayor parte del DNA se convierte a monocadena. Sin embargo, también se observa un porcentaje de DNA que queda separado del resto. Este corresponde a DNA repetitivo y renaturalizado (doble cadena) que en ese entonces se le llamo DNA **satélite**. En las regiones teloméricas y centroméricas es donde se concentra la mayor parte de los satélites. Los satélites son de 500-1000 bases, mientras que los **microsatélites** son secuencias repetidas más cortas (2-10 bases).

Los microsatélites típicos son nucleótidos AT repetidos en tándem de cinco a veinte veces. A los microsatélites también se les llama SSRs (**Simple Sequence Repeats**). Su polimorfismo viene dado por el número de veces que se repite su matriz. Se pueden diseñar *primers* que amplifiquen estas regiones para discriminar las diferentes variantes alélicas. Se efectúa entonces un PCR y después se corre una electroforesis en gel de agarosa o acrilamida para separar las variantes por su tamaño. En el genoma de maíz, la abundancia de los microsatélites es elevada, sus polimorfismos son múltiples y su distribución es más o menos uniforme a lo largo de los cromosomas. Son marcadores codominantes, lo que permite distinguir heterocigotos de homocigotos. Normalmente se encuentran en regiones no codificantes, son marcadores neutros. (Figura 4-2).

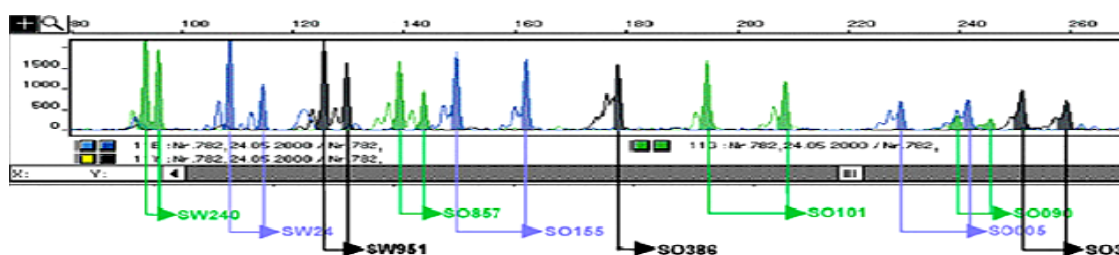


Figura 4-2. Ejemplo de microsatélites

Los marcadores SW240, SO357 y SO101 (verdes) y el SW 26 y SO005 (azules) presentan dos picos separados. El SO386 (negro) es el único homocigótico ya que presenta un solo pico. Si en este rango no hay ningún otro marcador significa que no hay heterocigosis. Cuanto más repeticiones en tándem, más separados estarán los alelos.

¿Cómo se generan los polimorfismos de microsatélites?

La polimerasa va replicando el DNA, al trabajar con tándems puede ser que se forme un “loop” (vuelta) con lo que al unirse con la cadena complementaria habrá un número elevado de repeticiones. Si este *loop* se

forma en la cadena molde, el fragmento que se generaría tendría una repetición menos. A esto se le llama un **deslizamiento** “*slippage*” de la **replicación** (Figura 4-3).

Figura 4-3. Modelo de *slippage* de la polimerasa

Aplicaciones de los microsatélites:

Gracias a su polimorfismo y su rápida evolución (dan lugar a muchos alelos), existen muchos en el genoma. Muchos son de *locus* único y neutro (no codificante). Además, se encuentran en el genoma de casi todos los organismos.

Pruebas de paternidad:

Normalmente se evalúan unos 15 microsatélites. Si dos o más no coinciden se puede descartar la paternidad. Pero también sirve para encontrar altos grados de parentesco o bien saber de quién proviene un cierto material genético. La técnica se basa en la amplificación de los microsatélites mediante **PCR** (utilizando cebadores específicos para cada uno de los 15 microsatélites utilizados). Al ser secuencias nucleares, cada individuo posee dos copias de cada marcador, una de origen paterno y otra de origen materno, que pueden ser iguales o diferentes entre sí. Los 15 marcadores poseen de 5 a 22 variantes cada uno (es prácticamente imposible encontrar 2 individuos iguales). La identificación es inequívoca y es independiente del tipo de muestra que se utilice. Accede directamente al genotipo de los individuos y, por tanto, no se ve afectado por las relaciones de dominancia ni por las variaciones de penetrancia o expresividad ya que estos marcadores se heredan de forma codominante permitiendo así, determinar también el genotipo (Figura 4-4).

Figura 4-4. Prueba de paternidad con marcadores

Estudio del origen de poblaciones (migraciones humanas):

Se analizan diferentes poblaciones y se intenta encontrar caminos lógicos para explicar cómo han pasado de un lugar a otro. ¿Cuánta diversidad hay en una población determinada? ¿Qué individuos pertenecen a una subespecie o a otra? Los marcadores moleculares pueden dar respuestas a este tipo de preguntas.

Problemática de los microsatélites:

Son relativamente difíciles de encontrar en una especie nueva. Las colecciones que se tienen ahora han sido una labor de muchos años. Antes se hacían en geles de agarosa, pero era muy difícil diferenciar entre 5 pb. Hoy muchos se hacen en geles de acrilamida. Algunos SSRs son complicados de interpretar por lo que aparecen otras bandas inespecíficas. Muchas veces no es posible automatizar por lo que se necesita anotación manual.

4.1.3.3. *Random Amplified Polymorphic DNA (RAPD)*

Se diseñan *primers* cortos (10-16 bases) con una secuencia aleatoria. Con cada par de *primers* escogidos al azar se hace un PCR con DNA genómico a una temperatura de alineamiento (T_m) bastante baja (35-40° C) y después se corre en un gel de agarosa. Los RAPDs son útiles cuando no hay información previa del genoma ni de los genes. Como los *primers* son pequeños habrá cientos de sitios en el genoma donde hibridarán. Con suerte puede ser que los *primers* se unan a regiones del genoma no muy lejanas. Si se unen en una orientación complementaria habrá amplificación. Si la orientación no es correcta, entonces no habrá amplificación exponencial.

La metodología se ha aplicado más a DNA de plantas, que de animales. Por lo regular los RAPDs dan a lugar de 5 a 20 bandas que muchas veces no son muy específicas, es decir, que varían bastante según las condiciones de cada experimento. Bajo condiciones ideales, se usa un par de *primers* con DNA de muchas variedades diferentes y se obtienen bandas de diferente tamaño para cada variedad. Muchas veces son las mismas bandas en las muestras, pero a veces en una sí se amplifica y en la otra no.

Ventajas: Para diseñar los *primers* no hace falta tener información previa del genoma del organismo. De hecho, se pueden usar los mismos *primers* para diferentes especies.

Problema: Al hacerlo a menor temperatura hay problemas de repetibilidad en PCR. Por ejemplo, la amplificación varía mucho en función de la cantidad de DNA genómico. En los RAPDs a veces aparecen o desaparecen bandas, y muchas veces esto no tiene que ver con la secuencia del DNA. Es una técnica poco robusta. Por eso, cuando se encuentra 1 banda consistente se diseñan *primers* más específicos para una PCR normal.

Tarea 4-4:

Calcule cuántas variantes posibles existen para un *primer* de 10 bases. Asumiendo un tamaño de genoma como el del arroz (450Mb), ¿cuántas veces en promedio se pegará un primer aleatorio de 10 bases?

Respuesta:

Existen $4^{10}=1048576$ variantes diferentes de un primer de 10 bases.

Para calcular las veces que se pega asumimos que el primer se puede unir a cualquiera de las dos hebras del ADN. Es decir, se tienen dos formas de escanear el genoma ($2 \times 450\text{Mb} = 900\text{Mb}$)

En promedio, un primer de 10 bases se pega $(900'000'000/1'048'576) = 858$ veces

4.1.3.4. Restriction Fragment Length Polymorphism (RFLP)

A finales de los años 80, los marcadores moleculares más usados fueron los RFLPs. El procedimiento es el siguiente: Se extrae el DNA genómico y se hace una digestión con enzimas de restricción. Después se hace una electroforesis y se transfiere el DNA a un filtro de nylon. Se hace un marcaje e hibridación del DNA digerido con una sonda radiactiva. Se observan diferentes patrones de bandas entre individuos. Los polimorfismos de DNA tenían ventajas sobre los polimorfismos bioquímicos porque permitían descubrir muchas más diferencias; el nivel de variabilidad de polimorfismo aumentó mucho. Individuos que se veían muy parecidos a nivel del fenotipo, podían ser muy bien distinguidos con marcadores de DNA. Se acuñó entonces el término de “*fingerprinting*”. Dos individuos diferentes no comparten el mismo patrón de bandas (Figura 4-5.).

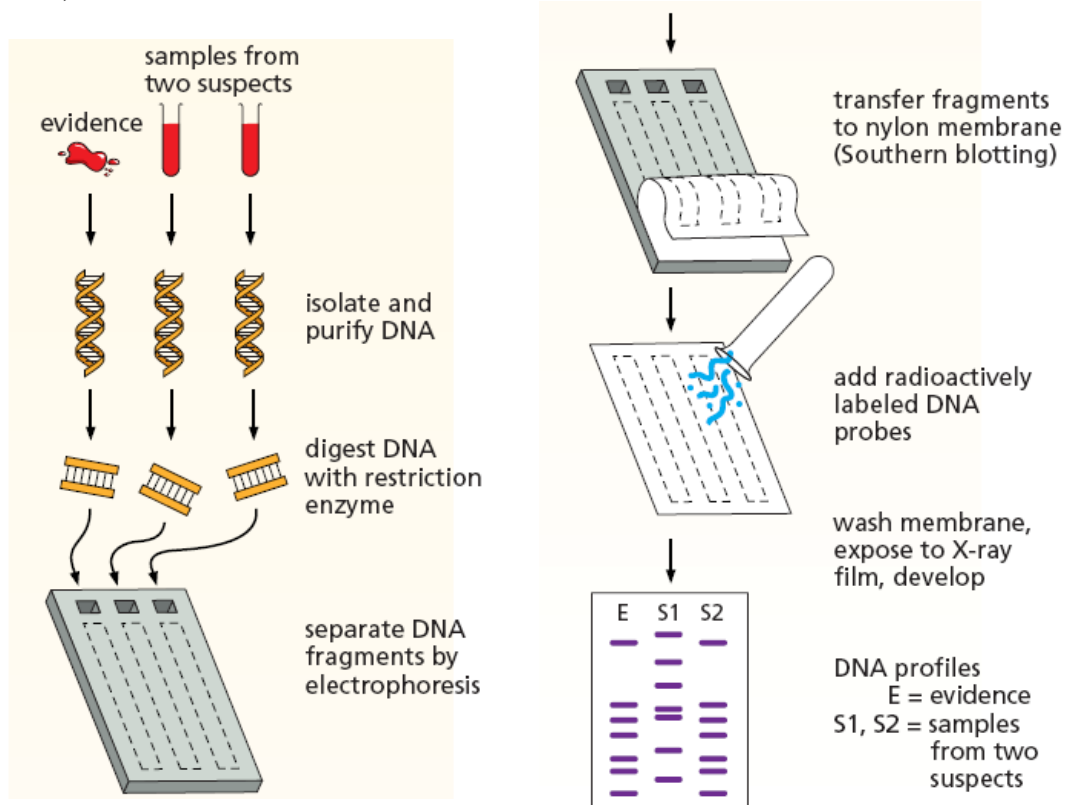


Figura 4-5. Pasos para realizar una prueba de RFLP

4.1.3.5. *Amplified Fragment Length Polymorphism (AFLP)*

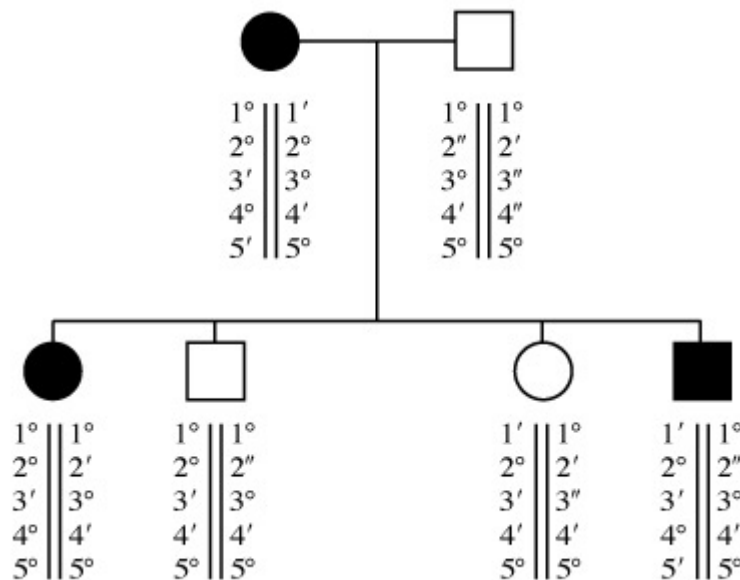
Los AFLPs tratan de superar las limitaciones prácticas de otras técnicas. De la falta de robustez y repetitividad de los RAPDs por un lado, y de la cantidad de DNA necesario para los RFLPs por otro lado. Son parecidos a los RFLPs, pero en lugar de usar enzimas de restricción se usa amplificación por PCR. Se hacen dos rondas de PCR con *primers* anidados. En la primera ronda de PCR se usan *primers* universales, y en la segunda se usan *primers* con extensiones específicas. La ventaja de los AFLPs es que se requieren de mucho menos DNA de buena calidad para empezar.

Primero se usan **enzimas de restricción** para digerir DNA. Por ejemplo *EcoRI* (reconoce 6 bases) y *M* (reconoce 4 bases). Una de las enzimas lleva la etiqueta de corte frecuente y la otra de corte menos frecuente. Este es un punto crítico de la técnica, porque **E y M dejan extremos cohesivos**, de forma que hecha la digestión, usaremos adaptadores (fragmentos de DNA de doble cadena que serán cohesivos en los extremos generales para E y M, de forma que en 3' y 5' tendrán secuencias que conocemos).

Con esto conseguimos millones de fragmentos de los que conocemos secuencias de los extremos 3' y 5' y que tienen tamaños variables. Gracias a esto podemos amplificar diseñando *primers* específicos de M y E. Si hacemos PCR con estos *primers* será un fracaso porque habrá demasiados fragmentos. Por esto aún metiendo primer E + M tampoco podríamos ver nada. Todos los fragmentos digeridos quedan unidos por los extremos a los adaptadores. Se conoce su secuencia porque la ha diseñado el investigador. Obtenemos un conjunto de fragmentos con secuencias de los extremos conocidos. Podremos hacer un PCR específico. Los *primers* de este PCR serán complementarios a la secuencia de los diferentes adaptadores. No sólo amplificaremos los fragmentos que tienen los dos tipos de adaptadores diferentes sino amplificaremos mucho y en el PCR veremos un barrido de bandas.

Tarea 4-5.

The gene for the autosomal dominant disease shown in this pedigree is thought to be on chromosome 4, so five RFLPs (1-5) mapped on chromosome 4 were tested in all family members. The results are also shown in the diagram; the superscripts represent different alleles of the RFLP loci.



- (a) Explain how this experiment was carried out.
 (b) Decide which RFLP locus is closest to the disease gene (explain your logic).
 (c) How would you use this information to clone the disease gene?

Método: Se hace un PCR pero diseñando cuatro *primers* E de la misma secuencia excepto por el nucleótido en 3' que en un caso será A, en otro T, C y G. Se hace lo mismo con los *primers* de M. Se hacen 4 alícuotas de primer E y cuatro de los primers M. En total se hacen 16 PCR diferentes con las combinaciones de *primers*. Así en cada PCR se logra amplificar un subconjunto de fragmentos. Al hacerlo, comenzarán a intuirse bandas pero seguirán habiendo demasiados fragmentos. Fue la **PCR pre-selectiva** donde habríamos enriquecido un subconjunto de moléculas del DNA genómico del que habíamos partido. Entonces se diseña un nuevo subconjunto de *primers* que además de tener secuencia complementaria a E y la A, dos nucleótidos más adicionales, más la secuencia complementaria a los adaptadores, igual con los otros 3 *primers* de E. Así, se tendrán 16 combinaciones, 16 *primers* dEA, 16 de EC, 16 de ET y 16 de EG logrando 64 PCR's diferentes en total. Por tanto 64 *primers* MSEI. Estamos ante la **PCR selectiva**. Ahora sí que el número de bandas era discreto (100-200 bandas) y hacía falta usar **gel acrilamida** o bien marcar primers fluorescentemente.

Ventajas: A diferencia de RAPD, se obtienen 10-20 veces más bandas (genera mucha información). Es una técnica robusta y repetitiva ya que usa oligos largos y el Tm es más alto y específico. No es necesario conocer el genoma, incluso se puede hacer *fingerprinting* de un DNA desconocido. Es una buena técnica para obtener marcadores.

Problema: Es obligatorio hacer las dos PCR's anidadas porque sino

habrá demasiada amplificación. No se puede distinguir homocigoto dominante y hetero (genera las mismas bandas para ambos). No hay softwares que permitan automatizar las lecturas.

Pruebas de paternidad: El hijo debe tener una banda del padre o de la madre. si no tiene ninguna banda de los dos no es hijo suyo. El problema es que se tardaba más tiempo comparando el gen del hijo con el de los padres. Si hubiera softwares mejores los AFLP se harían más. Si el DNA está degradado o de mala calidad generará problemas en la técnica, ya que el 1er paso de la técnica es hacer la digestión del DNA genómico, si este ya viene cortado o está en muy mal estado, no funcionará bien, no será tan robusta ni repetitiva.

4.1.4. Estrategias de genotipaje

4.1.4.1. Técnica de extensión de primers

Unas de las más usadas son las que utilizan un secuenciador. Usa el *primer* que queremos extender. Diseñamos un *primer* que en 3' tenga el nucleótido adyacente a la posición de la mutación del DNA. Dejamos hibridar, colocamos la DNA polimerasa y 4 dideoxido nucleótidos con fluorescencias diferentes.

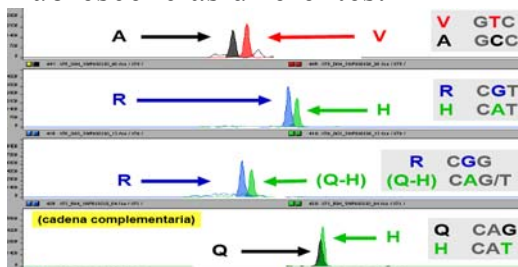


Figura 4-6

Heterocigoto: 2 picos. La DNA polimerasa pondrá ddG y a otros ddA. Pero cabe recordar que al ser dideox, sólo podrá añadir 1 nucleótido dd más a continuación del “*primer*”. Se pueden detectar diferencias de migración debido a las marcas fluorescentes.

Se puede usar una extracción rápida de DNA (15-20') en condiciones alcalinas para degradar el RNA. Seguidamente **PCR** (1:30') y tratamiento con fosfatasa y exon (**ExoSAP** = cóctel de las dos enzimas). Tan sólo 30 minutos de digestión. Ponemos el cóctel con los dideox (1:20') inactivadas previamente las otras enzimas. El paso **CIP** es para cuando analizamos muchos SNP's (elimina los dideox no hibridados). Si nos lo saltamos, no sólo veremos el primer del pico, sino también otras

señales que son los *dideox*. La eficiencia con la que la DNA polimerasa incorpora un *dideox*, no es la misma de que los picos tengan la misma altura. Las diferencias de migración en las gráficas son debido a las fluorescencias. Todo este proceso tarda unas 12 horas/placa (384 muestras/placa).

Ventajas: Técnica muy rápida.

Problema: Es necesario purificar *primers*/DNA: Si no, al ponerlos con los *dideox* los *primers* y DNAs mas cortos igualmente se alargarían por lo que veríamos muchos picos con muchas fluorescencias. Se puede usar enzimas exonucleasas para eliminar DNA/*primers* y fosfatasa alcalina para inactivar nucleótidos. Se puede hacer “*in situ*” pero después se deben inactivar estas enzimas antes de la reacción de PCR. Tambien se puede usar una columna de exclusión molecular donde quedará retenido el DNA y se eluirán *primers* cortos y nucleótidos.

4.1.4.2. PCR en tiempo real con sondas TaqMan

Este método requiere del monitoreo de la cantidad de producto en cada ciclo del PCR. La fluorescencia se puede detectar:

Por método óptico normal con un solo filtro tipo *highpass*: Se estimula con una lámpara a varias longitudes de onda simultáneamente a una molécula marcada fluorescentemente y se detecta la fluorescencia total. Esta técnica no permite detectar varios tipos de fluorescencia a la vez, y es por eso que solo se puede ver una sola marca por carril.

Por secuenciadores automáticos: Se usan diferentes láser con diferentes longitudes de onda de excitación. Esto permite leer varias fluorescencias dentro del tubo de forma continua y simultanea. Además, el láser de fluorescencia permite ser lo suficientemente sensible para trabajar con volúmenes muy pequeños dentro de tubos de vidrio capilares.

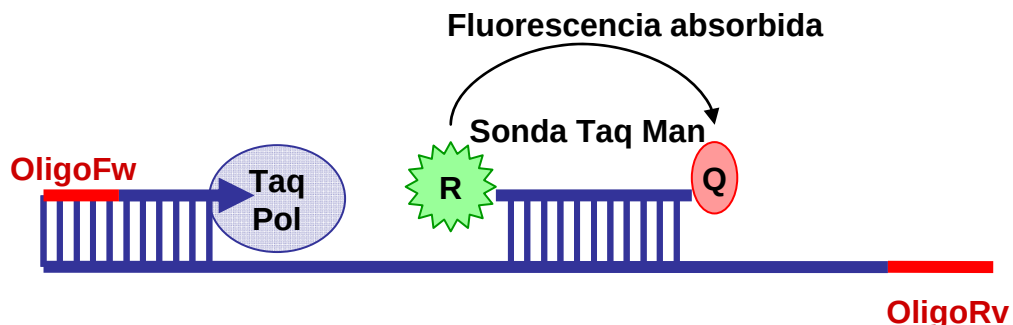


Figura 4-7. Fluorescencia absorbida son una sonda Taq Man

Sondas Taqman: La polimerasa Taq, además de extender, elimina, pues la sonda está marcada en 5' por un fluorocromo y en 3' por otro. Cuando la Taq extienda el *primer* se acabará encontrando con la

Taqman que tendrá activo exon y la separará (los fluorocromos quedarán separados no anulándose entre si). **La cantidad de fluorescencia** que se detecta **es directamente proporcional al número de moléculas molde que se están extendiendo**. Es una **técnica cuantitativa**. También se utiliza para genotipar en tiempo real (Figura 4-7).

Método para genotipar:

Añadimos dos sondas TaqMan diferentes en el PCR y deben contener el nt que estoy genotipando y cada una será complementaria a uno de los dos alelos, además, cada uno tendrá una fluorescencia diferente, por tanto, en esta reacción tendremos 4 *primers*; 2 *primers* R y Q y 2 sondas TaqMan. Al final del PCR sólo queremos saber qué fluorescencia será mayoritaria. Este método lo podemos hacer con PCR normal y después PCR cuantitativa para que lea la fluorescencia.

Homocigoto: Hibridará una de las dos sondas porque será totalmente complementaria. 90/10 fluorescencia.

Heterocigoto: Habrá un 50% de eficiencia de hibridación para cada sonda. 50/50 fluorescencia.

Ventajas: No se requiere purificación. En esta se hace todo en un tubo. Es eficiente y rápida.

Problema: No se genotipan mas de 1 SNP. Sería difícil que la hibridación sea correcta si se usan más (no se puede hacer multiplex). Es necesario comprar 2 sondas TaqMan con 2 fluorocromos diferentes, lo que lo encarece. Sólo vale la pena si se tiene que genotipar muchas muestras con el mismo marcador. Las sondas deben llevar modificación química (MGB) en el segundo fluorocromo para ganar especificidad. Muy importante con las sondas TaqMan porque son muy pequeñas y es mejor aumentar su T_m y especificidad a la hora de hibridar. No es eficiente si tenemos poca muestra pero muchos SNP's.

4.1.4.3. PCR en tiempo real con sondas FRET

Las sondas FRET emiten fluorescencia cuando dos fluorocromos están juntos. Se diseñan dos *primers* adyacentes a la posición de los SNP's a genotipar, si los dos *primers* ligan y se enganchan emitirán fluorescencia. Aquí volvemos a tener 4 *primers*, los dos de la PCR y los dos que discriminarán un alelo de otro. **El extremo 3' de la sonda reportero está bloqueado y no se extiende**. También podemos hacer RT-PCR a tiempo real mediante cinética de desnaturalización del

producto amplificado. Un solo nucleótido cambiado hace suficiente diferencia entre la T_m como para que la cinética cambie.



Figura 4-8. Ejemplo de una sonda FRET

Ventajas: No es necesaria la degradación porque se emite fluorescencia sin eso. Tan sólo es necesaria hibridación.

Problema: Es costoso.

4.1.4.4. PCR alelo-específica

Sistema de PCR que permite la amplificación de uno alelo específico individual y no del resto de alelos posibles. Esto debido a que el primer *forward* es complementario solo a uno de los dos alelos del gen. Con **fluoróforos intercalantes (SYBR Green)** la calidad de la detección depende de la especificidad de los *primers*, de la interacción entre *primers*. La fluorescencia se emite cuando se intercala en el surco menor del DNA de doble cadena (dsDNA). Por lo que hay que diseñar un primer “forward” que en 3’ acabe con el nucleótido que se está evaluando. Sabemos que si esta última base del *primer* no se complementa la PCR bajará mucho su eficiencia.

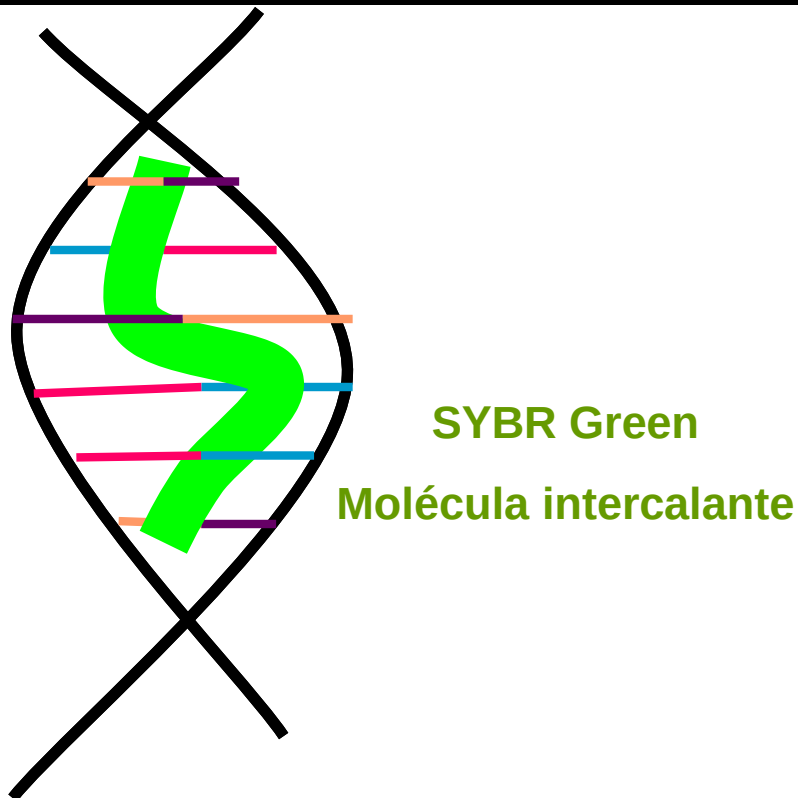


Figura 4-9. SYBR Green II es una molécula que flúorese cuando se une a DNA de doble cadena.

Homocigoto: Habrá amplificación y sólo fluorescencia en un tubo, de lo contrario será heterocigoto y con fluorescencia en ambos tubos. Podría ser homocigoto mutante. Para comprobarlo se tiene que hacer un segundo PCR con un nuevo *primer*.

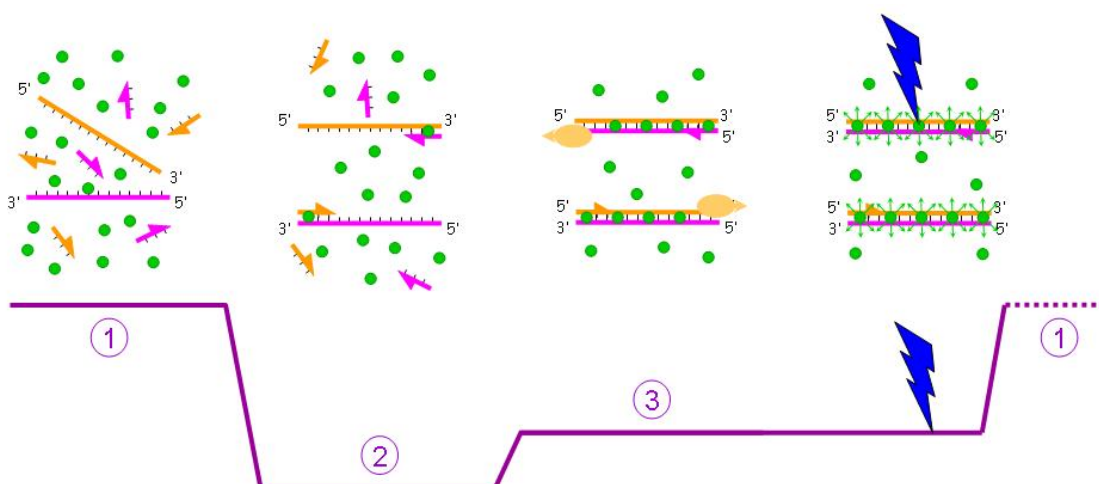


Figura 4-10. Se muestra una molécula verde intercaladora que se une a la doble hélice del DNA

SYBR Green I emite fluorescencia cuando se une al DNA de doble cadena. Es decir, no reacciona con el DNA de cadena sencilla. Si hacemos

PCR en un medio donde hay SYBR esté se podrá unir al DNA en los momentos que esta presente como doble cadena, y así podemos ver la amplificación del DNA dándonos una señal fluorescente en tiempo real.

Ventajas: Se puede hacer con solo 2 *primers* sin estar marcados. El SYBR sirve para todos los SNP's ya que se intercala inespecíficamente en cualquier secuencia de DNA de doble cadena, por lo que sirve para cualquier genotipado. Permite ahorrarse el marcaje fluorescente de los *primer* Taqman o FERT, que son sondas específicas muy caras.

Problema: No se puede hacer multiplex en el mismo pozo. Es difícil optimizar varias PCR's al mismo tiempo.

Métodos de genotipaje:

Sabiendo la secuencia flanqueante al SNP queremos genotipar un polimorfismo A-G;

Homocigoto: Para A esperaríamos un pico el doble de alto.

Heterocigoto: Esperamos un pico bajo porque sólo se elonga una de las dos cadenas. Después de interrogar (A) lo hago con (G) por lo que lanzo (C), si vemos un pico equivalente a 1 incorporación nos indica que se alarga la cadena y que por tanto es heterocigoto. Segundamente interrogamos (C) (tendremos (G)); por lo tanto esperamos pico doble ya que este nucleótido está en ambas dos cadenas.

Podremos elaborar un pictograma teórico porque conocemos la secuencia flanqueante. Si en el alelo 2, en vez de (G) fuese (C), al añadir (G) veríamos un pico que sería el triple de (A) por lo que se incorporarían 3(G).

Secuenciar cadena simple: Puedo analizar haplotipos (si encontramos 2 SNP's muy cercanos entre sí con el dideox no podríamos saber que el individuo es doble heterocigoto ya que no distinguiría. En los 2 casos existe doble heterocigoto pero no podemos decir en qué orden y qué fase está el haplotipo. Con esta técnica sí podremos saberlo:

$$\frac{C}{T} \frac{G}{C} \text{ ó } \frac{C}{T} \frac{C}{G}$$

Pedigrí a través de probabilística en la población: Con las frecuencias alélicas de los marcadores. Pueden influir los haplotipos siempre y cuando las 2 posiciones de los SNP's estén más lejos de 100 pb (ya que la muestra no puede ser más larga).

Mezcla de muestras de DNA e interrogamos SNP: Si vemos una altura muy diferente entre el grupo control y el grupo enfermo podemos inferir que esa mutación está en la enfermedad.

Patrones de metilación diferencial en el DNA: Las C's pueden estar metiladas y estos patrones de metilación son importantes para determinar el "inprinting". El gene *inprinting* se refiere cuando individuos para un gen

determinado, sólo expresa uno de los dos alelos, ya sea el paterno o el materno. Pero, ¿cómo se estudia el patrón de metilación? Se hace una reacción con bisulfito que transfiere la C metilada en un análogo de base, después se hace *pyrosequencing* y en función del patrón y se determina si aquella C estará o no metilada.

Ventajas: Puede discriminar mezclas de 400 muestras (patologías, expresión) pero el DNA mezclado debe estar bien cuantificado. Se debe poner la misma cantidad de cada muestra. Si no es así, el genotipaje de la cual añadamos más cantidad estará sobrerrepresentado. Es una alternativa al dideox para análisis masivo de DNA.

Problema: La cantidad de información es mas baja comparada con “*dideox*”. Como máximo evalúa 100pb y en el otro 600-700 pb.

4.1.4.5.SNPlex

Para genotipar simultáneamente 48 SNP's en una muestra de DNA. En realidad no estamos obligados a hacer un PCR para cada par de *primers*. Para cada SNP diseño 3 *primers*; uno en 5' que comenzará en la posición 3' del SNP complementario al molde, pero su extremo 3' será universal. Los otros dos *primers* tendrán el último nucleótido en 3'; esto es, que uno será complementario al primer alelo, y el otro al segundo.

La discriminación de los dos alelos de un SNP se realiza mediante una ligación específica del alelo (OLA), en la que se utilizan dos oligos con el nucleótido del SNP en el extremo 3' y un oligonucleotido común que se une a la región adyacente al SNP. Los productos de la ligación se amplifican mediante una reacción de PCR múltiplex, que utiliza como *primers* las secuencias universales presentes en los oligonucleótidos. Tras la reacción de PCR, los amplicones se hibridan a una mezcla de sondas fluorescentes (sondas ZipChute), complementarias a las secuencias ZipCode (también presentes en los oligonucleótidos, y que muestran movilidades únicas y pre-optimizadas en una electroforesis capilar). La identificación de las sondas ZipCute que han hibridado se realiza en un analizador de DNA y los resultados se analizan con el software GeneMapper. Todas las reacciones se realizan en placas de 384 pozos, genotipándose hasta 48 SNPs simultáneamente en una muestra. Si hacemos esto en 48 muestras, todas tendrán en común las 2 regiones universales, los Zips serán diferentes para los 48 SNP's.

Homocigoto para C: Los 2 *primers* se podrán enganchar pero el de A no al último. La ligasa no lo permitirá. Con el PCR se amplificarán las señales de los *primers* ligados. Pero deberemos saber para cada uno de los 48 qué nt se ha ligado, si “A” o el *primer* “G”. Ahora sólo será necesario observar las diferentes fluorescencias pero no hay 48 parejas de fluorescencias si no que se combinan con la movilidad electroforética. Veríamos 48 puntitos. Heterocigoto para los 48 SNP’s: Veremos 96 puntitos.

- Ventajas:** Las 48 sondas utilizadas servirán para cualquier tipo de muestra. Es económico porque se puede genotipar muchas muestras a la vez. La plataforma puede producir unos 200,000 genotipos a la semana.
- Problema:** Como se realiza (OLA, ligación) se necesitan grandes cantidades de DNA genómico. Las reacciones de ligación entre dos regiones del genoma pueden ser muy discordantes, por lo habrá muchos SNP’s que no son compatibles entre muestras.

4.1.4.6. Sistema de Illumina

La tecnología de Illumina es una alternativa con mucho potencial y rendimiento. Permite genotipificar muchos SNP’s simultáneamente de la misma muestra. Es una técnica de microarreglo (microesferas unidas a *primers* de secuencia definida). Las microesferas con diversos *primers* unidos (todos ellos de la misma clase para una esfera) se colocan en unos tubos con filtros. Si se tienen 1536 esferas diferentes y en la superficie del arreglo hay unos 30 filtros, los filtros se unen entre sí formando fibras más grandes. Cada arreglo engloba 50,000 fibras ópticas por lo que se tendrán 90,000 arreglos/mm².

Figura 4-11. Microbolas que se unen a los oligos.

Método de genotipaje:

Para cada SNP se diseñan tres oligos (dos serán idénticos excepto en el último nucleótido que será complementario 1 a cada variante), una parte de estos *primers* “forward” será una secuencia universal. El *primer* que situaremos al 3’ del SNP tiene una región complementaria al DNA, seguida de una secuencia complementaria a la secuencia de la microesfera bola y una 3ª región que será universal. Con el genómico y el trío de 1,500 oligoelementos haremos PCR, (sólo harán falta los 2 *primers* hibridados) gracias a las 2 secuencia universales de los P1 y P2. Con una fluorescencia distinta cada uno y un tercer primer que será complementario a la universal de reverso.

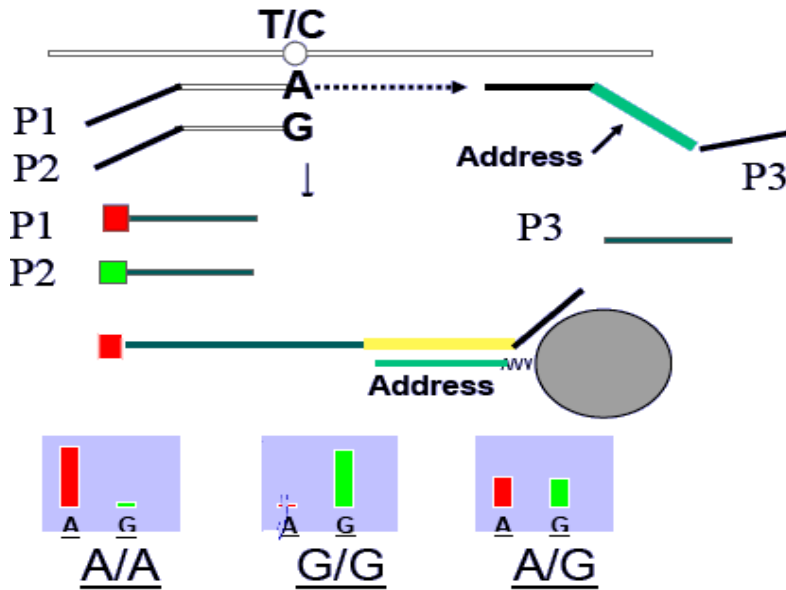


Figura 4-12. Técnica de genotipaje con microbolas.

Homocigoto: Sólo se amplificará a partir de un *primer* (una única fluorescencia). ¿Cómo se verá? Hay que pasar el producto amplificado por los filtros con esferas, en función de la secuencia. El que lleve el *primer* se unirá con unas esferas u otras (será necesario no perder de vista en aquel arreglo en qué la fluorescencia es mayoritaria y el *software* nos indica el genotipo).

Heterocigoto: Veremos dos fluorescencias.

Las Fases: Extensión del alelo específica por PCR. Esto se puede sofisticar en una misma esfera donde se pueden colocar diferentes secuencias, aumentando así el número de SNP's que se puede genotipificar. Illumina igual que SNPLEX tiene dopantes, es decir, algunos SNP's que no hibridan correctamente. Utilizando este aparato se pueden hacer un millón de SNP's al día. Más recientemente, ha salido un chip de Illumina donde las esferas se colocan de forma distinta.

Ventajas: Es barato. Cuesta ~0,05 USD/ SNP y ~400 USD por *array*. Se considera una técnica muy potente.

Problema: El equipo para leer es caro. Si hay mutaciones alrededor de los SNP's. Los primers no hibridarán bien y por eso hay algunos que no funcionan.

4.1.4.7.SNP Chips

Los chips más conocidos los maneja Affimetrix, aunque también existen de otros proveedores. Se colocan oligos de 15 nucleótidos sintetizados en un punto determinado por fotolitografía. Gracias a un láser se pueden colocar muchos oligos por unidad de superficie. Así, se pueden hacer

arreglos con decenas de miles de oligos. Son servicios comerciales. Los científicos en los laboratorios sólo deben aislar mRNA, pasar a cDNA, marcar con fluorescencia y dejar hibridar, para ver cuáles de aquellos genes se expresan más o menos.

La técnica no está limitada por el número de oligos, no consiste en una sola hibridación y sólo hace falta una PCR (con *primers* universales).

Método de genotipaje:

Se pueden colocar 2 oligos que únicamente varíen por el SNP, pero hay que tener en cuenta que los oligos son cortos y la hibridación podría ser muy sutil. Otra manera, es que, por cada SNP se diseñe un conjunto de oligos diferente, en el que haya 0 *missmatch*, 1, 2, 3, asegurando así, que en aquel oligo no haya tan sólo una diferencia de un SNP. Entonces, se puede obtener un patrón de intensidad de hibridación dependiente del genotipo de la muestra. La máquina interpretará este patrón (gracias a que en esta técnica no hay limitación de espacio) y nos dirá cuál es el genotipo. Actualmente se cuenta con un *GeneChip* para 100,000 SNP.

Figura. 4-13. SNP-Chip-Based Genome-Wide Analysis

Tarea 4-6

Consulte la siguiente página y haga una presentación de la información con sus propias palabras: http://www.mun.ca/biology/scarr/DNA_Chips.html

Es necesario fraccionar el DNA genómico (a partir de 250 ng de buena calidad) y unir adaptadores en cada extremo. Se hace un PCR universal y se rescata la región de interés. Después estos se vuelven a fragmentar y se marcarán con fluorescencia. Al hacer PCR, sólo amplificarán bien los fragmentos pequeños (reducción de complejidad). Para prevenir esto, se pueden prever los tamaños de la restricción, haciendo esta, *in silico*, es decir, en programas especializados. De esta manera, se asegura que el SNP quede en un fragmento de longitud adecuada. Existen chips que genotipifican todas las mutaciones conocidas por ciertos genes en específico. Por ejemplo, los genes CYP2C19 y CYP2D6 que están relacionados con el metabolismo de fármacos en humanos.

Método MIP (Universal): por cada SNP se diseña una sonda, que consiste de oligo muy largo, en 5' y 3' con regiones complementarias a las regiones flanqueantes del SNP a genotipificar. Cuando la sonda híbrida forma un semicírculo, dejando un espacio libre (SNP a analizar). Este se rellena con nucleótidos con distintas fluorescencias. Se usa la enzima exonucleasa para degradar el DNA. La secuencia semicírculo tiene dianas, con secuencias adyacentes conocidas y entonces, se diseña un *primer*. Se abre el círculo y

se amplifica por PCR. Con el producto se diseña un chip donde se introduce la secuencia del oligo. El producto de la PCR hibridará y se podrá observar la fluorescencia. En estas sondas del chip no serán complementarias al SNP, sino a la secuencia bioinformática generada (diferente para cada sonda MIP, color verde). Si se diseñan 12,000 sondas con 12,000 secuencias bioinformáticas, en función del SNP que se quiera analizar, se le añade una región roja a otra, así que el chip no sólo se podrá usar para una sino sino para cualquier especie. Se tienen cuatro tubos (cada uno con un círculo no cerrado). En el primero se añade (A), en el segundo (C) no marcados. Entonces se amplifica por PCR.

Homocigoto: Sólo habrá amplificación en un tubo si es homocigoto, en dos si es heterocigoto. Como se sabe si el SNP es entre A y C, sólo se usan 2 fluorescencias. Se dispone del mismo primer marcado con fluorescencias diferentes. En el chip las sondas serán complementarias a estos *primers*.

Ventajas: Se pueden diseñar sondas específicas que tan solo difieran de la región flanqueante por el SNP pero manteniendo el mismo blanco de ER y secuencia flanqueante a la diana. Los arreglos no poseen oligos específicos de los SNPs.

Problema: No es eficiente marcar cada nucleótido con una fluorescencia diferente. Las sondas que se usan ahora tienen un solo marcaje fluorescente al final.

4.1.4.8. Espectrometría de masas (MALDI-TOF)

Se mezcla el DNA con un solvente y después se coloca una microgota en un chip. Por lo regular se usa una solución de algún compuesto que absorba la energía láser como el ácido fórmico (matriz). El chip se expone al vacío y con la energía de un rayo láser se volatiliza y se ionizan las moléculas. De ahí el nombre de *Matrix-Assisted Laser Desorption Ionization* (MALDI). Las moléculas con cargas eléctricas positivas o negativas puedan ser aceleradas hacia el detector de masas. Los polimorfismos son clasificados en un espectrómetro de masas por su tiempo de vuelo, *Time of Flight* (MS-TOF). Con el mismo campo eléctrico aplicado a moléculas con diferente masa, se generan diferentes aceleraciones o velocidades, y esto se mide por el tiempo que tardan en recorrer la longitud del detector. Los dos trozos de DNA de igual longitud pero con un nucleótido diferente, tendrán una masa diferente, y por tanto, diferente tiempo de vuelo.

Ventajas: Analiza fragmentos amplificados por PCR

Problema: El equipo de MALDI-TOF es costoso, y además no todos los modelos se pueden manejar de manera automática, ya que muchos necesitan la interacción manual del usuario para apuntar el láser en el punto correcto.

4.2. Técnicas de comparación genómica

4.2.1. Differential Display (DD)

Por medio de esta técnica se comparan genes de expresión diferencial. Es parecida a los AFLPs, pero en lugar de partir de DNA genómico, se inicia con RNA, que se transcribe a cDNA usando un oligodT con nucleótidos específicos adicionales. Se buscan bandas de marcadores que distingan a las variedades. Con esta técnica se encuentran genes con expresión diferencial entre dos muestras.

Tarea 4-7

Lea el siguiente texto en inglés. En base a lo que entendió, explique ahora con sus propias palabras en español como funciona la técnica de *differential display*.

Differential Display allows rapid, accurate and sensitive detection of altered gene expression (Liang, P. and A.B. Pardee. 1992. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction, Science 257, 967-971). The mRNA Differential Display technology works by systematic amplification of the 3' terminal portions of mRNAs and resolution of those fragments on a DNA sequencing gel. Using anchored primers designed to bind to the 5' boundary of the poly-A tails for the reverse transcription, followed by PCR amplification with additional upstream primers of arbitrary sequences, mRNA sub-populations are visualized by denaturing polyacrylamide electrophoresis (See schematic illustration of differential display). This allows direct side-by-side comparison of most of the mRNAs between or among related cells. The differential display method is thus far unique in its potential to visualize all the expressed genes in a eukaryotic cell in a systematic and sequence-dependent manner by using multiple primer combinations. More importantly, the new method enables the recovery of sequence information and the development of probes to isolate their cDNA and genomic DNA for further molecular and functional characterizations. Because of its simplicity, sensitivity, and reproducibility, the mRNA Differential Display method is finding wide-ranging and rapid applications in developmental biology, cancer research, neuroscience, pathology, endocrinology, plant physiology, and many other fields.

Ejemplo: Se parte de una extracción de mRNA que se pasa a cDNA vía RT-PCR. Para ello se usan *primers* aleatorios de secuencia corta (decámeros) junto con un *primer* de oligodT₁₂MN, y se le agrega la transcriptasa reversa. De cada muestra se hacen cuatro alícuotas, y como *primer* se usa un oligodT12 con dos nucleótidos extras como la G. Sólo las moléculas que tengan dos C (complementarias de las dos G) antes de la

cola poliA serán eficientemente transcritas. En la primera alícuota estará pasando de los RNAs a cDNA que antes tengan C, en las otras alícuotas se usan otros *primers* con un nucleótido adicional G, A o T. Todo esto se hace con las dos muestras en paralelo.

Las moléculas de RNA que estén en A y B se transcribirán en los 2 tubos. Una vez hecho esto, para cada una de las tres parejas de tubos se hace una amplificación y se usa un oligo de secuencia aleatoria de 10 nucleótidos. Se hacen geles y se analizan bandas que estén en A y no en B o viceversa. Esto se repite con más *primers* aleatorios y con todas las alícuotas. Se recomienda usar geles de poliacrilamida para poder ver bien las bandas. La separación eficiente de fragmentos es necesaria para distinguir las bandas de interés. También cabe decir, que esta técnica nos puede indicar, no tan sólo genes que se expresan en uno y en el otro, sino también genes que se expresan más en A que en B o a la inversa. Toda banda candidata se puede recortar del gen y ser verificada usando técnicas convencionales como *northern*, *RT-PCR*. Es una técnica de arreglos para pobres, porque un chip de expresión es muy caro.

Ventajas: Permite comparación múltiple de las muestras. Se puede partir de pequeñas cantidades de RNA que estará limitada por las alícuotas necesarias. Esta técnica puede determinar genes que están expresados diferencialmente entre muestras mediante la intensidad de la banda (**técnica cualitativa**).

Problema: Se presentan muchos falsos positivos, por lo que se recomienda hacerlo por duplicado.

4.2.2. *Representational Difference Analysis (RDA)*

La técnica de RDA, permite comparar muestras de dos genomas e identificar regiones que presentan diferencias entre ellas. Se llama así, porque se usa exclusivamente una fracción del DNA (representativo). Antes de hacer el PCR, se hace una substracción por hibridación con un DNA *tester*. Esto permite amplificar la fracción del DNA que tiene los adaptadores correctos. Se ha usado para comparar DNA de tejido tumoral con DNA del mismo paciente de tejido sano.

Lisitsyn N, Lisitsyn N, Wigler M. (1993), cloning the differences between two complex genomes. Science, 259, 946-951

Metodológicamente, la técnica se basa en **hibridación substractiva** de ácidos nucleicos. Se tienen 2 DNA's, uno es el DNA prueba o **tester** (en el que se pretende identificar las diferencias), el otro es el **driver** (o DNA de

referencia). Inicialmente, ambos DNA son digeridos y ligados a adaptadores. Los adaptadores del *tester* tienen que ser diferentes a los del *driver* (aunque los 2 DNAs hayan sido preparados con la misma enzima de restricción). Así pues, en el *tester* habrá algún fragmento que no se encuentre en el *driver*, o al revés. Por esto se recomienda hacerlo en paralelo y que en uno, el primer caso el *tester* sea *tester* y en el otro el *tester* pase a ser el *driver*. Como digerimos con la misma enzima de restricción, la mayoría de fragmentos tienen una longitud promedio similar. No es recomendable usar enzimas que corten con frecuencias diferentes.

Representational Difference Analysis

RDA

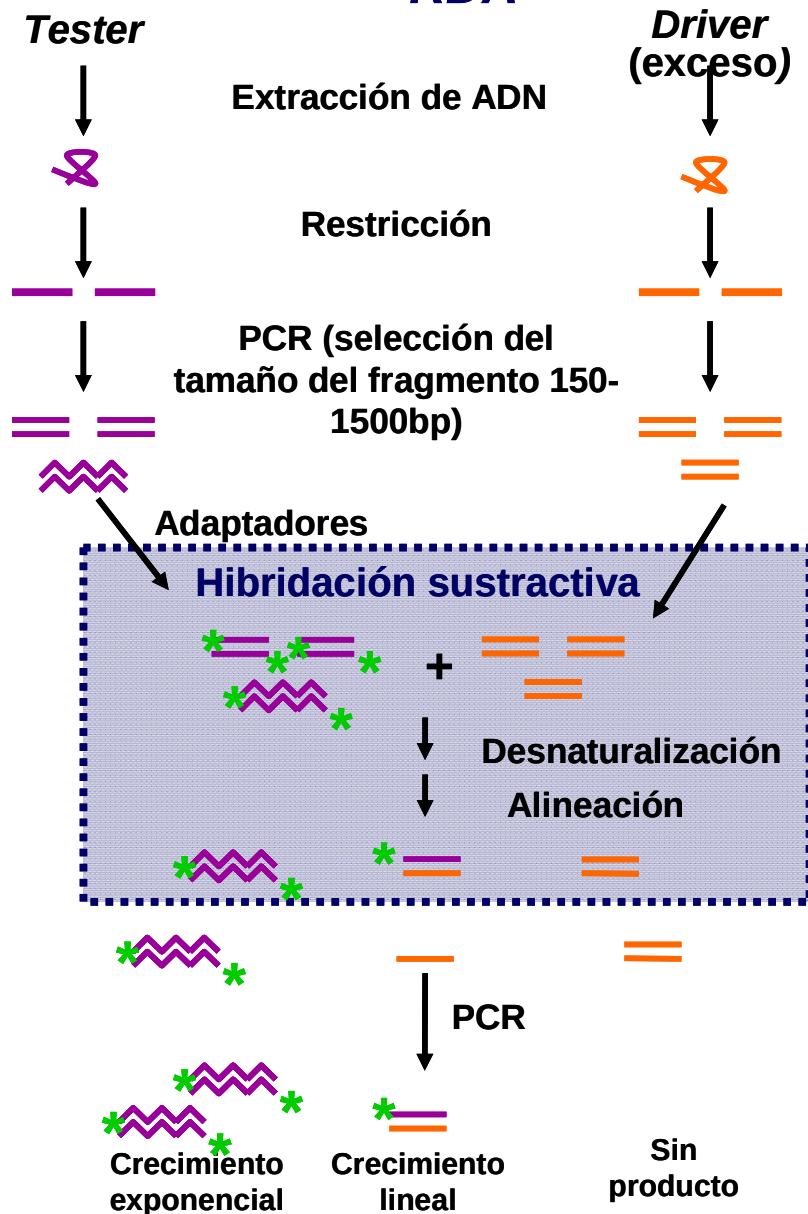


Figura 4-14. Representational Difference analysis (RDA).

Si el DNA viene del mismo individuo no habrá polimorfismos de restricción, pero si los *driver* y *tester* son de individuos diferentes, sí que

podremos obtener fragmentos de diferente tamaño, porque puede haber polimorfismos de restricción. No solamente habrá deleciones e inserciones que nos podemos encontrar si son del mismo individuo. Después de poner adaptadores, se tiene el *driver* en exceso, por lo que habrá hibridación entre la cadena sencilla del *driver* y la cadena sencilla del *tester*. Si el fragmento está en las dos muestras, es más probable que hibride el *driver* con el *tester* que el *tester* con él mismo, porque el *driver* está en exceso. Como resultado, los fragmentos que solo están en el *tester* no encontrarán ninguna cadena *driver* complementaria, y por tanto volverán a hibridar con su cadena complementaria. Esto implica que aquellos fragmentos de *tester-tester* son los que no están presentes en el *driver*.

Se hace un PCR con *primers* complementarios a los adaptadores *del tester*, de tal manera que sólo estos amplifican de manera exponencial. Cabe decir que puede ser que un fragmento que esté en los dos se haya colado porque haya vuelto a hibridarse con *tester* en lugar de *driver*. Por eso es que la hibridación subtractiva y el PCR se tienen que repetir varias veces. Después se corren las bandas en geles, se recortan y se mandan secuenciar. Seguro que habrá una fracción del genoma que no será amplificable por PCR debido a que los fragmentos son más grandes de la cuenta, por lo que se recomienda digerir el DNA con diferentes enzimas.

Ventajas: Es una técnica potente.

Problema: Muchos falsos positivos, pero se pueden disminuir haciendo alícuotas del producto de PCR y hibridarla otra vez con un exceso de DNA *driver*. Así se descartan posibles moléculas inespecíficas que hayan hibridado.

SSH

Es equivalente a la RDA, pero a diferencia de este, se parte de RNA y no sobre DNA genómico. También se habla de *tester* y *driver*. En esta técnica se tienen 2 alícuotas diferentes de *tester*, cada una con adaptadores de secuencias diferentes. Al *driver* no se le liga ningún adaptador. Se desnaturalizan las muestras y se hibrida el *tester A* y *B* con exceso de *driver* en diferentes tubos. Después se juntan ambos tubos, se desnaturalizan y se vuelven a hibridar. De todas las moléculas que se formen, interesarán aquellas que contengan una cadena de *tester A* con 1 cadena de *tester B*. Si ahora se hace PCR, en estos dos tubos, aquellos *tester* que no estén presentes en los *drivers*, no hibridarán con ellos y quedarán en forma de cadena sencilla, de tal manera, que si juntamos los dos tubos e hibridamos de nuevo, el *tester* de los dos tubos que no habían hibridado, hibridarán entre ellos y los que estaban hibridados. Finalmente se hace PCR con un *primer* complementario al adaptador A y otro al B, para que sólo se amplifique el *tester A+B*.

Ventajas: Resultados normalizados, se amplifica cDNA expuesto diferenciado, pequeñas cantidades de RNA.

Problema: No es cuantitativa.

4.2.3. Genome Mismatch Scanning (GMS)

A veces interesa encontrar fragmentos de DNA comparables para dos muestras, y que los hayan heredado de los antecesores. El objetivo de esta técnica es encontrar secuencias idénticas de DNA.

Se debe analizar el **pedigrí** y cuatro o cinco generaciones atrás encontrará al individuo afectado. Probablemente pensemos que han recibido del antepasado ese fragmento de DNA. Así que se necesita una técnica que permita identificar los fragmentos de DNA que estos dos hijos tengan en común (fragmentos iguales,) candidatos de la enfermedad y que por lo tanto, habrán heredado. Además este fragmento compartido no tendrá que tener SNP's porque tan sólo han pasado tres o cuatro generaciones, lo que se considera poco para que surjan SNP's. Se Quiere una colección de fragmentos de identidad por descendiente, *identity by descent* (IBD) que no presenten cambios nucleotídicos. El GMS descarta secuencias con nucleótidos diferentes y se queda con las secuencias iguales (Griffiths, *et al.* 2004)

Se parte de dos DNA genómicos (A y B), de los cuales, uno será metilado y el otro no (tan solo tendrá dos moléculas de DNA genómico metiladas las regiones normales). Enseguida, se utilizan enzimas de restricción como PstI (que puede corta tanto secuencias hipermetiladas como no) se mezclan las 2 digestiones, se desnaturaliza y se renaturaliza. De tal manera el DNA A se hibrida con DNA B. Puesto que nos interesan los DNA mixtos (pero moléculas de igual secuencia), se debe descartar los homohíbridos. ¿Cómo? Aprovechando el mismo grado de metilación que tienen las cadenas homohíbridas. Se digiere con DpnI (reconoce cuatro nucleótidos y sólo corta cuando las dos cadenas están metiladas) y MboI (reconoce la misma secuencia que DpnI y corta las dos cadenas cuando no hay metilación). Usando estas dos enzimas digeriremos homohíbridos. No se usan otras enzimas de restricción, porque son de mayor frecuencia de corte. Los heterohíbridos que no presenten ningún cambio de nucleótidos no se cortarán eficientemente, ni por una enzima ni por la otra y quedarán en un tamaño mayor, por lo que serán moléculas candidatas a ser responsables de esa característica fenotípica. Ahora se compara cuales no presentan *mismatch* de estas moléculas grandes. ¿Cómo se hace esta discriminación? Si se incuban los heterohíbridos con enzimas de reparación estos buscarán *mismatch*, si lo encuentran harán *nicks* en la molécula, y por tanto, si

después de hacer esto quedan fragmentos largos, estos serán candidatos a ser el heterohíbrido sin *missmatch*. Después se requiere recuperar las moléculas de doble cadena grandes que son candidatas a ser IBDs.

4.3. Ejemplos de aplicaciones

4.3.1. Demostración de Pedigrí

Anteriormente, las pruebas de descendencia (entre ellas las de paternidad) se hacían con marcadores microsatélites, pero ahora se está utilizando más los SNP's. Los microsatélites tienen un grado de polimorfismo mas elevado que los SNPs. Con 8 microsatélites es suficiente para discriminar un individuo en las pruebas de paternidad. Se necesitan más de 40 SNPs para lograr una discriminación con el mismo grado de confiabilidad.

Pruebas de paternidad con microsatélites:

Algunos laboratorios recomiendan usar un par concreto de marcadores, para que todo el mundo los haga igual. Suelen ser paneles de 12 marcadores. Cuando se hace la prueba de paternidad y se compara el par padre-hijo, no se puede hacer la exclusión de paternidad solo por un microsatélite que no coincida, como mínimo hacen falta 2. Si medimos 12 microsatélites y sólo sale 1 diferente, ampliaremos la batería de marcadores a comparar hasta 16. La pregunta es: si realmente el individuo es del padre ¿Cómo puede ser que el microsatélite no coincida? Porque hubo mutación en la línea germinal ($P = 10^{-4}$). Hay microsatélites en donde la tasa de mutación es más elevada que las ordinarias. Cuando la DNA polimerasa se encuentra con repeticiones puede haber un deslizamiento en la replicación, lo que no será reconocido por los sistemas de reparación. Los sistemas de reparación de la célula si reconocen una mutación de SNP y en la mayoría de los casos la reparan. Por eso hay una tasa de mutación más elevada en los microsatélites que en los SNPs.

Criterios de selección de microsatélites:

Cuando elegimos marcadores para hacer pruebas de paternidad, se escogen estos de algunas listas estándares. Por ejemplo, para animales se usa la de la *Internacional Society of Animal Genetics* (ISAG). Esta tiene unos 32. Los criterios que se siguen para la selección de marcadores de DNA son:

- Que no den problemas de lectura.
- Puedan amplificar en múltiplex.
- Evitar aquellos que tengan alelos nulos. Estos son individuos que tienen mutación justo donde se empareja el primer, que no tiene nada que ver con el microsatélite. Si el individuo es Aa y pasa esto, uno de los 2 alelos no amplifica y el otro sí: pensaremos que es (AA) y en realidad es (Aa).

- Evitar utilizar marcadores demasiado cercanos en un mismo cromosoma.
- Tratar de que tengan segregación independiente.

4.4. Haplotipos

4.4.1. Desequilibrio por ligamiento

Un haplotipo es un conjunto de alelos ligados por su proximidad en un cromosoma. Los genes que están en loci cercanos (menos de 5 centimorgans) se heredan en bloque ya que son pocas las recombinaciones que se realizan entre ellos. A mayor distancia física entre los genes, mayor tasa de recombinación. El término de haplotipo se emplea para referirse a ese **bloque de polimorfismos ligados**. Por ejemplo, si los abuelos paternos tienen el haplotipo ABCDEF mientras que los abuelos maternos tienen el haplotipo abcdef, los nietos tendrán uno de esos dos haplotipos. Serán muy pocas las veces en que los nietos tengan un haplotipo combinado tipo abcDEF o ABCdef. Solo cuando las distancias entre los genes son mayores a 10 centimorgans es cuando se da un desligamiento entre genes. En inglés a esto se le llama *linkage disequilibrium*. Al estudiar un trozo de DNA, secuenciarlo y alinearlo con una secuencia referencia encontramos varias posiciones polimórficas. Cada una de estas combinaciones de polimorfismos ligados recibe el nombre de haplotipo. En una escala de tiempo evolutiva, los haplotipos se rompen en los alelos individuales que los conforman, por lo que decimos que se heredan de forma independiente e individual. Sin embargo, en escalas de tiempo cortas (menos de 20 generaciones) los polimorfismos se agrupan en haplotipos.

Tarea 4-8

Un mejorador hizo una cruce entre dos padres, después autofecundo la F1 y la F2 para generar una población segregante de familias F3. Usando esa población genotipo las diferentes líneas para 3 genes de interés. Los alelos que detecto eran: Gen 1: A y a, Gen 2: B y b, Gen 3: C y c.

Obtuvo los siguientes resultados:

a b c	a B C	A b c	A b c
A b c	A b c	A B C	a b c
a B C	A b c	A b c	A b c
A b c	A B C	A B C	a B C
a B C	a b c	A b c	A b c
A B c	A b c	a b c	A B C
A B C	a b c	A b c	A B C

En la tabla anterior cada celda representa el genotipo de un individuo de esa población. Un padre tenía el genotipo ABC y el otro padre el genotipo abc. Calcule las distancias genéticas entre los diferentes genes A, B y C en unidades centimorgan.

4.4.2. Proyecto HapMap

El proyecto HapMap fue Iniciado en 2002 con el objetivo de describir las variaciones en la secuencia de DNA de la población humana. Debido a que estas variaciones se dan en bloques, la meta entonces es definir los diferentes haplotipos del genoma humano y de esta forma facilitar el estudio de algunas enfermedades.

Al secuenciar cromosomas de diferentes personas se encontró que en ciertas regiones de genoma humano había combinaciones de SNP's que son más frecuentes que otras. Esto se debe al ligamiento de secuencias de DNA que están muy cercanas y que casi siempre se heredan en grupo. Es decir, los genes no se segregan cada uno independientemente, sino que en bloques. Un haplotipo es un conjunto de polimorfismos que se conservan juntos aún después de muchas generaciones. Por lo regular, un haplotipo abarca de 10 a 1000 genes.

¿Cuántos SNP's son suficientes para darnos información de los diferentes haplotipos? No hace falta genotipar todos los polimorfismos: analizando tan sólo 3 SNP's podremos determinar cuál de los haplotipos tenemos.

4.5. Pros y contras de las herramientas moleculares

4.5.1. Ventajas de las herramientas moleculares

4.5.1.1. Rapidez

El uso de marcadores moleculares se traduce en una rapidez mucho mayor en comparación con los métodos convencionales. Por ejemplo, con marcadores moleculares se pueden lograr la introgresión por retrocruza de un gen en tan solo 3 ó 4 generaciones, mientras que con el método sin marcadores esto mismo llevaría hasta 6 ú 8 generaciones.

Esta rapidez es muy importante para las compañías de semillas, y es por eso que no es de sorprender que las empresas privadas como Monsanto o Pioneer usen los marcadores moleculares muchísimo más que las instituciones de investigación pública como el INIFAP o la Universidad Antonio Narro. Mismo en una institución internacional de mejoramiento como el CIMMYT los marcadores moleculares sólo se usan en casos especiales y de manera esporádica, muy lejos de la rutina masiva que usan las compañías semilleras de maíz. *Time is money*. Muchas veces los

investigadores no tienen mucho dinero de fuentes de financiamiento generosas, y es por que eso que suelen tardarse un poco más para lograr las mismas metas de mejoramiento en comparación con sus colegas en las empresas.

4.5.1.2. Alta heredabilidad

Los marcadores moleculares ofrecen la oportunidad de obtener información de manera más precisa que al medir el fenotipo. El fenotipo es muy variable, ya que depende del ambiente y del estado de desarrollo (ver página 14). Si decimos que el genotipo es una información digital con valores discretos (A, T, G o C), el fenotipo es información análoga con valores continuos que dependen de muchos otros factores y circunstancias (ver página 17). Un marcador molecular por consiguiente, tiene un valor de heredabilidad máximo (cercano a 100%), ya que toda la varianza se debe al polimorfismo y no hay influencia del ambiente. En cambio, el fenotipo puede tener una heredabilidad mucho más baja, por ejemplo el rendimiento de grano bajo sequía suele tener una heredabilidad de menos de 30%.

La heredabilidad es un término que se usa para describir la proporción de la varianza que puede atribuirse a los genotipos, en relación con la varianza fenotípica total, que incluye la varianza ambiental, y el error residual. La heredabilidad puede también entenderse como “repetibilidad” de la medición del carácter de interés, en nuestro caso, el rendimiento de grano bajo sequía.

4.5.2. Limitaciones de las herramientas moleculares

4.5.2.1. Costos

Un argumento muy válido en contra de los marcadores moleculares es que actualmente los costos son demasiado altos. Muchas veces, su uso no se justifica por simples razones económicas. Las herramientas moleculares pueden requerir una inversión inicial muy elevada y un costo de operación muy alta, lo que impide su uso de manera general para todos los cultivos y todas las características. La inversión inicial implica la compra de equipos sofisticados de secuenciación o de genotipaje, que en la mayoría de los casos cuestan más de dos cientos mil dólares. También se requiere de personal calificado y entrenado, ya que las metodologías no son tan fáciles y no siempre funcionan a la primera. El costo por *datapoint* de los marcadores moleculares puede ser de 3 a 10 dólares, lo que es mucho más caro que sembrar varios genotipos en el campo para evaluarlos. Muy pocas instituciones públicas aplican rutinariamente los marcadores moleculares para el mejoramiento genético.

El uso de los marcadores moleculares únicamente se justifica en algunos casos en que las metodologías tradicionales no han funcionado muy bien, o cuando la velocidad es lo mas crítico y el costo no importa tanto (p. ej.

empresas de semillas transnacionales). Uno de los retos más grandes para el futuro es desarrollar metodologías que sean mucho más económicas y eficientes. Para que valgan la pena, los costos tendrían que bajar entre dos a tres órdenes de magnitud.

4.5.2.2. Escepticismo de los mejoradores clásicos

Algunos genetistas clásicos piensan que no es necesario utilizar las herramientas moleculares modernas. Su justificación es que con sus técnicas clásicas han conseguido incrementar la producción sin ninguna necesidad de secuenciar el DNA o usar la información genómica. El éxito histórico del mejoramiento clásico, así como los resultados contundentes que obtienen de forma continua les da la razón. En promedio, el avance genético que se ha obtenido en el mejoramiento de maíz es de 10% por década aproximadamente.

Sin embargo, se debe tener en cuenta que no en todos los caracteres se ha conseguido lo mismo, y que en algunos casos, el progreso genético se podría acelerar al usar la información molecular. Es decir, hoy en día no solo importa el resultado, sino también la velocidad y la eficiencia. Todavía no existen muchos casos documentados que demuestren que con marcadores moleculares se logró un progreso genético mas rápido y mas eficiente que sin los marcadores. La eficiencia hay que medirla en términos de una relación de costo-beneficio.

Pocos fitomejoradores han dicho que ahorraron dinero por usar marcadores. Por otro lado, si hay varios que han invertido bastante dinero en marcadores sin realmente tener muchos beneficios en términos genéticos. Actualmente se están realizando estudios comparativos para demostrar la superioridad de los marcadores, pero todavía no hay mucha evidencia publicada. Hay que recordar que los marcadores moleculares se propusieron hace mas de dos décadas, y sin embargo, hasta la fecha no han dado tan buenos resultados como muchos investigadores prometieron desde el inicio.

Algunos piensan que todavía es cuestión de tiempo demostrar que los marcadores funcionan eficientemente en la práctica de un fitomejorador de campo. Otros piensan que todo lo molecular suena muy bonito e interesante en teoría, pero que ya en la práctica es muy diferente. Algunos han llegado a pensar que los marcadores son una promesa que se quedará en los laboratorios pero que nunca llegará realmente a tener impacto en el campo. Las opiniones varían según las experiencias de los grupos y la participación de las instituciones.

Una de las limitantes más grandes de la aplicación de los marcadores es la brecha tan grande que existe entre los científicos y los diferentes usuarios. Por lo regular los fitomejoradores trabajan solo en el campo, mientras que

los investigadores lo hacen en el laboratorio. Existen muy pocos investigadores que realmente sepan hacer bien las dos cosas: hacer ensayos de campo con evaluaciones fenotípicas por un lado, y por otro lado hacer experimentos de laboratorio y análisis bioquímicos o moleculares. Son actividades que requieren de diferentes destrezas. Hay muy pocas personas con vocación y talento en las dos áreas.

Muchas veces, los dos grupos se echan mutuamente la culpa. Es típico escuchar de algunos investigadores que dicen que los fitomejoradores no hacen caso de los descubrimientos que ellos han publicado en las revistas. Por otro lado, algunos fitomejoradores dicen que los agricultores no se van a comer el papel de las publicaciones. Otros piensan que los moleculares viven encerrados en sus laboratorios y que no saben realmente de que están hablando cuando se trata de los problemas del campo. Es muy fácil hacer promesas en teoría, pero no es tan sencillo cumplirlas en la realidad. Algunos dicen que a partir de un genoma secuenciado ya automáticamente se pueden generar mejores variedades. Otros piensan que el desarrollo de toda la tecnología de secuenciación de los últimos 20 años fue una tarea relativamente sencilla, comparado con el grandísimo reto de generar variedades mejoradas a partir de la información de una secuencia.

“Es un camino mucho más largo y sinuoso... la del gen a la semilla, de la semilla al campo, y del campo al plato, que del modelo de la doble hélice de Crick al genoma secuenciado de Watson”

Cita anónima

Nunca sabremos quien tiene más razón, ni se trata de eso. Lo que si es verdad es que se necesitan más personas calificadas en los dos temas, agrícolas y moleculares, para que un ejército de estudiantes jóvenes, motivados y capacitados se conviertan en verdaderos puentes de unión, para que lleven los logros de la ciencia de vanguardia a todos los rincones del medio rural. Sin esos jóvenes y sin esa educación de calidad, todos los sueños de la biotecnología agrícola se quedarán solo en papel.

4.5.2.3. El vínculo entre el laboratorio y el campo

Hay muchos científicos que saben hacer PCR. Hay otros que saben tomar datos agronómicos. Sin embargo, no existen muchos científicos que quieran hacer las dos cosas, aislar un gen y calificar mazorcas. Hay muchos estudiantes que prefieren ponerse la bata y los guantes, y muchos mas los que deciden ponerse las botas y el sombrero para trabajar bajo el sol. Hay muy pocos que saben hacer las dos cosas. Y mucho menos los que realmente son buenos en las dos áreas. Necesitamos más científicos que puedan publicar en las más altas revistas científicas, y a su vez estén generando semilla mejorada de alguna variedad vegetal.

Pareciera que existe una barrera psicológica entre los científicos y los mejoradores que evita que los dos se complementen de una manera eficiente. Por un lado los científicos reclaman: “es que los mejoradores no aplican nuestros conocimientos sobre el DNA”. Por otro lado los agrónomos contestan: “es que los científicos investigan solo lo que les interesa. Después de graduar y publicar, ya no se esfuerzan en aterrizar las cosas”. A su vez los científicos responden “yo publico en revistas para que los demás lean mis artículos y apliquen el conocimiento”. El problema es que a veces no hay gente capacitada para hacerlo, y otras veces, los resultados que se generaron son demasiado costosos para una aplicación de campo.

Los científicos se guían por el interés y se motivan por la curiosidad de responder preguntas, mientras que los mejoradores se guían por la necesidad y la urgencia de resolver problemas. Un científico busca contestar las preguntas de una manera novedosa, mientras que un mejorador trata de resolver un problema de la manera más rápida y con el menor costo posible. Si un científico no responde la pregunta inicial, no importa porque habrá aprendido mucho, y tal vez haya contestado muchas otras preguntas muy interesantes. Si un mejorador no logra su objetivo de incrementar el rendimiento, o la calidad, entonces ha fracasado. Es decir, la eficacia y la eficiencia son parámetros primordiales para el mejorador, mientras que para el científico no tanto.

Un ejemplo muy característico son los marcadores moleculares. Se pueden realizar estudios sofisticados sobre la función de los genes para identificar QTLs y generar marcadores moleculares. Eso se puede publicar muy bien. Sin embargo, muy pocas veces resulta útil esa información para el mejoramiento en términos prácticos, debido a que los QTLs son específicos de las poblaciones experimentales que se utilizaron, y por lo regular no aplican para las generaciones mas avanzadas de los materiales elite de los mejoradores. La complejidad de algunas características, como el rendimiento de grano bajo sequía, es tal que resulta muy poco eficaz utilizar marcadores para un gen que solo contribuye con menos del 1% de la variabilidad genética.

Además, todo se complica aun más, con los fenómenos de heterosis, epistasis, interacción genotipo-ambiente, etc. Bajo esas circunstancias resulta mucho más rápido, barato y eficaz confiar en un buen fenotipo y un análisis bioquímico, que gastar demasiados recursos en el genotipo y selección por marcadores. En teoría, los marcadores suenan muy bien, pero ya en la práctica resultan más costosos y menos confiables para lograr los objetivos de mejoramiento deseados. En teoría, la secuenciación de un genoma es una noticia fabulosa, pero en la práctica, no se necesita esa información para hacer mejoramiento. Más serviría desarrollar protocolos de fenotipo más eficientes, confiables y baratos. Sin embargo, no hay suficiente dinero, ni apoyo para esos trabajos, ni tampoco hay recursos humanos capacitados.

Hoy en día es muy fácil secuenciar miles de genes. Lo que es difícil es encontrar la forma de aprovechar la información nucleotídica para generar una variedad mejorada. Lo que se necesita es una oleada de estudiantes inteligentes y motivados para que se pongan a trabajar en esos temas para que resuelvan los problemas que están pendientes. Lo que se requiere es formar un ejército de futuros fitomejoradores para que trabajen con los diversos cultivos, y se enfoquen en desarrollar variedades adaptadas para los diferentes suelos, climas y ambientes. Sólo así se lograrán avances

significativos para satisfacer las necesidades de los agricultores y los consumidores.

5. Genómica y regulación

5.1. Antecedentes de la genómica

La genómica surgió en 1990 con el proyecto del genoma humano (HUGO). El objetivo era caracterizar todos los genes de la especie humana. Se sabía que algunas mutaciones en determinados genes eran responsables de algunas enfermedades o padecimientos. Debido a que los genes sólo representan una pequeña parte del genoma, secuenciar todo el DNA, incluyendo los fragmentos repetitivos, implicaba un gran esfuerzo de tiempo y dinero. El interés de curar y diagnosticar algunas de estas enfermedades fue lo que finalmente movilizó los recursos financieros para lograrlo. El objetivo era obtener tanto el mapa genético, como el mapa físico y la secuencia de DNA. Se redujeron los costos y aumentó la velocidad de secuenciación.

Después de HUGO se trataron de incluir a muchas más especies diferentes. Se introdujo el concepto de especies modelo. El primer genoma bacteriano secuenciado fue el de *Haemophilus influenzae*. En 1996 se secuenció la levadura *Saccharomyces cerevisiae*. Este fue el primer eucariota secuenciado con 6,000 genes. En el 1998 se secuenció un nemátodo (*Caenorhabditis elegans*) que ya posee diferenciación celular. En el 2001 se publicó el genoma de *Arabidopsis* con 23 mil genes.

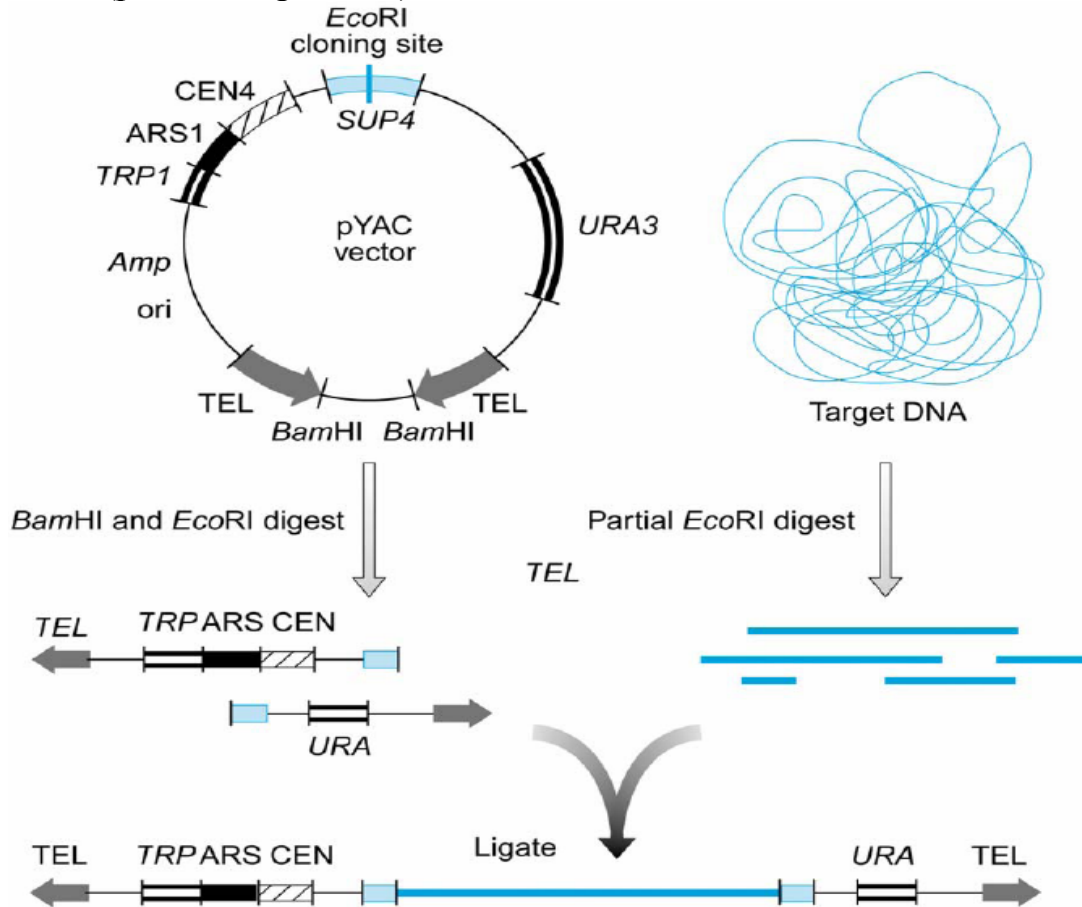
Entre los genomas secuenciados figuran los cloroplastos, las mitocondrias de mamíferos, diferentes organismos de laboratorio como levaduras, nematodos, la mosca de la fruta, *Drosophila melanogaster* la planta *Arabidopsis thaliana*, el ratón, *Mus musculus*. También se han secuenciado una gran variedad de patógenos, como las bacterias causantes del cólera, la tuberculosis, la sífilis, la gonorrea, la enfermedad de Lyme, el virus de la viruela y el del mal de Epstein-Barr. Con la conclusión de estos y muchos otros proyectos genomas se espera lograr una mejor comprensión sobre el mecanismo de acción de los organismos patógenos, para disminuir su virulencia, o lograr erradicarlos.

Los proyectos genómicos se pueden dividir en tres fases:

- **Genómica estructural:** Es una caracterización física. Por ejemplo ¿qué tamaño?, ¿cuántos cromosomas?, ¿cuántos genes?, ¿cuántas secuencias repetitivas? Se desarrollan mapas genéticos, mapas físicos, BACs, marcadores, se averigua como están organizados los genes en clusters, centrómeros, telómeros, bandeo, etc.
- **Secuenciación:** Es la determinación del orden de las letras del DNA. Esto incluye la secuenciación (ya sea por Sanger, Pirosecuenciación, Solid o Solexa), el ensamblado bioinformático en *contigs* y

almacenaje de las secuencias en bases de datos. La secuenciación se puede hacer en fragmentos por azar (método *shot-gun*) o de manera organizada por YACs, BACs o *contigs* (método tradicional).

- **Genómica funcional:** Ya una vez que se tiene la secuencia, se tienen que identificar los genes (exones e intrones), y después anotar las proteínas con sus respectivas secuencias regulatorias. Después se debe de determinar la función de cada gen por medio de mutantes, transgénicos, expresión heteróloga, etc. Finalmente se pretende entender la función de cada uno de los elementos del genoma (promotor, gen, etc.).



Figuras 5-1. Yeast Artificial Chromosomes. PYAC con DNA blanco

5.2. Estrategias de ensamblaje genómico

Para la secuenciación de un genoma de manera tradicional, primero se tienen que obtener clonas que estén ordenadas dentro de un mapa físico. Primero necesitamos tener el genoma clonado en forma redundante en cromosomas artificiales. Después, necesitamos obtener un mapa físico preferentemente de traslape mínimo, “*minimum tilling*” y finalmente, necesitamos obtener la secuencia de cada una de las clonas elegidas.

Otra forma es buscando los pasos apropiados “*chromosome walking*”. Comenzando con un clon, se clona solamente un extremo y se busca en la biblioteca otro(s) que traslapen. Este proceso se repite, avanzando “pasos” en el cromosoma.

Los mapas de restricción de los vectores son una alternativa al “*Chromosome walking*”. Se pueden hacer mapas de restricción de los vectores para identificar secciones parcialmente traslapadas. Con ello se puede ordenar la biblioteca de manera que se obtengan grupos traslapados que pertenezcan al mismo cromosoma.

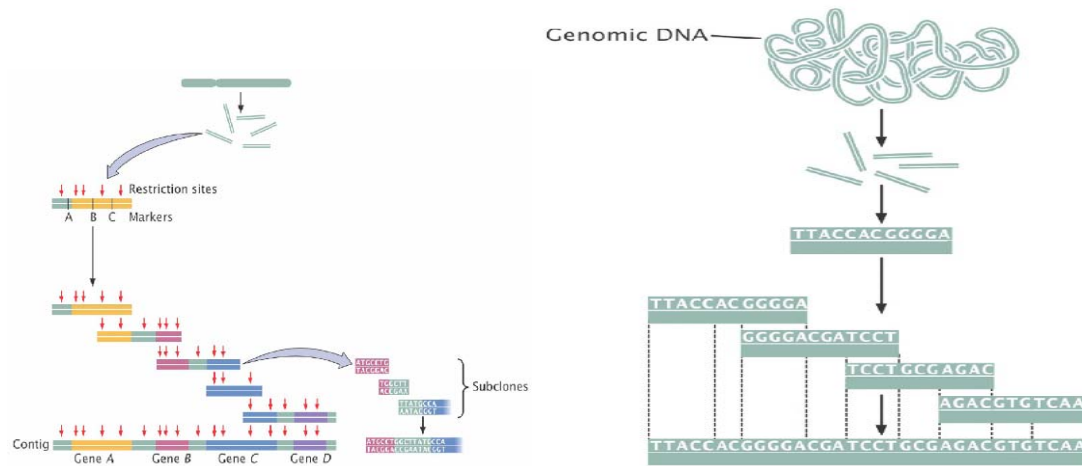
Los vectores son plásmidos, microcromosomas bacterianos. Sus principales características es que tienen un origen de replicación y son estables, es decir deben poseer secuencias que permitan su propagación dentro de una célula bacteriana. Los vectores tienen sitios específicos para insertar el DNA a clonar y sitios con secuencias únicas para varias enzimas de restricción. Los vectores también deben tener marcadores de selección; que confieren propiedades fenotípicas que permiten seleccionar a los portadores.

Se tiene varios tipos de vectores clonados, tales como:

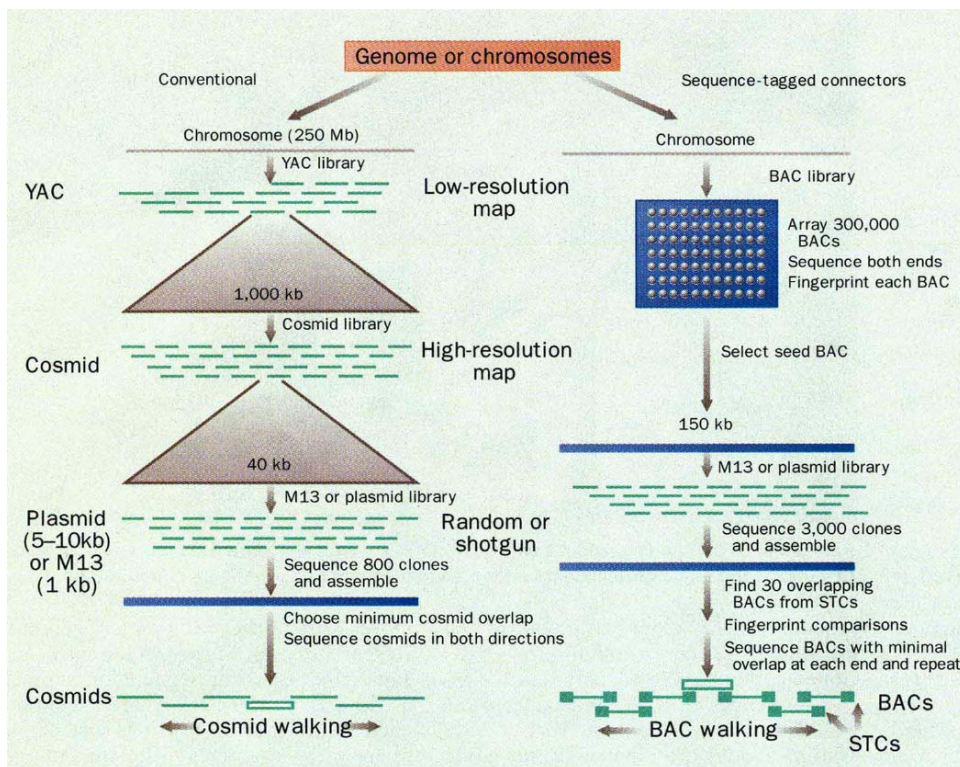
- a) Plásmidos; cuyo límite de clonación es de 0.1 a 10 kb.
- b) Fagos; derivados del bacteriófago lambda, de 8 a 20 kb.
- c) Cósmidos; combinación de fago I y plásmido, de 35 a 50 kb
- d) BAC: por sus siglas en inglés *Bacterial Artificial Chromosome*, derivado de plásmido F (plásmidos de fertilidad) de *E. coli*, que interviene en el control de la replicación. El sitio de clonación en el gen LacZ (selección azul/blanca). Su transformación es por electroporación. De 75 a 300kb.
- e) Vectores derivados de bacteriofago P1; de 100 kb.
- f) YAC; por las siglas en inglés, *Yeast Artificial Chromosome*. Cromosoma artificial de levadura,

Hablaremos un poco más de los YACs. Son segmentos largos (más de 100 kb), se ligan con los dos brazos, y cada brazo termina con un telómero de hongo, de manera que el producto puede ser estabilizado en la célula fúngica. Los YACs más largos son también más estables. Los brazos tiene una secuencia de replicación autónoma (ARS), un centrómero (CEN) y un marcador de selección (*trp1*), el otro brazo tiene otro marcador de selección (*ura3*) esto asegura que la célula ha recibido un cromosoma artificial con ambos telómeros, dada la complementariedad de los mutantes, y el cromosoma artificial el ADN insertado. Los problemas con los YACs: El más importante es que de 40 al 60% de ellos sufren rearrreglos, el 40% de los YACs se pierden, tienen baja eficiencia de transformación y son extremadamente difíciles de manipular

Figura 5-1 Esquema del vector YAC en *Escherichia coli*.



Figuras 5-2. Secuenciación basada en mapas y secuenciación por *Shotgun*.



Figuras 5-3. Método de Shotgun. Figura tomada de los artículos de Craig Venter sobre *shotgun* (Nature 1996)

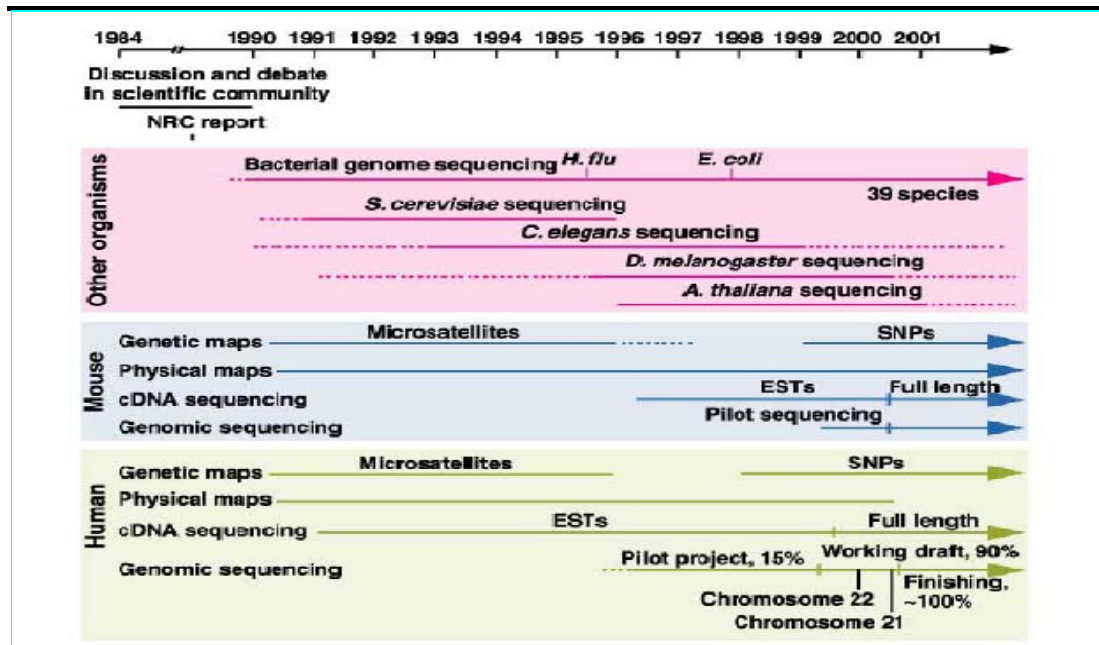


Figura 4 2: Historia de los mapas genéticos y secuenciación genómica.

5.3. Ejemplos de proyectos de genómica

Son proyectos de caracterización global de los genes de una especie. Se trata de determinar la secuencia, el número de genes, la posición de los mismos dentro de los cromosomas, la estructura intrón-exón, las secuencias regulatorias, promotores, la función de los genes, su patrón de expresión en diferentes tejidos, estados de desarrollo e influencia de los factores ambientales.

Los organismos modelo que se han escogido para los proyectos de genómica suelen ser aquellos con un genoma relativamente pequeño, ya que entre menos nucleótidos, mas barato resulta secuenciarlos.

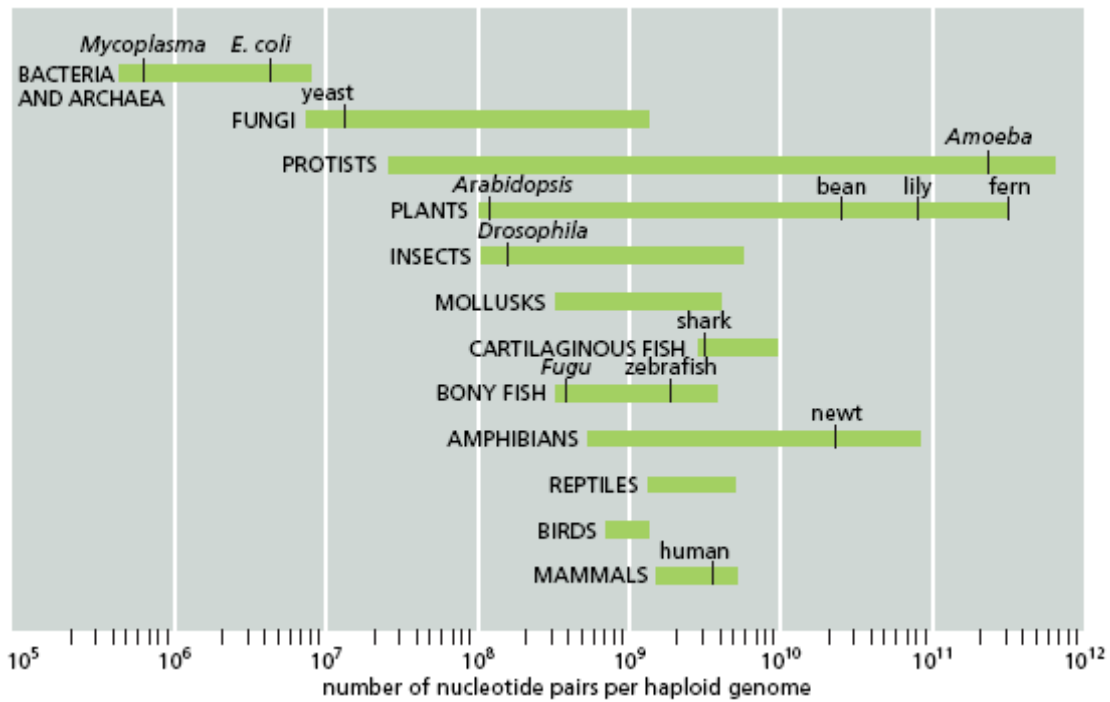
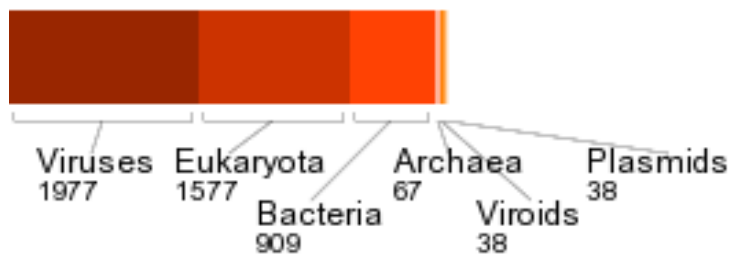


Figura 5-2 Tamaño del genoma haploide de varias especies

Total species (4606)



Total records (9726)

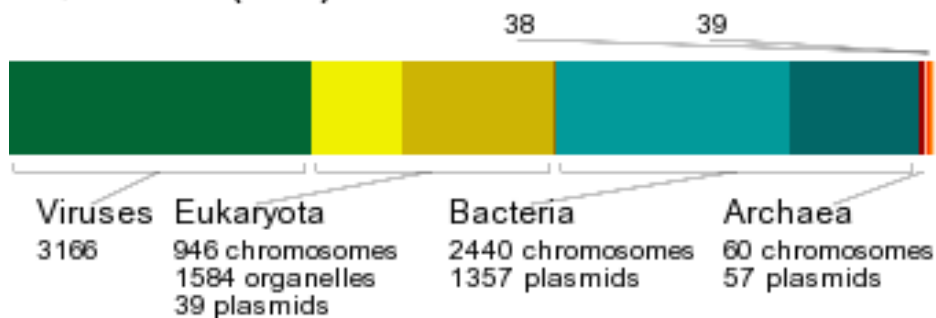
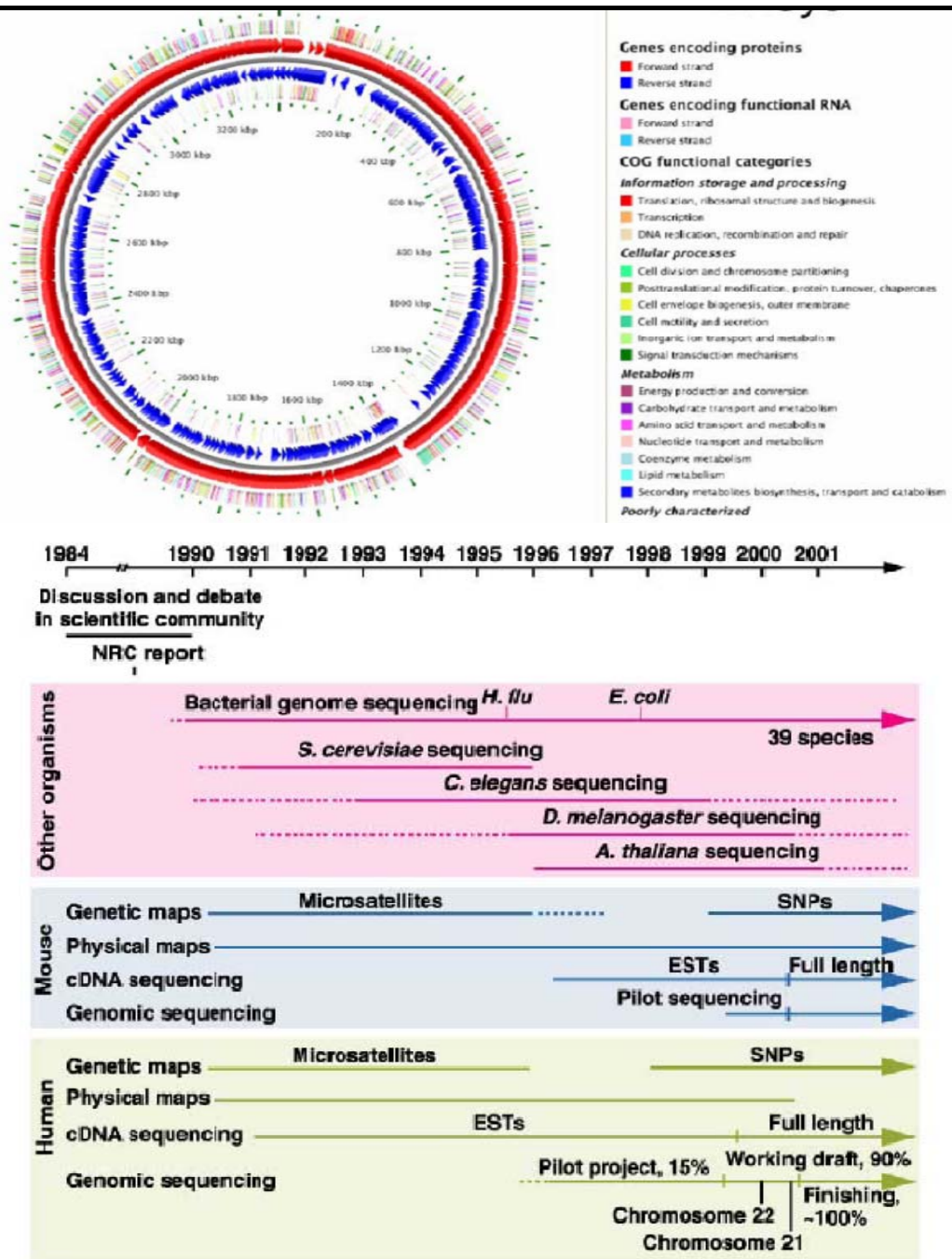


Figura 5-3 Número de genomas completos en las bases de datos de Genbank



Tarea 5-1:

Usando los datos de las figuras anteriores conteste las siguientes preguntas:

¿Qué son los dominios taxonómicos?, ¿Por qué en la actualidad se usan más que los 5 reinos?, ¿Cuáles son estos? y ¿cuáles son las características de cada uno?

¿En qué tipo de organismos existen más genomas completos? ¿Puede usted postular algunas razones de ello?

5.3.1. Genomas vegetales

5.3.1.1. Arabidopsis

Tarea 5-2:

Lea el siguiente texto en inglés y conteste las siguientes preguntas:

¿Por qué se escogió *Arabidopsis* como modelo genómico?

¿Cuántos genes nuevos se agregaron en la versión TAIR8?

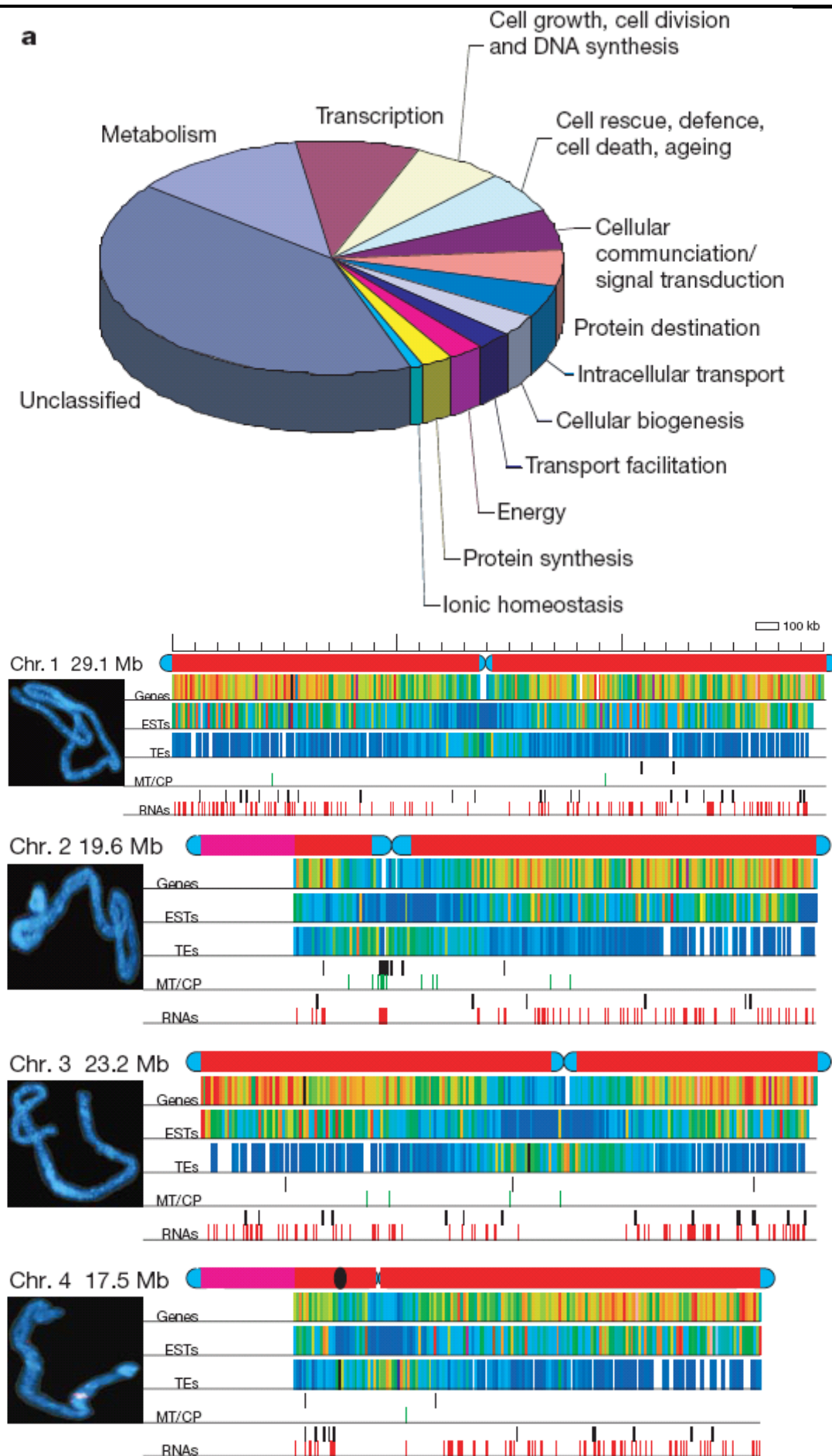
Arabidopsis thaliana is a small flowering plant of mustard family, brassicaceae. It was selected as a model organism for genome sequencing in plants based on the fact that it has (1) a small genome of ~120 Mb with a simple structure having few repeated sequences and high gene density (2) short generation time of six weeks from seed germination to seed set (3) produces large number of seeds, and (4) is easy to transform. The sequencing was done by an international collaboration, collectively termed the Arabidopsis Genome Initiative (AGI). It consisted of research groups in the U.S., Europe and Japan. The project was initiated in 1996 and completed in 2000 (Nature, 408:796-815).

The *Arabidopsis* Information Resource (TAIR) has announced the release of the latest version of the Arabidopsis genome annotation (TAIR8) at TAIR and NCBI. The latest release builds upon the gene structures of the previous TAIR7 release.

The TAIR8 release contains 27,235 protein coding genes, 4759 pseudogenes or transposable elements and 1288 ncRNAs (33,282 genes in all, 38,963 gene models). A total of 1291 new genes and 2009 new gene models were added. Thirteen percent (4330) of Arabidopsis genes now have annotated splice variants. Updates were made to 3811 gene structures of which 625 gene models had CDS updates; a total of 4007 exons were modified and 683 new exons incorporated. There were 33 gene splits and 41 gene merges. Overall 23% of all existing TAIR7 genes (7380 genes) were updated (includes split, merged and deleted genes as well as locus type changes, structural updates and sequence updates) for TAIR8. Changes to chromosome sequences include removal of vector/Ecoli contamination and 1425 single nucleotide substitutions. Additional details can be viewed at the TAIR website.

Texto tomado de:

http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=3702



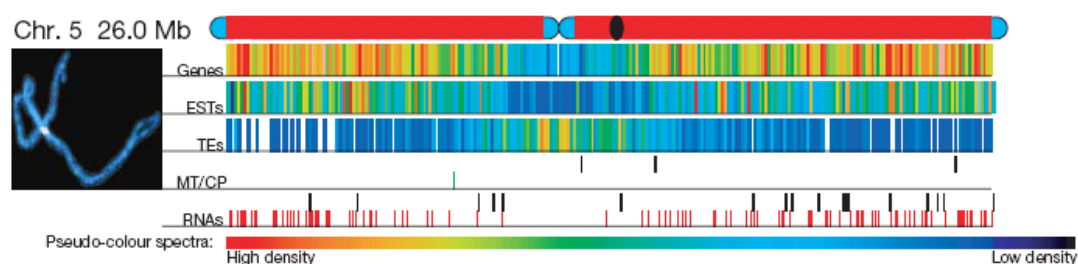


Figura 5-4. Representación de los cromosomas de arabidopsis. Figura tomada de Nature vol 408 p797

Each chromosome is represented as a coloured bar. Sequenced portions are red, telomeric and centromeric regions are light blue, heterochromatic knobs are shown black and the rDNA repeat regions are magenta. The unsequenced telomeres 2N and 4N are depicted with dashed lines. Telomeres are not drawn to scale. Images of DAPI-stained chromosomes were kindly supplied by P. Fransz. The frequency of features was given pseudo-colour assignments, from red (high density) to deep blue (low density). Gene density ('Genes') ranged from 38 per 100 kb to 1 gene per 100 kb; expressed sequence tag matches ('ESTs') ranged from more than 200 per 100 kb to 1 per 100 kb. Transposable element densities ('TEs') ranged from 33 per 100 kb to 1 per 100 kb. Mitochondrial and chloroplast insertions ('MT/CP') were assigned black and green tick marks, respectively. Transfer RNAs and small nucleolar RNAs ('RNAs') were assigned black and red ticks marks, respectively.

Table 4 General features of genes encoded by the three genomes in *Arabidopsis*

	Nucleus/cytoplasm	Plastid	Mitochondria
Genome size	125 Mb	154 kb	367 kb
Genome equivalent/cell	2	560	26
Duplication	60%	17%	10%
Number of protein genes	25,498	79	58
Gene order	Variable, but syntenic	Conserved	Variable
Density (kb per protein gene)	4.5	1.2	6.25
Average coding length	1,900 nt	900 nt	860 nt
Genes with introns	79%	18.4%	12%
Genes/pseudogenes	1/0.03	1/0	1/0.2–0.5
Transposons (% of total genome size)	14%	0%	4%

5.3.1.2. Arroz

Tarea 5-3

Lea el siguiente texto en inglés y conteste las siguientes preguntas:

¿Cuántas veces más grande es el genoma de arroz comparado con el de *Arabidopsis*?

¿Cuántos proyectos genómicos de arroz existen actualmente?

Rice is one of the most important food crops in the world and feeds more people than any other crop. The international community decided to use rice as the model monocot plant to study cereal genomics for two reasons: 1) the

small genome size, ~450 Mb, and 2) the understanding that many of the commercially important cereals have genomes that are very syntenic with rice. There have been commercially funded efforts as well as publicly funded efforts to determine the nucleotide sequence of rice. The first commercial effort, by Monsanto, resulted in a database of genomic sequence and SSR objects. The genomic sequence has been shared with the International Rice Genome Sequencing Project members. The second commercial effort, a collaboration between Myriad Genetics and Syngenta, used the whole-genome shotgun method. Despite a press release in January 2001 noting the completion of the "rice genome map", little if any genomic sequence from this project has been deposited into the public domain. A response to the impediments and shortcomings of this is available.

The two publicly funded rice genomic sequence projects have each produced draft genome sequence. The International Rice Genome Sequencing Project is a collaboration of publicly funded investigators. The project began at the 1997 meeting of the International Symposium on Plant Molecular Biology in Singapore and is described. All of the PAC and BAC libraries for the effort have been produced from DNA prepared from seeds from a single plant of *Oryza sativa* ssp. *japonica* c.v. Nipponbare. The work has been apportioned among the various collaborators. The current status of sequence acquisition by the contributors to the sequence acquisition is routinely updated at the coordination organization, IRGSP. As the Monsanto sequence is brought to phase 2 qualities by IRGSP members, this material is published in public databases. The nucleotide sequence for twelve chromosome assemblies has been deposited in DDBJ by IRGSP. Various groupings of subsets of the IRGSP have deposited in GenBank the nucleotide sequence for four alternative assemblies.

The other publicly funded project, directed by the Chinese Academy of Sciences, Beijing, China, is to sequence the genome of *O. sativa* ssp. *indica* and *O. sativa* ssp. *japonica* by the whole genome shotgun (WGS) method. This project began in April 2000, as a WGS project to sequence the *Oryza sativa* L. ssp. *indica* genome. The first peer-reviewed report was in the April 5, 2002 issue of Science. This has been deposited at GenBank/EMBL/DDBJ under the project accession AAAA00000000. The revision history for this data is available. Sequence data are retrievable through Entrez. The collection of contigs is available as a BLAST database for public query. This project has expanded to include the WGS sequencing of the *Oryza sativa* L. ssp. *japonica* genome. This has been deposited at GenBank/EMBL/DDBJ under the project accession AACV00000000. The revision history for this data is available. Sequence data are retrievable through Entrez. The collection of contigs is available as a BLAST database for public query.

Texto tomado de:

http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=39947

5.3.1.3. Maíz

El maíz no solo es una especie económicamente importante, sino que también puede ser un buen modelo para los estudios de evolución genómica. Existe mucha información sobre esta especie debido a que muchos genetistas clásicos y moleculares han trabajado con esta especie desde hace más de un siglo. Si bien su secuencia completa no ha sido aun

publicada, ya se han identificado gran parte de los genes. No fue secuenciado con anterioridad debido al gran tamaño de su genoma. El maíz tiene 10 cromosomas ($2n=20$).

Tarea 5-4:

Lea el siguiente texto en inglés y conteste las siguientes preguntas:

¿Cuáles es el tamaño mínimo del genoma de maíz? ¿Cuántos mapas genéticos de maíz están públicamente disponibles actualmente?

The maize nuclear genome is thought to be amphidiploid - a fusion of two diploid genomes (Helentjaris T, Weber D, Wright S. Genetics 1988 118:185-188) - or allopolyploid - (Wilson et al. 1999) with knob heterochromatin. The knob heterochromatin has been characterized as a duplication of a 185-bp element (Peacock et al. Proc Natl Acad Sci, USA 1981; 78:4490-4494) and a 350-bp element (Ananiev et al. 1998) with the amount of duplication varying up to 3 to 4 orders of magnitude between different races of corn. This leads to a wide range of genome sizes among corn. The minimum size is estimated to be 2400 Mbp with a high proportion of repetitive DNA. At a NSF sponsored workshop it was decided that sequencing maize genes and placing them on a cross-referenced physical-genetic map is important to improving the agronomic characteristics of the plant. The effort to determine the nucleotide sequence of the *Zea mays* subspecies *mays* (maize) has begun.

Currently there are seven genetic maps available. The data for six maps were provided by MaizeDB. The data are currently available from [MaizeGDB](#). These six are 1) the IBM interim genetic map of 2000, 2) version 2 of the IBM genetic map from 2002, 3) the Emerson genetic map, 1935, and the 4) UMC-98 map with the two maps with map coordinates dependent upon the UMC-98 map, 5) QTL map and 6) bin-mapped loci. [Gardiner J et al. \(1993\)](#) described the UMC-98 map and the use of loci on the map to create a framework to localize non-Mendelian loci or incompletely positioned Mendelian loci. The IBM map was presented in [Sharpova N et al. \(2001\)](#). The initial data for the 1999 Wilson maize map [Wilson WA et al. (1999)] was provided by RiceGenes with input from MaizeDB/MaizeGDB used to complete the display. As data for other maps are requested or new map data are made available, these maps will be displayed.

Texto tomado de:

http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=4577

Un pariente del maíz, el sorgo, tiene un contenido mucho menor de DNA (800Mb). Gaut y Doebley (1997), utilizando 14 pares de genes que se encuentran duplicados en el maíz, lograron fechar la duplicación hace entre 16,5 y 11,4 millones de años. Sin embargo, esta poliploidización no explica totalmente el incremento del genoma del maíz.

En el maíz son muy abundantes los elementos móviles (McClintock, 1946; Fedoroff, 2000). Estos elementos móviles contribuyen al gran tamaño del genoma del maíz. Por ejemplo, Hake y Walbot (1980) estimaron que entre

60 y 80% del total del genoma del maíz se encuentra constituido por elementos móviles, alrededor del 20% está formado por elementos altamente repetitivos con hasta 80,000 copias, y 40% está dado por elementos moderadamente repetitivos, con más de 1000 copias.

Un estudio detallado de la dinámica de los elementos móviles fue realizado por San Miguel *et al.* (1998) en el cual analizaron una región de 240kb que incluye al gen *Adh1*. Alrededor del 62% del total de esta sección está formado por transposones y 6% por repeticiones de bajo número. En esta sección del genoma encontraron 23 retrotransposones pertenecientes a 10 familias distintas. De estos 23 elementos, 10 se encontraban a su vez insertos dentro de otro elemento. La edad del elemento móvil más antiguo en esta región, según aproximaciones de reloj molecular, sería de hace unos 5.2 millones de años, aunque el promedio es de alrededor de 3 millones de años. Dado que esta región en el sorgo carece de retrotransposones, se considera que la proliferación de éstos en el maíz ha sido reciente y posterior al evento de tetraploidización. Los datos con que se cuenta indican que la situación de la región de la *Adh1* es representativa de la mayor parte del genoma del maíz. Así estimaron que el 50% del total del genoma del maíz está constituido por las familias de retrotransposones que se encuentran en esta región, y que su distribución a lo largo de los diferentes cromosomas es bastante uniforme. De hecho, tres de estas familias representaría algo así como el 25% del total del genoma de esta planta {Aldridge, 2009 #176}.

Gaut *et al.* (2000) y Tenaillon *et al.* (2001) realizaron experimentos para explorar a fondo la variación genética a nivel de secuencias en el maíz. Como ya se mencionó, los niveles de variación genética son muy altos en esta especie. Así, en el cromosoma 1 en 25 colectas, encontraron una $q = 0.0144$ para 4 genes cerca del centrómero y una $q = 0.0170$ para 11 genes más alejado del centrómero. q es un estimado de $4N_e\mu$ que depende del total de sitios polimórficos y del largo de la secuencia, y si la evolución del gen es neutra, debe de ser igual a p y se interpreta de manera parecida, (Cummings y Clegg, 1998).

Otro patrón interesante en el maíz y sus parientes silvestres cercanos del mismo género y conocidos como **teosintles**, es la falta de congruencia entre las filogenias obtenidas para diferentes genes. Por ejemplo, las secuencias derivadas de plantas de diferentes localidades pertenecientes al teosintle *Zea luxurians* forma un grupo muy bien definido con los genes *glb1* y *adh1*. Estos genes se encuentran cercanos en el cromosoma 1, pero las mismas plantas presentan secuencias más parecidas a las de otros taxa del género *Zea* para la *adh2* (del cromosoma 4) y para *c1* (del cromosoma 9), sugiriendo una historia evolutiva compleja mediada por diferentes tipos de

selección natural y patrones de flujo génico, que posiblemente sean comunes en la mayor parte de las angiospermas con polinización cruzada.

5.3.1.4.Sorgo

El sorgo no solo es una especie económicamente importante, sino que también puede ser un buen modelo para...

Tarea 5-5

Lea el siguiente texto en inglés y conteste las siguientes preguntas:

¿De que tamaño es el genoma de Sorgo? ¿Qué características agronómicas del sorgo son interesantes?

Sorghum is the fifth most important cereal crop in the world and third in the United States. Some members of this genus are weeds but the domesticated varieties are used as human food and fodder. It exhibits C4 photosynthetic pathway. It is one of the most drought-tolerant crop that grows well in semiarid regions. It exhibits deep root system and thick leaf wax that help it survive hot dry conditions. *S. bicolor* is native to Africa and is grown throughout the world. Certain varieties are sweet where grain and stem are used for making syrup, jaggery and sugar.

Sorghum has a small genome size of about 760 Mb with $2n=20$ chromosomes, which falls between the genomes of rice (~420 Mb) and that of the other crops with larger and more complex genome. This makes Sorghum a good model for comparative genomic studies between rice and sorghum and for structural genomics analysis.

Sorghum is also an interesting model to study drought tolerance as it is known to grow even if water becomes limiting showing a "stay-green" phenotype. It is currently under investigation at Texas A&M University as part of an NSF funded project.

Sequence the genome of sorghum will not only help develop better sorghum varieties but will also help in understanding the genomes of maize, millet and sugarcane. Sugarcane is a close relative to sorghum and is an important biofuel source and also an important source of sugar.

Tarea 5-6

Consiga uno de estos artículos sobre sorgo y lealo:

Paterson AH et al., "The Sorghum bicolor genome and the diversification of grasses.", *Nature*, 2009 Jan 29;457(7229):551-6

Saski C et al., "Complete chloroplast genome sequences of *Hordeum vulgare*, *Sorghum bicolor* and *Agrostis stolonifera*, and comparative analyses with other grass genomes.", *Theor Appl Genet*, 2007 Aug;115(4):571-90

Bowers JE et al., "A high-density genetic recombination map of sequence-tagged sites for sorghum, as a framework for comparative structural and

evolutionary genomics of tropical grains and grasses.", *Genetics*, 2003 Sep;165(1):367-86

5.3.1.5. Tomate

Tarea 5-7:

Lea el siguiente texto en inglés y conteste las siguientes preguntas:

¿Ya se secuenció totalmente el genoma de tomate? ¿Qué países están participando?

¿Entre que especies de tomate se hicieron cruza para determinar QTLs y para hacer mapas genéticos?

The genus *Solanum* includes the cultivated tomato, *S. lycopersicum*, together with its wild relatives. Tomato has long been used as a model system for plant genetics. It has been used for a number of firsts in molecular genetics of plants. The use of RFLP (restriction fragment length polymorphism) to generate a linkage map of a complete plant genome was done with tomato (Bernatzky and Tanksley. *Genetics* 1986; 112:887-898). Tomato was the organism where quantitative traits were resolved into Mendelian factors ([Paterson et al.](#)). The use of allelic variation of RAPD (random amplification of polymorphic DNA) to facilitate map-based walking to agronomically pertinent loci was first done in tomato ([Wing et al.](#) and [Martin et al.](#)). The creation of a high-density molecular linkage map and the use for comparative genomics were done with tomato and potato ([Tanksley et al.](#)). There is an NSF-funded effort to create an ordered collection of BACs that is to be superimposed upon the tomato genetic map. This resource will more rapidly allow access to the underlying genetic and functional basis of the extensive collection of alleles of already known loci affecting fruit development and ripening, interactions between plants and pathogens, and other biological processes of plants.

International consortiums of researchers as part of the Solanaceae Genomics Project are in the process of sequencing the gene rich regions of all 12 chromosomes on a BAC-by-BAC basis and annotate it. In US, majority of the research is funded by NSF and USDA-NRI Plant Genome program. It will be sequencing chromosomes 1, 10, and 11. Other participating member countries are Korea (chromosome 2), China (chromosome 3), UK (chromosome 4), India (chromosome 5), The Netherlands (chromosome 6), France (chromosome 7), and Japan (chromosome 9).

Map Viewer Data

There are four maps currently available. The gene genetic map is from the data presented by [Tanksley et al.](#) The other three are maps built with loci where the allelic state is determined by examining the nucleic acid of the genes. These maps are built using backcross populations from an interspecific cross between *Solanum lycopersicum* and *Solanum pennellii*. For all the maps the data either are publicly available or have been published in a peer-reviewed journal. The 1986 map data are from Bernatzky and Tanksley (*Genetics* 1986; 112:887-898). The 2000 map data are entirely from the [Solanaceae Genome Network](#).

Texto

tomado

de:

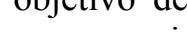
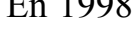
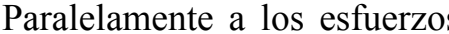
http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=4081

Tarea 5-8:

El nombre científico del tomate era *Lycopersicum esculentum*, mientras que hoy en día se le trata de llamar *Solanum lycopersicum*. Haga una búsqueda bibliográfica y averigüe más sobre el tema, y dé una razón de este cambio. ¿Qué experimento haría usted para demostrar que la papa (*Solanum tuberosum*) es un pariente muy cercano del tomate? Es decir, porque no se justifica tener un género separado para el tomate (*Lycopersicum*) y para las diferentes especies de papa (*Solanum*). Haga un árbol taxonómico para explicarlo.

Tabla 5-1. Lista de especies y el número de genes tentativos consenso y secuencias únicas (*singletons*). Datos sacados el 14 de enero 2009 de la página http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/tc_count.pl?category=plant

Species	TCs	Singletons	TC + Singletons
<i>Malus x domestica</i>	28,148	26,317	
<i>Aquilegia</i>	13,556	7,389	
<i>Arabidopsis thaliana</i>	34,155	47,671	
<i>Hordeum vulgare</i>	41,206	39,517	
<i>Beta vulgaris</i>	4,784	12,367	
<i>Brassica napus</i>	47,634	42,693	
<i>Chlamydomonas reinhardtii</i>	11,139	20,469	
<i>Citrus clementina</i>	10,269	29,614	
<i>Theobroma cacao</i>	843	1,972	
<i>Coffea canephora</i>	7,420	10,212	
<i>Phaseolus vulgaris</i>	5,351	8,738	
<i>Gossypium</i>	49,233	65,161	
<i>Gossypium raimondii</i>	9,508	15,383	
<i>Vitis vinifera</i>	33,638	45,338	
<i>Mesembryanthemum crystallinum</i>	3,627	6,772	
<i>Euphorbia esula</i>	10,727	15,769	
<i>Lactuca sativa</i>	12,505	17,380	
<i>Lotus japonicus</i>	15,472	18,793	
<i>Zea mays</i>	85,624	151,387	
<i>Medicago truncatula</i>	29,273	38,190	
<i>Ipomoea nil</i>	11,754	9,760	

<i>Physcomitrella patens</i>	28,254	35,527	
<i>Nicotiana benthamiana</i>	5,861	10,266	
<i>Allium cepa</i>	4,063	8,182	
<i>Citrus sinensis</i>	26,081	72,817	
<i>Prunus persica</i>	8,439	15,690	
<i>Capsicum annuum</i>	4,661	9,588	
<i>Petunia hybrida</i>	2,230	6,499	
<i>Pinus</i>	34,181	27,683	
<i>Populus</i>	49,638	50,013	
<i>Solanum tuberosum</i>	31,567	29,805	
<i>Lactuca serriola</i>	8,047	13,958	
<i>Oryza sativa</i>	77,158	104,638	
<i>Secale cereale</i>	1,471	4,116	
<i>Sorghum bicolor</i>	23,442	22,583	
<i>Glycine max</i>	42,647	40,856	
<i>Picea</i>	42,051	38,443	
<i>Saccharum officinarum</i>	40,016	76,572	
<i>Helianthus annuus</i>	12,347	24,616	
<i>Festuca arundinacea</i>	6,686	13,251	
<i>Nicotiana tabacum</i>	26,151	56,932	
<i>Solanum lycopersicum</i>	25,764	21,085	
<i>Triphysaria</i>	17,442	17,043	
<i>Triphysaria versicolor</i>	7,165	5,647	
<i>Triticum aestivum</i>	91,464	124,988	

5.3.2. Genomas animales

5.3.2.1. Genoma Humano

El proyecto HUGO inició en 1990 con el objetivo de obtener el mapa genético completo del genoma humano, que no se consiguió hasta el 1996 con 5264 marcadores y 1,6 cM de distancia. En 1998 se consiguió el mapa físico con 30.075 STS localizados en contig maps. En el 2000 se anunció el primer borrador de la secuencia completa del genoma humano y en 2001 se publicó este borrador con el 90% de la secuencia. En 2003 se dio por acabada la secuencia del genoma humano. Paralelamente a los esfuerzos públicos de HUGO, una compañía privada (Celera Genomics de Craig

Venter) secuenció el genoma de varios humanos (mujeres y hombres de diferentes razas) en un tiempo récord. Esto se debió a que el grupo de Venter escogió una estrategia de secuenciación masiva a partir de fragmentos de DNA aleatorios.

La **secuenciación tipo shot gun** no necesita generar mapas genéticos previos. Sin embargo si necesitaba un software muy poderoso de ensamblado. Hoy en día se han desarrollado programas muy sofisticados de ensamblado que pueden usar fragmentos tan cortos de 100 bases para generar *contigs* de millones de bases.

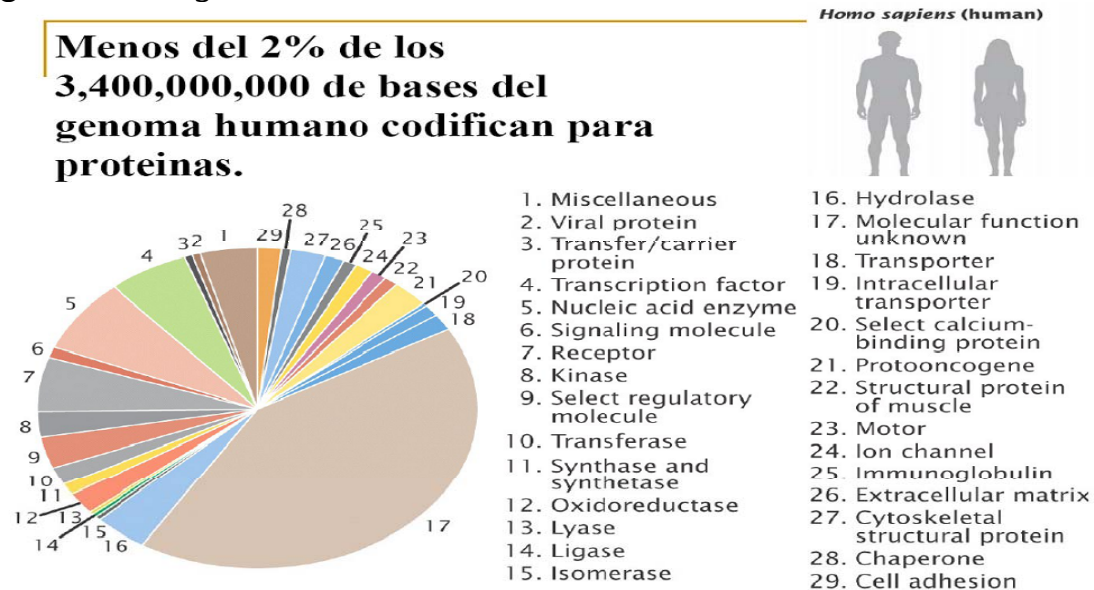


Figura 5-5. Menos del 2% de los 3400 millones de bases del genoma humano codifican para proteínas.

Tabla 5-2:

Tabla de estadísticas sobre el Genoma Humano

Topic	Statistic
Total size of the genome:	approximately 3,200,000,000 bp
Percentage of adenine (A) in the genome:	54%
Percentage of cytosine (C) in the genome:	38%
Percentage of bases not yet determined:	9%
Highest gene-dense chromosome:	chromosome 19 with 23 genes per 1,000,000 bp
Least gene-dense chromosomes:	chromosome 13 and Y with 5 genes per 1,000,000 bp
Percentage of DNA spanned by genes:	between 25% and 38%
Percentage of exons:	1.1 to 1.4%
Percentage of introns:	24% to 37%
Percentage of intergenic DNA:	74% to 64%
The average size of a gene:	27,000 bp
The longest gene:	dystrophin (a muscle protein) with 2,400,000 bp
Average length of an intron:	3,300 bp
Most common length of an intron:	87 bp

Occurrence rate of SNPs:	roughly 1 per 1,500 bp
Occurrence rate of genes:	about 12 per 1,000,000 bp

Tarea 5-9

Usando los datos del genoma humano mostrados en la tabla anterior, conteste las siguientes preguntas:

¿Cuál es el porcentaje de bases de guaninas (G) en el genoma humano? Dedúzcalo.

La longitud promedio de los intrones es de 3300 bp pero la longitud más común es de tan sólo 87 pb. Explique ¿por qué? Para ello dibuje una grafica tipo histograma con la distribución de la longitud de los intrones.

Tarea 5-10

Observe el video de las siguientes páginas web para que comprenda mejor los datos anteriores.

<http://www.elmundo.es/especiales/2001/02/ciencia/genoma/grafico>

<http://www.virtual.unal.edu.co/cursos/ciencias/2000088/lecciones/Foros/Genoma/GenomaHuman>

http://es.wikipedia.org/wiki/Proyecto_Genoma_Humano

Tarea 5-11

¿Cuántos cromosomas hay en una célula humana diploide? ¿Hay diferencias entre el sexo masculino y femenino? Investigue en otros seres vivos (mínimo 10), haga un cuadro comparativo y explíquelo.

Tarea 5-12

¿Cómo se explica que en la mitocondria tengamos material genético? ¿Postule una hipótesis de porque se ha podido acumular DNA que no tiene función aparente?

El genoma haploide humano tiene 23 cromosomas con 3200 millones de pares de bases. La cantidad de información en el núcleo es muy elevada: si puntuéramos cada nucleótido de una célula abarcaríamos una longitud de Norteamérica a Londres. Dentro de la mitocondria tenemos un genoma muy pequeño (16 kb en mamíferos) que sólo codifica para el rRNA y tRNA propios de la mitocondria y para unos pocos genes más. Se podría dividir en sus componentes: Genes y secuencias relacionadas y DNA intergénico (64%). Los genes representan el 1.4% del genoma (regiones codificadas), y el resto son intrones, transposones, etc. El DNA intergénico engloba elementos repetidos, copias dispersas, etc. El genoma es muy grande, pero las regiones codificantes son muy pequeñas. Al parecer hemos acumulando DNA que no tiene una función aparente.

Por eso se ha querido caracterizar una parte muy pequeña del genoma: los genes. Dentro del conjunto de los genes, a los mejoradores les interesan

principalmente aquellos con una repercusión en el fenotipo. Hace falta un mapa para no perderse, así nació el concepto de mapa genético: es una descripción de la posición y la distancia relativa de los genes y otros marcadores en los cromosomas de una especie. También se introdujo el concepto de Haplotipo {Aldridge, 2009 #176; Gallais, 2009 #175}.

5.3.2.2. Genoma de animales domésticos

Algunas especies domésticas se han secuenciado gracias al proyecto *Tree of Life* (<http://www.tigr.org/tol/>). Este proyecto pretende secuenciar los genomas de diferentes especies que pertenezcan a diferentes ramas del árbol evolutivo de la vida. Cuando es muy costoso secuenciar el genoma completo se plantea obtener secuencias con una cobertura de tan solo 2X. Esto es suficiente para hacer un análisis comparativo del genoma.

Ejemplo: El perro fue el primer genoma de especie doméstica en estar secuenciado. En el 2003 la empresa *Celera Genomics* secuenció el genoma del perro y lo publicó como un borrador, pero con una redundancia 1.5X (78% del genoma). Pese a esto fue útil para comparar tamaño, genes ortólogos y ver que el 5% del genoma está conservado entre el perro y el ratón. Se vio pues que ~3% eran genes codificantes, y el resto de las secuencias no lo eran, pero estaban igualmente conservados.

Desde que el perro, un descendiente del lobo, se convirtió (hace ~15.000 años) en un animal doméstico, los criadores han desarrollado varios cientos de razas. Hoy existen en el mundo cerca de 400 millones de perros de compañía. Hay una gran variabilidad en las razas de perros. Aproximadamente hay 300 razas con características morfológicas y de comportamiento diferentes. Muchas de estas razas derivan de unos pocos fundadores provocando aumento de la consanguinidad. Las razas son muy similares entre sí, pero muchas diferencias con las otras razas. En perros se han descrito 360 enfermedades hereditarias muy similares a las humanas.

El que fuera cazador del ser humano se convirtió cada vez más en un perro guardián, ayudante o faldero. Los simpáticos y ladrones mejores amigos del hombre, presentan variaciones de tamaños de 50 a 1 o incluso más. Era necesario averiguar por qué. El que investigó la causa de este extraordinario fenómeno fue Nate Sutter, un genetista del Instituto Nacional de Investigación del Genoma en Bethesda, Maryland.

Comenzó tomando radiografías de los huesos de 500 perros de aguas portuguesas y haciendo prolijas mediciones de 91 partes críticas de sus esqueletos. Hecho esto, caracterizó a cada ejemplar como pequeño o grande con respecto al promedio de su propia raza. Por último, quedaba solamente averiguar en qué difería el ADN de un perro de aguas grande del de uno pequeño. Este trabajo era fácil: el mismo Sutter es uno de los responsables del mapeo completo del genoma de Tasha, así que no tuvo que esforzarse

mucho. La diferencia entre un perro pequeño y uno grande consistía en... ¡un solo gen! El gen en cuestión, previsiblemente, ocupaba una posición en las zonas del genoma que presentan muchas variaciones entre ejemplares de la misma especie (Figura 5-6).

Cuando se descubrió que el hígado era un órgano productor de hormonas, se identificaron tres sustancias secretadas por él: una de ellas era el Factor de Crecimiento Similar a la Insulina (IGF-1, según su sigla más usual). El descubrimiento del IGF-1 demostró que la hormona de crecimiento o somatotrofina, secretada por la hipófisis, no activaba el crecimiento corporal por sí misma. En realidad, la somatotrofina disparaba la secreción de IGF-1, responsable, en última instancia, del crecimiento y regeneración de los tejidos. El nombre de la hormona proviene de la enorme similitud estructural de su molécula con la de la hormona insulina, secretada por el páncreas.

Figura 5-6. Evolución genética del perro a partir del Lobo
www.clarin.com/diario/

5.3.2.3. Genoma de la gallina

La revista "Nature" presentó en 2004 la secuencia del genoma de la gallina, el primero decodificado de un ave y de un animal de granja. Este logro de científicos de una docena de países. El DNA analizado fue el de un ejemplar "red jungle fowl" (*Gallus gallus*), especie salvaje asiática de la que derivan todas las variantes de gallinas domésticas. El animal fue criado en la Universidad de Michigan y se trata de una hembra. Estas tienen los dos tipos de cromosomas sexuales, mientras que los machos sólo tienen uno. "Esta investigación ha revelado que gallinas y humanos comparten más de la mitad de sus genes. La secuenciación de este genoma nos ayudará a descubrir los genes que aumentan la resistencia a enfermedades en las aves", asegura Jerry Dodgson, genetista que ha trabajado en el proyecto durante los últimos 17 años.

"Somos más parecidos a los pájaros de lo que creíamos. Alrededor del 60% de los genes que codifican proteínas en la gallina tienen su equivalente en el hombre", afirma Peer Bork, del Laboratorio Europeo de Biología Molecular, una de las 49 instituciones que han participado en el proyecto. La gallina es un importante modelo de investigación biomédica porque es fácil de mantener, se reproduce rápidamente y es fácil determinar sus linajes por las características físicas.

Es un animal ideal para investigar el desarrollo embrionario. Mucho de lo que se sabe del desarrollo de nuestras extremidades se debe al estudio de gallinas dentro del huevo. Se han creado estirpes que sufren patologías

genéticas propias de humanos, como la distrofia muscular. Además, como gallinas y mamíferos tienen respuestas inmunitarias muy parecidas, el genoma de esta ave ayudará a luchar contra enfermedades que saltan entre especies, como la gripe aviar.

Los investigadores dicen que la secuenciación permitirá a los productores saber por qué algunas gallinas ponen más huevos y la carne de ciertos pollos asados contiene menos grasa. "Si sabemos los genes que influyen en esas características, podremos seleccionar las gallinas que mejor cubran las demandas de los consumidores y que lo hagan de una manera más sana", dice Dodgson (Figura 5-7).

El último ancestro común del hombre y la gallina vivió hace unos 310 millones de años. Desde ese momento, los antepasados de ambas especies emprendieron caminos separados en la evolución. "El genoma de la gallina llena un hueco crucial del conocimiento científico. Localizada entre los mamíferos y los peces en el árbol de la vida, la gallina está bien colocada para proporcionarnos nuevas perspectivas de la evolución del genoma y la biología humana", señala Francis Collins, del Instituto Nacional para la Investigación del Genoma Humano.

Figura 5-7. El genoma del pollo aportará datos de enorme valor para mejorar la productividad, aunque no de inmediato.

El genoma de la gallina está compuesto por 1,000 millones de pares de bases -un tercio de los del hombre- que se reparten en 39 pares de cromosomas, uno de los cuales determina el sexo: los machos tienen dos cromosomas Z, mientras que en las hembras ese par está formado por uno Z y uno W. Esta ave de corral tiene entre 20,000 y 23,000 genes, cuyos primeros estudios ya han dado algunas sorpresas. Así, se ha descubierto que la gallina presenta un importante número de receptores olfativos en su ADN, cuando hasta ahora se creía que su sentido del olfato era muy pobre. En cambio, apenas tiene sentido del gusto.

El genoma de la gallina se obtuvo con una redundancia de 6,63X. En China hubo un proyecto con gallina que permitió determinar 2,000,000 de SNP's. Una conclusión preliminar del genoma de la gallina es que aunque es tres veces más pequeño que el humano (tiene unas 1000 Mb) no es menos complejo: es más compacto porque hay menos secuencias repetitivas, menos intrones.

5.3.3. DNA extranuclear

5.3.3.1. Genoma Mitocondrial

El DNA mitocondrial está presente como estructura circular de 17,000 pb. Se dice que es DNA citoplasmático o extracromosómico, es decir, no está

en los cromosomas del núcleo. La comparación de secuencias y detección de cambios en determinadas posiciones del genoma mitocondrial es útil para estudios evolutivos sobre DNA antiguo debido a su tamaño, estabilidad y grado de conservación. Además, el DNA mitocondrial es abundante ya que hay alrededor de 1,000 mitocondrias por célula. Una característica adicional es que se transmite por vía materna dando información sobre la dispersión de **linajes maternos** antiguos.

- **DNA pequeño y circular:** mayor estabilidad frente degradación.
- **Regiones muy conservadas:** Ejemplo de 16s de RNA ya que los cambios no son muy tolerantes evolutivamente aunque suelen producirse ciertos cambios útiles para comparar poblaciones.
- **Origen materno.** Permite establecer el linaje materno. Todos los humanos compartimos un solo ancestro femenino. La Eva mitocondrial. (Contraparte masculina: El Adán del cromosoma Y)
- **No recombina.** El DNA de una mitocondria paterna no se mezcla con el de las mitocondrias maternas.
- **La tasa mutagénica es diferente.** El mecanismo de reparación del DNA es diferente a los nucleares. Con menos mutaciones se dan menos polimorfismos.
- **Ediciones especiales.** En la mitocondria en algunos genes se dan procesos frecuentes de edición de RNA antes de ser traducidos a una proteína, y además, en algunos casos también se presentan variantes al código genético universal.
- **Regiones hipervariables:** Su replicación es en forma de “D-Loop” (replica una cadena antes que la otra), la región cercana a este inicio de replicación (D-Loop) acostumbra a ser muy variable (muy útil).
- **Poco DNA:** Es el único marcador que se puede utilizar con poquísimos DNA (cadáveres antiguos...)

Aplicaciones del DNA mitocondrial:

Inferir relaciones entre poblaciones y especies:

Para el conocimiento del **origen de razas animales domésticas**. Se usan regiones del Citocromo B del DNA mitocondrial para estudiar por ejemplo el origen de razas animales:

Ejemplo - Distinción de razas de cerdos:

Queremos saber si tres especies de cerdo; Ibérico, canario y mallorquín están relacionados. Para ello, obtenemos muestras biológicas de un grupo de individuos, extremos DNA, amplificamos con PCR y lo secuenciamos. Al comparar los polimorfismos nos damos cuenta que el cerdo ibérico y el mallorquín sólo tiene haplotipos europeos, mientras que la raza canario

posee haplotipos tanto europeos como asiáticos. De esta forma podemos inferir de dónde vienen estas poblaciones porcinas.

Ejemplo – Lugar de domesticación del cerdo:

La domesticación del cerdo se pudo dar independientemente en Europa, Asia y otros lugares al mismo tiempo. El estudio del DNA mitocondrial permitió confirmar la existencia de múltiples centros de domesticación y que ha habido poca migración de estos cerdos domesticados hacia otros lugares. Esto se comprobó gracias a que el DNA mitocondrial del cerdo asiático domesticado se parecía más al del cerdo asiático salvaje, mientras que el cerdo europeo se parecía mas al cerdo europeo salvaje.

Tarea 5-13

Busca otros ejemplos de especies animales como de plantas en donde se haya usado el DNA mitocondrial para conocer sus ascendencia. ¿Cómo se explica la evolución del Hombre?, ¿Dónde se supone que se origino y por qué?

Comparación de especies y subespecies:

También permite realizar estudios de **filogenética** y **taxonomía**. Se estudia una región no codificante en la que las mutaciones se pueden considerar neutras (no intervienen en la selección). Esto permite calcular las tasas de divergencia entre especies. Para cada gen se estima una tasa de mutación por unidad de tiempo. Este análisis también puede ser respaldado por **fósiles**, para realizar una calibración molecular de la tasa evolutiva. Las evidencias de un grupo de genes se pueden contrastar comparándolas con otras secuencias independientes.

Modelo Árbol:

Relación entre diferentes haplotipos. Cuanta más diferencias polimórficas mayor será la distancia evolutiva, y esto se representa con una “rama” más larga. Son árboles de parsimonia que intentan agrupar las muestras de par en par de la mejor manera.

Tanto un árbol de parsimonia y uno de distancias (ME) para el mismo set de datos produce

Topologías y longitudes de ramas idénticas. La diferencia radica en que el árbol de parsimonia identifica qué sitio del alineamiento contribuye cada paso mutacional en la longitud de cada rama.

Como criterio de optimización de se usa la (máxima) parsimonia. Esto involucra la identificación de la topología con la menor longitud total del árbol, es decir, que requiere el menor número de cambios evolutivos (transformaciones) para explicar las diferencias observadas entre OTUs. El modelo de máxima parsimonia se justifica en filogenética dado que:
1) se asume que los cambios de estado de base (sustituciones) son poco

frecuentes y 2) no se puede conocer con exactitud el camino evolutivo de dichos cambios, por lo que se busca maximizar la similitud evolutiva que se puede explicar como homóloga (por ancestría compartida). De esta manera se busca de minimizar la homoplasia (similitud no heredada directamente del ancestro), ya que las hipótesis de homoplasia (convergencia, evolución paralela...) pueden ser juzgadas como intentos *ad hoc* de explicar por qué determinados datos no encajan en una hipótesis evolutiva (árbol filogenético) particular.

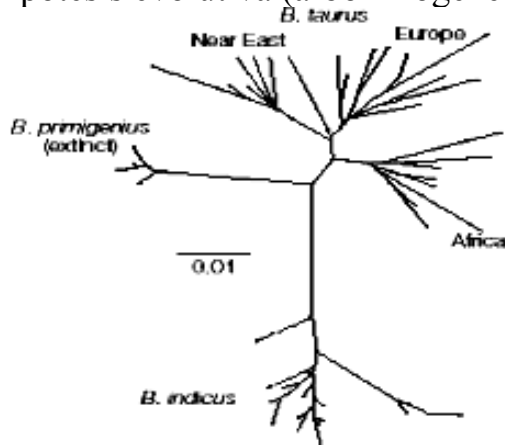


Figura 5-8.

Tarea 5-14.

En la página de Internet www2.uah.es/biologia_animal/EFyC2008/Practica%205.pdf realiza la práctica para la construcción de un árbol de parsimonia.

Mapas de migración:

Estudio del origen de poblaciones (por ej. migraciones bovinas). Se intentan encontrar caminos lógicos para explicar cómo han pasado de un lugar a otro. Las alturas de las barras nos indican la diversidad nucleotídica. En los lugares donde hay mayor diversidad se cree que son los centros de origen, y de ahí es dónde provienen las migraciones. Por ejemplo, los bovinos europeos entraron desde Turquía y no África, que además coincide con las migraciones humanas. En el mapa de rutas cada color indica una especie diferente. Vemos un foco amarillo importante con ciertas rutas de migración bovina. El rojo y negro representan orígenes de domesticación independientes.



Figura 5-9. Mapas de migración bovina.

5.3.3.2. Genoma plastídico

Los cloroplastos son un tipo de plastos, orgánulos exclusivos de las células vegetales. Tienen una forma alargada y presentan una organización similar a la de las mitocondrias. Su importancia se debe a que son los orgánulos donde se realiza la fotosíntesis.

Estructura

Presentan tres sistemas de membranas, dos de los cuales forman la envoltura externa. Estas membranas son:

- La membrana externa, que contiene porinas que le proporcionan una gran permeabilidad.
- La membrana interna, menos permeable. Junto con la membrana anterior forma la envoltura externa.
- La membrana tilacoidal, se localiza en el interior del cloroplasto y está altamente plegada para formar los tilacoides. Estos se apilan unos sobre otros para formar grana. En esta membrana se localizan los fotosistemas, responsables de la captación de la energía solar, los componentes de una cadena de transporte electrónico y una ATP sintasa. Esta membrana es muy fluida debido a que contiene una elevada proporción de ácidos grasos.

Figura 5-10. A) El cloroplasto presenta tres compartimentos diferentes: un espacio intermembranoso, un espacio interno, ocupado por el estroma o matriz del cloroplasto, y un espacio tilacoidal.

El espacio intermembranoso se localiza entre las dos membranas que constituyen la envoltura externa del cloroplasto, y posee una composición semejante a la del citosol. El estroma está situado entre la membrana interna y la membrana tilacoidal. En él se encuentran ribosomas, enzimas, varias moléculas de ADN, distintos tipos de ARN, gránulos de almidón y gotas de lípidos. B) El espacio tilacoidal corresponde al espacio interno de los tilacoides que, al estar interconectados, delimitan un único compartimento común.

Las mitocondrias y los cloroplastos son orgánulos celulares autónomos, ambos poseen ADN que es capaz de replicarse, transcribirse y traducirse independientemente del nuclear. A la vista de este hecho, podríamos pensar, si esos genes contenidos en el cromosoma de mitocondrias y cloroplastos, tuvieran o no influencia en el fenotipo de un individuo. De ser así, su herencia no sería de tipo mendeliana. Los genes nucleares segregan en meiosis, pero los genes de estos orgánulos segregaran en mitosis, cuando se produce la distribución al azar de los orgánulos celulares. En la fecundación la aportación citoplásmica del gameto masculino es muy pequeña, sí no nula, por lo tanto hemos de pensar que los individuos reciben sólo aportación de mitocondrias y cloroplastos vía materna, es decir

sólo recibe los genes extranucleares de la madre. Un último aspecto a considerar es que si el individuo empieza su desarrollo en el citoplasma del óvulo, ¿tendrán influencia las proteínas (productos génicos de genes nucleares maternos) allí presentes sobre el fenotipo del individuo?

Es un genoma circular, filogenéticamente muy cercano al genoma de las cianobacterias, como demuestra gran cantidad de estudios de evolución molecular. En plantas con flores el cloroplasto se hereda de manera casi exclusiva por vía materna, y por lo tanto, es un genoma haploide (cada organismo solo tiene una copia de este genoma). Es interesante notar que en coníferas (Gimnospermas) usualmente los cloroplastos se heredan por vía paterna así, mientras que en las angiospermas el cloroplasto describe la historia evolutiva por parte de la función femenina (patrones de dispersión de las semillas), en coníferas se relaciona con la función masculina (patrones de dispersión del polen).

Existen varias secuencias completas del genoma del cloroplasto en plantas con flores, incluyendo especies monocotiledóneas como el arroz (*O. sativa*) y el maíz (*Zea mays*), y varias dicotiledóneas, como el tabaco (*Nicotiana tabacum*), la espinaca (*Spinacia oleracea*), la especie experimental de la familia de la mostaza *Arabidopsis thaliana* y la planta parásita no fotosintética *Epifagus virginiana*.

El cloroplasto en las plantas con flores tiene una estructura muy conservada. Generalmente mide entre 120 y 217kb (miles de pares de bases de DNA), aunque en *E. virginiana* mide tan solo 70kb. Esta planta, al ser parásita, no necesita de la maquinaria fotosintética y ha perdido muchos de los genes involucrados en este mecanismo. El número de genes presente en el cloroplasto es también muy conservado, entre 110 a 113 (Tabla 5-3), aunque *E. virginiana* sólo tiene 42 genes. Así, el número de genes del cloroplasto es menos del 25% del que presentan los genomas bacterianos más pequeños, posiblemente debido a la pérdida de una elevada cantidad de los genes que tenían inicialmente las bacterias que dieron origen a los actuales cloroplastos. Por ejemplo, la bacteria con genoma "mínimo" *M. genitalium*, tiene 470 genes que codifican para proteínas.

Tabla 5-3. Comparación de genomas terminados de cloroplastos *
<http://www.scielo.org.ve/scielo.php?pid>

Los cloroplastos eucarióticos derivan evolutivamente de las cianobacterias, su DNA es una molécula circular, bicatenaria y que contiene entre 120 y 160 kbp. Este contenido en DNA es inferior al que presentan las algas verdes, por lo que se supone que a lo largo de la evolución se ha reducido el tamaño del mismo, bien por pérdida de material o bien por integración de genes dentro del genoma eucariótico. El número de moléculas de DNA

por cloroplasto es muy variable. En plantas este número oscila entre 20 y 60 moléculas por cloroplasto, mientras que en algas superiores el número es mayor. La organización es muy sencilla ya que apenas posee secuencias repetidas, aunque en la mayoría de las especies vegetales se ha demostrado la presencia de una duplicación invertida que incluye genes de tRNA. La secuenciación del DNA de los cloroplastos ha revelado que existen alrededor de 150 genes. Además de los tRNA y los rRNA hay unos 90 genes que están implicados en la fisiología de los cloroplastos, principalmente en la fijación fotosintética del carbono. Estos genes suelen agruparse en pequeños nichos cuya organización se ha mantenido a lo largo de la evolución. El orden de estos genes dentro del genoma del cloroplasto, no se ha mantenido constante en todas las especies estudiadas, lo cual indica que en el proceso evolutivo han existido variaciones cromosómicas estructurales.

El cloroplasto, es el asiento de la fotosíntesis en una célula, contiene su propio genoma conocido como plastome. El plastome incluye los genes que cifran para el RNA ribosomal, el RNA de transferencia y varias proteínas que alternadamente controlan fotosíntesis así como otros rasgos. Un desarrollo reciente en la manipulación de los genes del plástido es la introducción de genes extraños en los cloroplastos y esta técnica se conoce como ingeniería del Plastome. La transformación de plastome primero fue divulgada por Boynton y sus colegas en 1988 en *Chlamydomonas reinhardtii*. Esto fue logrado, bombardeando las células con las partículas de tungsteno cubiertas con el DNA. Ahora, las líneas transplastómicas (ésas que tienen un transgen) se han sintetizado ya en el tabaco, la soja, la colza, la espinaca, la patata dulce, etc.

El conocimiento del plastome completo puede ser útil para entender los procesos fundamentales de la biología del plástido y de la interacción del plastome con genoma nuclear y del proceso fotosintético. La diversidad genética de los cloroplastos de varias cosechas (en los niveles intervarietal, interespecíficos o intergenérico) puede ser revelada realizando el análisis de las secuencias. Los genes fotosintéticos codificados por el plastome se pueden manipular para lograr el nivel deseable de la eficacia fotosintética, así como la productividad. Puesto que los genes del plástido maternal se heredan los genes nuevamente incorporados no conseguirán transmitidos a los parientes o a las malas hierbas salvajes a través del polen. Esto es especialmente atractivo para los genes de tolerancia a herbicidas, puesto que el desarrollo de “malas hierbas estupendas” (las malas hierbas tolerantes a los herbicidas) se evita. Una vez que introduzcamos un gene nuevo en un solo plastome de una célula, consigue amplificarla varias veces durante la división y la multiplicación de plástidos.

Los genes para producir las proteínas útiles se pueden transferir en los cloroplastos para las expresiones múltiples y su producción enorme. Las líneas transplastómicas pueden ser desarrolladas para fines farmacéuticos.

5.3.3.3. DNA nuclear Vs. DNA citoplásmico

Ventajas del DNA nuclear frente al DNA mitocondrial:

El DNA nuclear por lo regular posee dos alelos; Esto permite hacer pruebas de paternidad.

La variabilidad del DNA nuclear es más alta que la del DNA mitocondrial (mitDNA) y dan mayor capacidad de discriminación: En una población muy parecida es difícil la discriminación, con suerte dos a tres haplotipos diferentes.

Con el DNA nuclear se pueden hacer dibujos más “finos” ya que si analizamos un número elevado de microsatélites tendremos una fuerza de discriminación muy alta. Esto nos sirve cuando por ejemplo queremos averiguar cuán diferentes son dos poblaciones, una salvaje y otra domesticada.

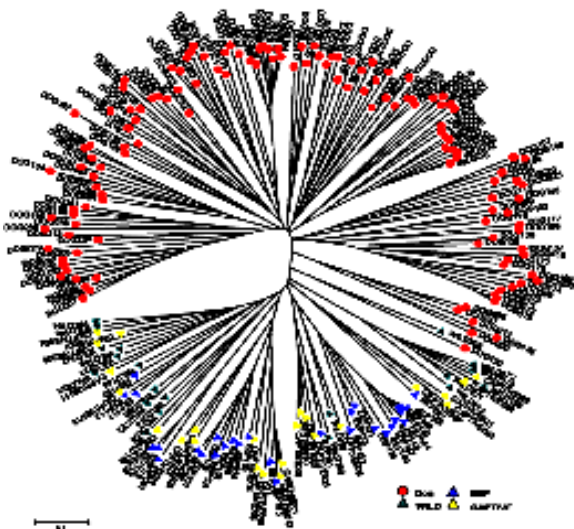


Figura 5-11. Diversificación entre el perro y lobo

Ejemplo de lobos y perros: Se quiere ver las poblaciones de perros en lugares donde también hay lobos. Se analizaron 74 muestras (lobos del zoo europeo; azul, Lobos en cautividad; amarillo, Perros; rojo). La pregunta que nos hacemos es si los lobos en cautiverio son de amplia diversificación y si sólo representan un grupo genético muy reducido. Las manchas amarillas están bastante distribuidas (muestras representativas de la diversidad genética, ya que además estas están intercaladas con lobos salvajes; verdes). En general vemos clara diferenciación genética entre perro y lobo (entre azul y rojo).

Elevado margen de confianza con pocos marcadores: Si utilizamos microsatélites ¿cuántos debemos utilizar? Queremos tener un nivel de confianza elevada de que los individuos no salgan iguales. Estamos

entonces frente a cifras de 3 a 20 marcadores. Además, tenemos que tener en cuenta el grado de consanguineidad. Para estimar divergencias genéticas entre individuos muy cercanos es difícil tener capacidad de discriminación con DNA mitocondrial. Los marcadores buenos son aquellos con suficientes número de alelos por *Locus*. Sin embargo, algunos marcadores son un poco delicados por las diferencias que pueden aparecer entre laboratorios.

5.3.4. Genomas comparativos

Cuando se habla de genómica comparativa se puede decir que se tiene un cromosoma (por ejemplo arroz) donde se ha identificado ciertos genes. Si estos genes se han encontrado también en maíz y trigo podríamos preguntarnos en qué cromosomas están repartidos esos genes. ¿Puede ser que algunos genes estén agrupados similarmente en especies cercanas? Al establecimiento de correspondencias entre genomas se le llama **análisis de sintenia**. Si el cromosoma nueve del arroz es homólogo a parte del cromosoma ocho del maíz es debido a que estas especies tuvieron un ancestro común en donde los genes estaban organizados de esa forma, pero con la evolución se han reorganizado en las diferentes especies. Podemos notar que esta reorganización es mayor en algunos géneros que en otros. La comparación entre especies de cereales puede ser buena, mientras que la comparación entre maíz y frijol puede ser no tan buena.

A los genes que supuestamente tienen la misma función en dos especies se les llama **ortólogos**. Cuando hablamos de comparar genomas queremos buscar similitudes o diferencias en el orden de los genes ortólogos. La **sintenia** es la similitud en la distribución de genes ortólogos a lo largo de cromosomas de especies diferentes.

Cuando hablamos de secuencias, es mucho más exacto hablar de similitudes que de homologías. Homología es un término que se usa en la evolución con una connotación muy específica. Las manos de los humanos y las alas de los murciélagos son órganos homólogos. Pero las alas de los pájaros y las alas de los insectos son órganos análogos y no homólogos. Los genes ortólogos, son genes que no solo son similares a nivel de secuencia, sino que además, cumplen todavía la misma función enzimática o estructural.

6. Herramientas bioinformáticas

6.1. ¿Qué es la bioinformática?

Tal vez no se pueda dar una sola definición de bioinformática, ya que la percepción de ella depende mucho de los tipos de usuarios. Por un lado están los biólogos que usan la bioinformática solo como algo accesorio, es decir, como un **conjunto de herramientas** para contestar **preguntas biológicas**. Mientras que por otro lado están los informáticos, programadores y estadistas para los cuales la bioinformática *per se*, es el tema central de su trabajo. Para algunos, la bioinformática es solo un medio, más no un fin, mientras que para otros, la bioinformática se justifica como una **disciplina científica emergente con una meta propia**.

Algunos consideran que la bioinformática utiliza la tecnología de la información para organizar, analizar y distribuir información biológica con la finalidad de responder preguntas complejas en biología. Esto incluye la colección, organización, almacenamiento y recuperación de información en bases de datos. Otros consideran que la bioinformática es un área propia de investigación multidisciplinaria, la cual puede ser ampliamente definida como la interfase entre dos ciencias: Biología y Computación. Involucra la solución de problemas complejos usando herramientas de sistemas y computación.

Según la definición del NCBI (2001); la bioinformática es un campo de la ciencia en el cual confluyen varias disciplinas tales como: biología, computación y tecnología de la información. El fin último de este campo es facilitar el descubrimiento de nuevas ideas biológicas, así como crear perspectivas globales a partir de las cuales se puedan discernir principios unificadores en biología.

Aparte de las definiciones formales de organismos o instituciones de referencia, los manuales de esta materia aportan sus propias definiciones operativas, lógicamente vinculadas en mayor o menor medida con las ya mencionadas. En otras palabras, cada quien define los términos según su propio interés o su área de trabajo. Dentro de este libro de texto, vamos a definir algunos términos de la siguiente forma:

- **Bioinformática:** Es un término difuso que enmarca de manera global todas las actividades, en donde la realización del experimento no es el elemento central, sino más bien, el análisis de datos y la interpretación de la información de origen biológico.

- Los **bioinformáticos**: Hacen investigación, desarrollo o aplicación de herramientas computacionales y aproximaciones para la expansión del uso de datos biológicos, genómicos, bioquímicos, agronómicos, médicos o de salud, incluyendo aquellas herramientas que sirvan para adquirir, almacenar, organizar, analizar o visualizar tales datos.
- **Biología computacional**: Es el desarrollo y aplicación de métodos teóricos y de análisis de datos, modelado matemático, estadística y técnicas de simulación computacional al estudio de sistemas biológicos, bioquímicos, agronómicos, conductuales y sociales.
- **Biocomputación**: Es la construcción y uso de computadores que contienen componentes biológicos o funcionan como organismos vivos. Esto incluye a las redes neuronales, los sistemas cibernéticos y la inteligencia artificial.
- La **bioestadística**: Es una serie de herramientas matemáticas que sirven para entender, interpretar y resumir mejor los datos a pesar de la grandísima influencia del azar en los sistemas biológicos. Las pruebas de bioestadística nos dan una idea cuantitativa de la probabilidad de que el azar es la causa principal de nuestros resultados. La bioestadística también tiene la finalidad de ahorrarle recursos a los experimentadores para obtener más datos confiables con menos repeticiones y experimentos.

Bajo estas definiciones, la bioinformática tiene más que ver con la **información**, mientras que la biología computacional lo haría con las **hipótesis**, y la bioestadística con el **azar**. Por otra parte, el término biocomputación suele enmarcarse en las actuales investigaciones con biocomputadores, es decir los dispositivos cibernéticos que usan principios biológicos u orgánicos para almacenar o procesar información.

Durante los inicios de la genómica el concepto de bioinformática se refería sólo a la creación y mantenimiento de bases de datos donde se almacenaba secuencias de nucleótidos y aminoácidos. El desarrollo de este tipo de base de datos no solamente significaba el diseño de las mismas, sino también el desarrollo de interfaces complejas donde los investigadores pudieran acceder los datos existentes, y suministrar o revisar datos por medio de la web.

Toda esa información debe ser combinada para formar una idea lógica y razonable, de tal manera, que los investigadores pudieran estudiar cómo estas actividades se veían alteradas en estados de una enfermedad o en dependencia de un factor ambiental. De allí viene el surgimiento de nuevos

campos de la bioinformática. El desarrollo de nuevos algoritmos (fórmulas matemáticas) y estadísticos, con los cuales se pueda relacionar partes de un conjunto enorme de datos, como por ejemplo, métodos heurísticos para localizar un gen dentro de una secuencia, predecir estructura o función de proteínas, y poder agrupar secuencias de proteínas en familias relacionadas. Por ahora el campo más popular es el análisis e interpretación de secuencias de nucleótidos y aminoácidos, dominios de proteínas y estructuras de proteínas. Los principales esfuerzos de investigación en estos campos incluyen el alineamiento de secuencias, la predicción de genes, montaje del genoma, alineamiento estructural de proteínas, predicción de estructura de proteínas, predicción de la expresión génica, interacciones proteína-proteína, y modelado de la evolución.

Los términos bioinformática, biología computacional y biocomputación son utilizados en muchas ocasiones como sinónimos. En realidad hacen referencia a campos de estudios interdisciplinarios muy vinculados, que requieren el uso o el desarrollo de diferentes técnicas que incluyen informática, matemática aplicada, estadística, ciencias de la computación, inteligencia artificial, química y bioquímica, para solucionar problemas, analizar datos, o simular sistemas o mecanismos, todos ellos de índole biológica, y usualmente (pero no de forma exclusiva) en el nivel molecular. El meollo principal de estas técnicas se encuentra en la utilización de recursos computacionales para solucionar o investigar problemas sobre escalas de tal magnitud que **sobrepasan el discernimiento humano**. La investigación en biología computacional se solapa a menudo con la biología de sistemas.

Una constante en proyectos de bioinformática y biología computacional es el uso de herramientas matemáticas para extraer información útil de datos producidos por técnicas biológicas de alta productividad, como la secuenciación del genoma. Un problema común en bioinformática es el montaje o ensamblado de secuencias genómicas de alta calidad desde fragmentos mas o menos cortos obtenidos tras la secuenciación del DNA a gran escala. Otros problemas comunes incluyen el estudio de la regulación génica para interpretar patrones de expresión génica utilizando datos de microarreglos de cDNA o de oligonucleótidos.

Principales áreas de investigación:

- Análisis de secuencias
- Anotación automática de genomas
- Genómica comparativa y ortología
- Biología evolutiva computacional
- Medición de la biodiversidad
- Análisis de la expresión génica

- Análisis de la regulación
- Análisis de la expresión de proteínas
- Análisis de la interacción entre proteínas
- Determinación de la estructura de las proteínas
- Predicción del plegamiento de proteínas
- Acoplamiento proteína-proteína
- Análisis de mutaciones en el cáncer
- Modelado de sistemas biológicos
- Simulación de rutas y redes metabólicas
- Análisis de imagen de alto rendimiento
- Herramientas de *software*
- Análisis estadístico de datos

6.2. Análisis de secuencias

6.2.1. BLAST

El *Basic Local Alignment Search Tool* (BLAST) es un programa informático de alineamiento basado en similitudes de tipo local. Las secuencias alineadas pueden ser de DNA o de proteínas. El programa es capaz de comparar una secuencia problema (comúnmente llamada *query*) contra una gran cantidad de secuencias que se encuentren en diversas bases de datos. El algoritmo encuentra las secuencias que tienen mayor parecido a la secuencia *query* y las alinea. Es importante mencionar que BLAST usa un algoritmo heurístico, por lo que no nos puede garantizar que ha encontrado la solución correcta. Sin embargo, BLAST es capaz de calcular una probabilidad de error de que el alineamiento se deba meramente al azar. Por lo que el *E value* (valor E), es un parámetro estadístico que nos permite juzgar el grado de significancia de los resultados que se obtienen. El *E value* depende del número de identidades de nuestro *query* (hits perfectos) pero también del número de secuencias en las bases de datos analizadas.

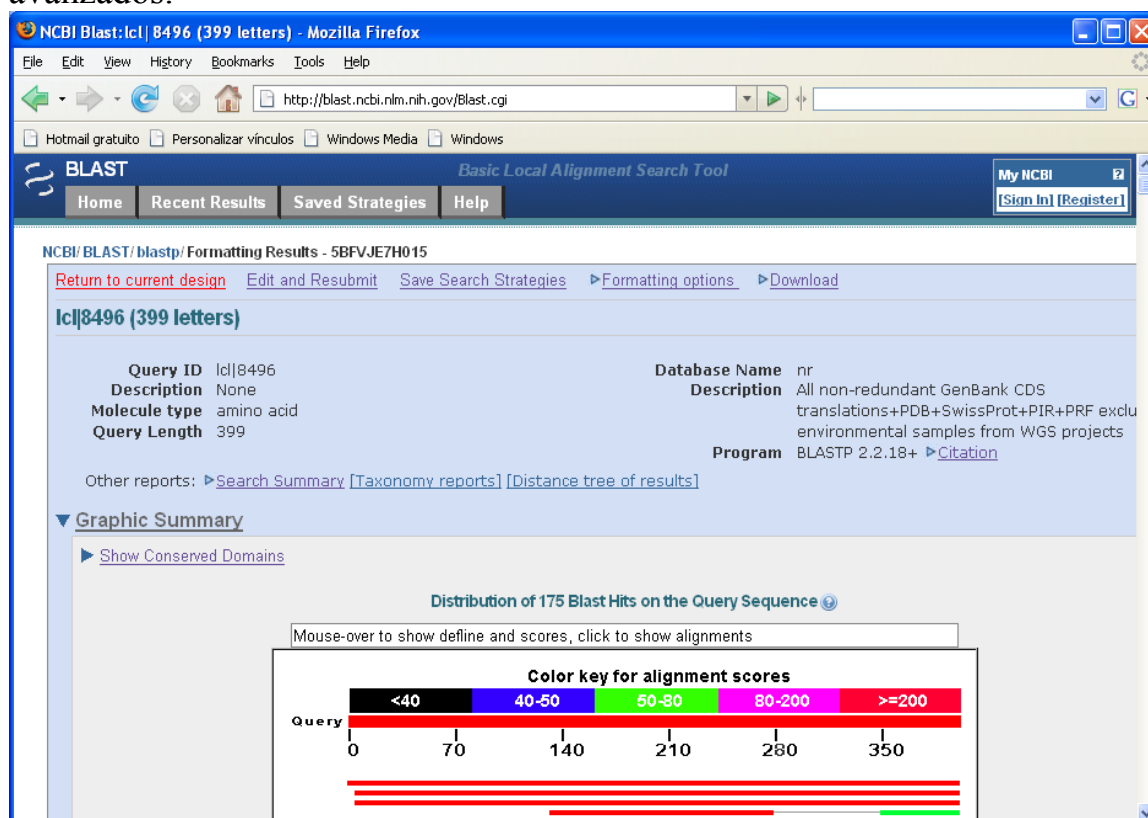
Normalmente el BLAST es usado para encontrar genes que tengan similitudes con nuestra secuencia en cuestión. Por lo general, cuando una nueva secuencia es obtenida, se usa el BLAST para compararla con otras secuencias que han sido previamente caracterizadas, para así poder inferir su función. La **predicción de función en base a similitud** es el paradigma central de la bioinformática, y por consiguiente, el BLAST es uno de los pilares de la genómica. El BLAST es la herramienta más usada para la anotación y predicción funcional de genes o secuencias proteicas.

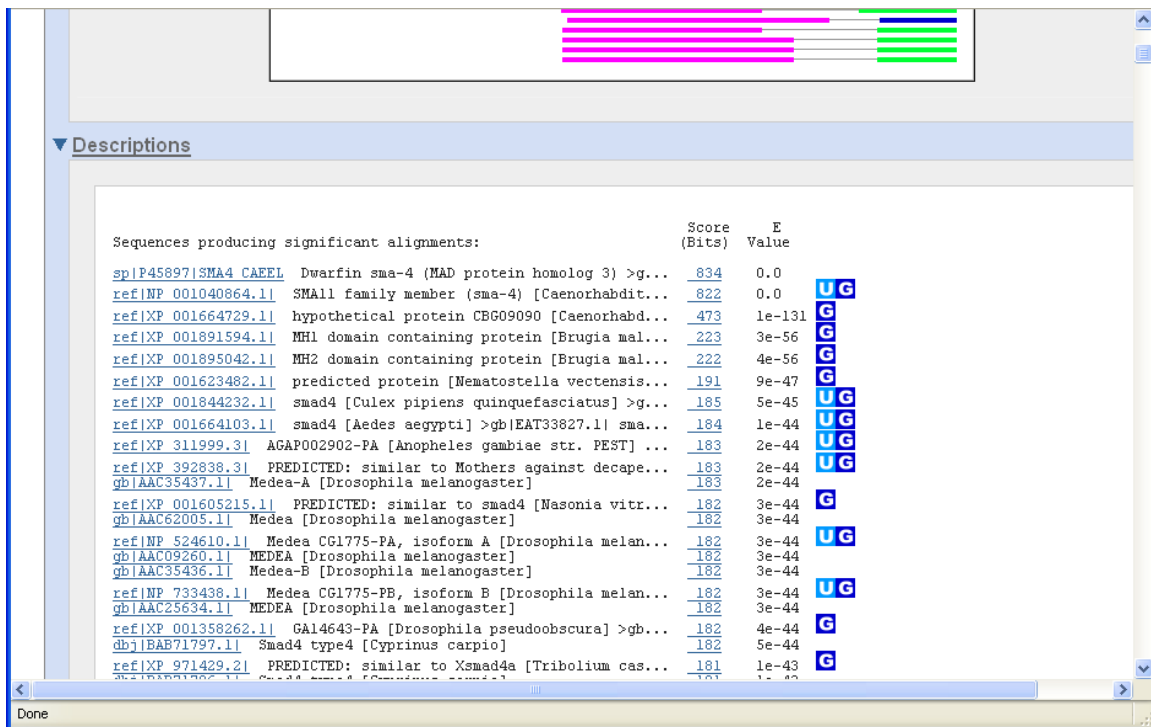
BLAST fue desarrollado por los Institutos Nacionales de Salud del gobierno de EE. UU., por lo que es de dominio público y puede usarse

gratuitamente a través de un servidor del Centro Nacional para la Información Biotecnológica (NCBI).

<http://blast.ncbi.nlm.nih.gov/Blast.cgi>

El programa BLAST también puede ser instalado en una computadora personal para buscar secuencias en bases de datos de manera local. Algunas ventajas de usar el servidor del NCBI son que el usuario no tiene que mantener ni actualizar las bases de datos y que la búsqueda se hace en un cluster de computadoras, lo que otorga rapidez. Las desventajas son: no se permite hacer búsquedas masivas dado que es un recurso compartido, no se puede personalizar las bases de datos contra la que busca el programa y las secuencias son enviadas al servidor del NCBI sin ningún tipo de cifrado, lo que puede ser un problema para quienes quieran mantener sus secuencias privadas. La aplicación local de BLAST tiene la ventaja de que permite manejar varios parámetros que en las búsquedas de NCBI están estandarizados, por lo que provee una mayor flexibilidad para los usuarios avanzados.





Sequences producing significant alignments:	Score (Bits)	E Value
sp P45897 SMA4 CAEEL Dwarf fin sma-4 (MAD protein homolog 3) >g...	834	0.0
ref NP_001040864.1 Small family member (sma-4) [Caenorhabdit...	822	0.0
ref XP_001664729.1 hypothetical protein CBG09090 [Caenorhabd...	473	1e-131
ref XP_001891594.1 MH1 domain containing protein [Brugia mal...	223	3e-56
ref XP_001895042.1 MH2 domain containing protein [Brugia mal...	222	4e-56
ref XP_001623482.1 predicted protein [Nematostella vectensis...	191	9e-47
ref XP_001844232.1 smad4 [Culex pipiens quinquefasciatus] >g...	185	5e-45
ref XP_001664103.1 smad4 [Aedes aegypti] >gb EAT33827.1 sma...	184	1e-44
ref XP_311999.3 AGAP002902-PA [Anopheles gambiae str. PEST] ...	183	2e-44
ref XP_392838.3 PREDICTED: similar to Mothers against decap...	183	2e-44
gb AAC35437.1 Medea-A [Drosophila melanogaster]	183	2e-44
ref XP_001605215.1 PREDICTED: similar to smad4 [Nasonia vitr...	182	3e-44
gb AAC62005.1 Medea [Drosophila melanogaster]	182	3e-44
ref NP_524610.1 Medea CG1775-PA, isoform A [Drosophila melan...	182	3e-44
gb AAC09260.1 MEDEA [Drosophila melanogaster]	182	3e-44
gb AAC35436.1 Medea-B [Drosophila melanogaster]	182	3e-44
ref NP_733438.1 Medea CG1775-PB, isoform B [Drosophila melan...	182	3e-44
gb AAC25634.1 MEDEA [Drosophila melanogaster]	182	3e-44
ref XP_001358262.1 GAL4643-PA [Drosophila pseudoobscura] >gb...	182	4e-44
gb BAB71797.1 Smad4 type4 [Cyprinus carpio]	182	5e-44
ref XP_971429.2 PREDICTED: similar to Xsmad4a [Tribolium cas...	181	1e-43

Consideraciones al usar BLAST:

- A pesar de que BLAST es un programa muy poderoso y casi siempre podemos confiar en sus resultados, se debe recordar que el programa es heurístico, y por lo tanto puede que no encuentre la solución óptima. En la actualidad, el abuso y la pobre interpretación de los resultados de BLAST ha llevado a múltiples errores de anotación. Una cosa a tener en cuenta al usar BLAST es que, cuanto más evidencia externa se pueda obtener para corroborar un alineamiento (fisiológica, filogenética, genética, etc.) es mejor.
- El programa de BLAST **NO** garantiza que las secuencias que alinea son homólogas y mucho menos que tengan la misma función (en el caso que sean proteínas), simplemente provee posibles candidatos. Se necesitan más análisis para anotar correctamente una secuencia.
- La puntuación BLAST depende del largo de la secuencia, una secuencia muy corta tendrá una puntuación menor que una grande, simplemente por la cantidad de bases que tiene. Así que siempre se debe interpretar la puntuación con respecto al largo de la secuencia.
- El *e-value* depende del tamaño de la base de datos. Para bases de datos muy pequeñas, *e-values* altos son menos significativos que para bases de datos muy grandes. Para la base de datos NCBI por lo general *e-values* de 0.01 o menos son considerados como

significativos, pero esto puede depender de la secuencia que se esté analizando.

- Se debe tener cuidado con los errores de anotación, es común que alguna secuencia que se anotó mal (ya sea porque se hizo automática o por error humano) sea utilizada como referencia para anotar otras secuencias similares, por lo que los errores de anotación se pueden propagar rápidamente. Siempre debemos especificar que la función de nuestra secuencia es posible o probable si fue asignada usando identidad con otras secuencias. Asimismo debemos tener en cuenta que la gran mayoría de las funciones asignadas en la actualidad son putativas y que pueden no ser una buena referencia para una asignación funcional.
- A pesar de lo que comúnmente se piensa, las secuencias con la mejor puntuación o el mejor e-value **NO** necesariamente es la mejor opción para la anotación funcional de un gen. Es importante analizar todos los alineamientos que encuentra el programa y sacar conclusiones en base al resultado global.
- BLAST tiene varios parámetros por defecto que en general funcionan bien para la mayoría de los casos, pero habrá situaciones en las que es necesario cambiarlos para obtener mejores resultados. No hay forma de saber exactamente qué parámetro es el óptimo, se tienen que realizar múltiples pruebas hasta encontrar las mejores condiciones.

Tarea 7-16-1

Haga una búsqueda en la base de datos del NCBI y averigüe mediante la herramienta BLAST, a qué proteína corresponde la siguiente secuencia: GRDKVPSSGSKMPRTVIALDGGGLYEHYKKFSSCVEATLT

Obtenga la secuencia completa de esa proteína. Tradúzcala a su secuencia nucleotídica.

Respuesta:

Para responder primero debes saber a qué base de datos se refiere. La podrás encontrar en la página: <http://www.ncbi.nlm.nih.gov/sites/entrez>. Una vez ahí busca en la barra de la izquierda la herramienta **BLAST** y ábrela. Como la secuencia que tienes es de aminoácidos usa la herramienta *protein blast*. Cuando la abras te pedirá que introduzcas la secuencia que quieres buscar y te la pide en formato FASTA. Este formato es un formato de fichero informático basado en texto, utilizado para representar secuencias bien de ácidos nucleicos, bien de péptido, y en el que los pares de base o los aminoácidos se

representan usando códigos de una única letra. El formato también permite incluir nombres de secuencias y comentarios que preceden a las secuencias en sí. Esto significa que metas la secuencia de aminoácidos sin espacios y sin puntos y comas, solo letras. Una manera es copiar del problema la secuencia y pegarla en la ventana. Una vez que hayas hecho esto oprime la instrucción BLAST, casi al final de esta ventana. Los resultados aparecerán en unos segundos, dependiendo de la red. Obtendrás una tabla con las secuencias que tienen un alineamiento significativo, es decir secuencias que se parecen a la que acabas de meter, y están ordenadas de las más parecidas a las que menos se parecen. El primer resultado es la hexocinasa de maíz

Hexocinasa de maíz (zea mays)

Secuencia completa:

```

1      mvkavvvgaa  vvacaavgva  avlvhrrrrr  daallgsaea
errrraaavi eeverslatp
61     tallrgiada  mvaemerglr  gdihaqlkml  isyv dnlptg
dehglfyald lggt nfrvlr
121    vqlggrekrv  vkqqyeevsi  pphlmvgtsm  elfdfiaaal
akfvgteged fqlpegrqre
181    lgftfsfpvn  qtsissgtli  kwtkgfsing  tvgedvysel
sramerqgld mkatalvndt
241    vgtlaggrym  dtdvvaavil  gtgt naayve  hanaipkwtg
llpksgkmvi ntewgsfksn
301    klplseydka  mdfeslnpge  qiyekmisgm  ylgeivrril
klahdaslf gdvvptkleq
361    pfilrtpdms  amhhdsshdl  ktlgsklkdi  vgvadtslev
ryitrhicdl vaergar laa
421    agiysilkki  grdkvpssgs  kmprt viald  gglyehykkf
sscveatltd llgeeasssv
481    vaklandgsg  igaallaash  sqygasd

```

6.3. Bases de datos

6.3.1. Genbank y el NCBI

El *National Center for Biotechnology Information* (NCBI) es parte de la *National Library of Medicine* (NLM), una rama de los *National Institutes of Health* (NIH). Está localizado en Bethesda, Maryland, en Estados Unidos. Fue fundado el 4 de noviembre de 1988 con la misión de ser una importante fuente de información en biología molecular. Almacena y constantemente actualiza la información referente a secuencias genómicas en **GenBank**. En **PubMed** se maneja un índice de artículos científicos referentes a biomedicina, biotecnología, bioquímica, genética y genómica, una recopilación de enfermedades genéticas humanas, además de otros datos biotecnológicos de relevancia. Todas las bases de datos del NCBI están disponibles en línea de manera gratuita.

<http://www.ncbi.nlm.nih.gov>

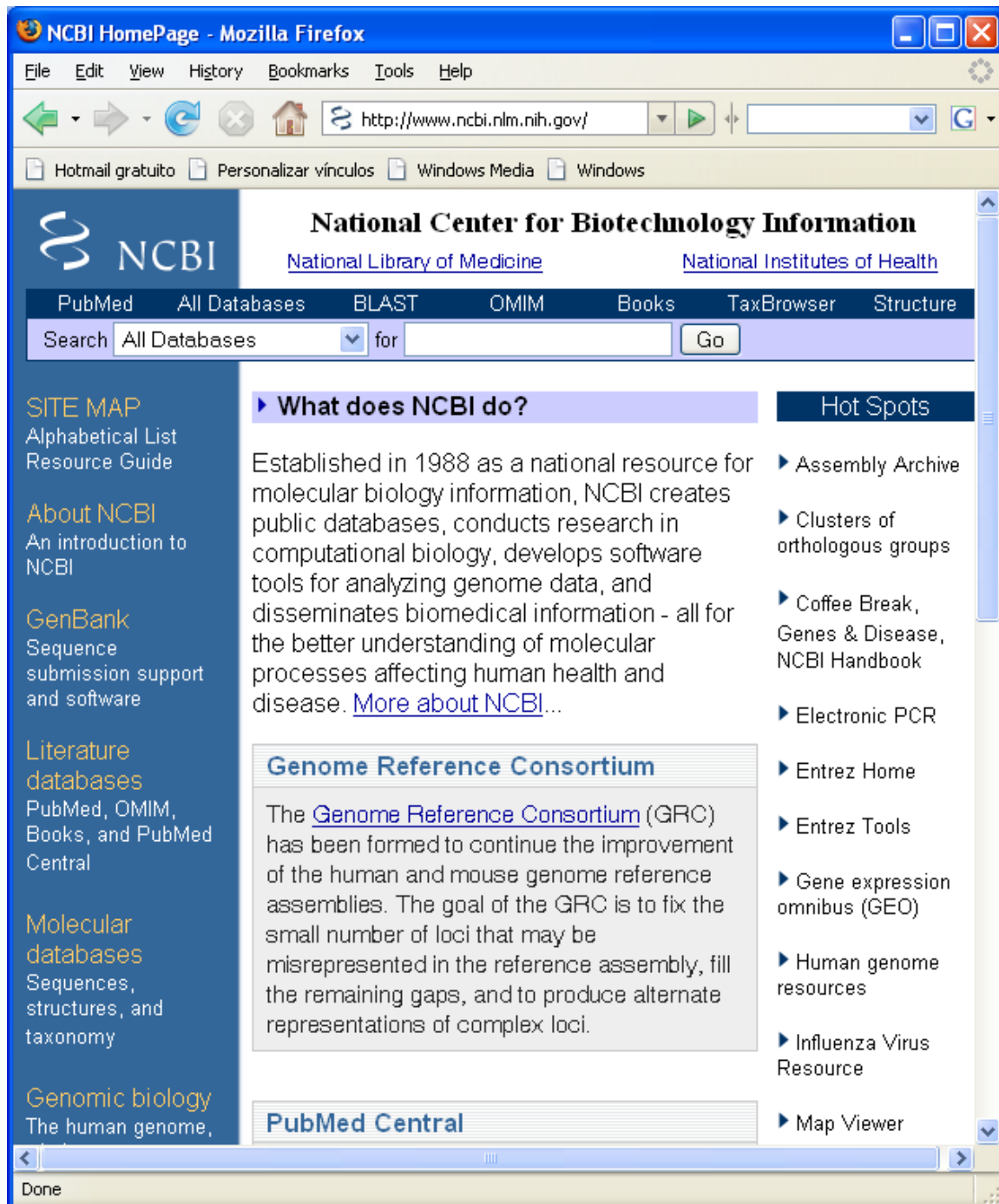


Figura 6-1. Página principal del NCBI

El NCBI ofrece además, algunas herramientas bioinformáticas para el análisis de secuencias de ADN, ARN y proteínas, siendo BLAST una de las más usadas. GenBank se coordina con otras bases de datos como la del instituto europeo de bioinformática (<http://www.ebi.ac.uk>) y la base de datos de DNA de Japón (<http://www.ddbj.nig.ac.jp/Welcome.html.en>). Otra instituto importante que alberga varias bases de datos nucleotídicas es el instituto de investigación genómica (*The Institute for Genomic Research* o

TIGR) que ahora ha sido absorbido por un consorcio liderado por Craig Venter.

<http://www.tigr.org>



Figura 6-2. Página principal de TIGR

6.3.2. Maize GDB

La página de *Maize GDB* es la base de datos más completa y más útil para los investigadores con un interés en fitomejoramiento. En *MaizeGDB* se pueden encontrar los resultados de proyectos de mapeo, QTLs e información genotípica y fenotípica sobre diversas poblaciones de maíz. Actualmente existen varios mapas genéticos disponibles para maíz, entre ellos

La línea B73 es una línea progenitora de varios híbridos que tuvieron bastante distribución agrícola y comercial en Estados Unidos. El híbrido B73xMo17 dio a lugar un set de líneas recombinantes (*Recombinant Inbred Lines* o RILs) que se conocen en el medio científico como población IBM (*intermated B73xMo RIL population*) y se usa mucho para proyectos de mapeo genético. Con referencia a las bases de datos de maíz, se puede recomendar el siguiente link:

<http://www.maizegdb.org>



Figura 6-3. Página inicial de MaizeGDB

6.3.3. Maize Gene Index

Actualmente se han secuenciado de manera masiva dos genotipos de maíz. En las bases de datos públicas de USA y Europa se usa el genotipo B73, que es una línea homocigótica para climas templados. Por otro lado, en **CINVESTAV Irapuato** se secuenciaron las islas de genes (regiones hipometiladas) de una mezcla de DNA de varios individuos de una población criolla de maíz mexicano (el maíz palomero cónico toluqueño, un genotipo del CIMMYT llamado *Mexi-5*). Sin embargo, el genoma de maíz palomero todavía no se ha ensamblado ni anotado totalmente, por lo que por el momento existe mucho más información pública para el maíz B73. Por ejemplo, ya se conocen casi la totalidad de los genes de B73, que son como 56 mil secuencias consenso tentativas.

Con referencia al genotipo maíz B73, las bases de datos con más información nucleotídica, incluyendo una lista casi completa de los genes de maíz (Maize Gene Index) se pueden consultar en los siguientes links:

<http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=maize>

<http://www.maizegenome.org>

<http://maize.tigr.org>

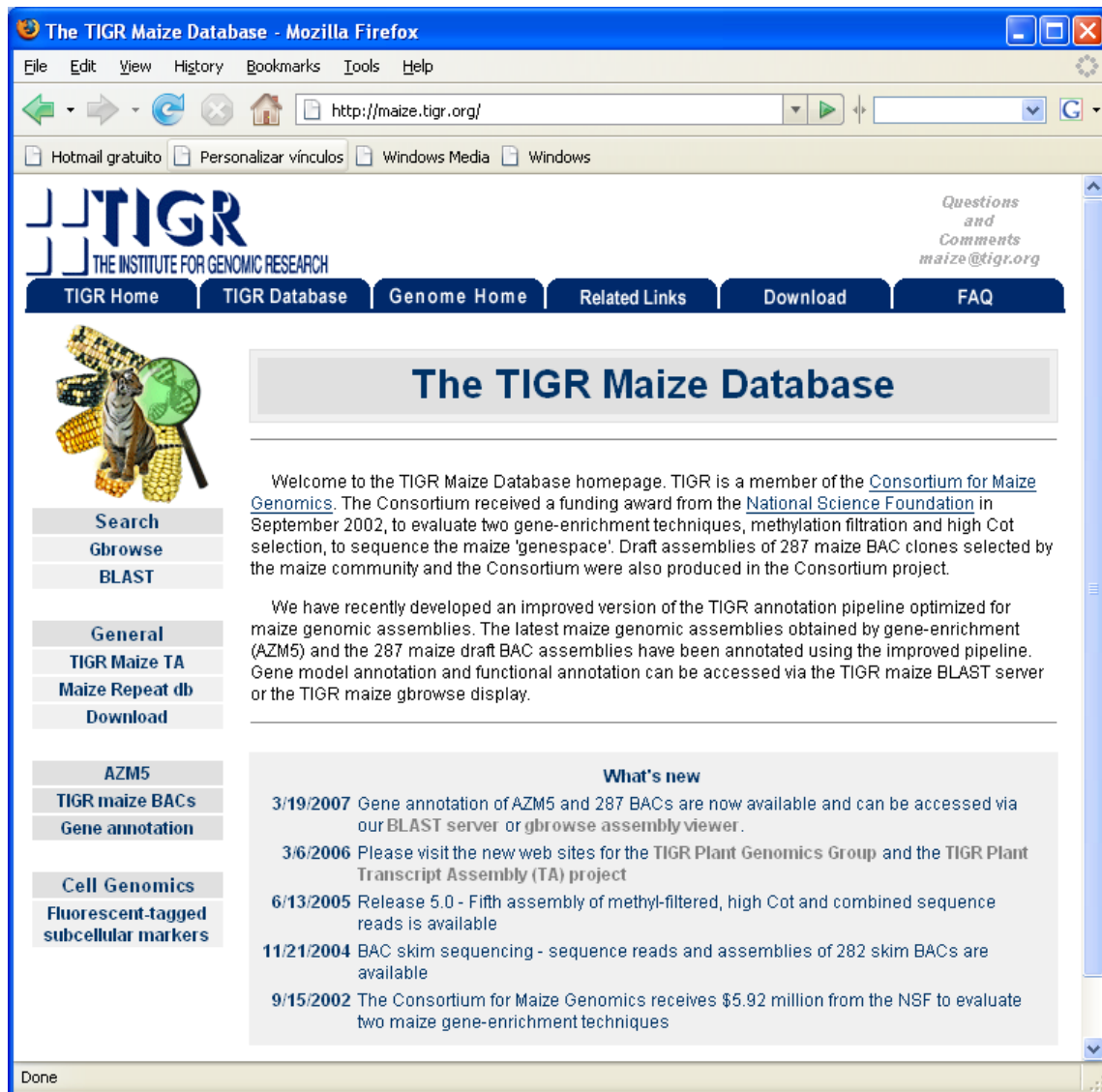


Figura 6-4. Página inicial de TIGR maíz.

6.3.4. CIMMYT

Una de las instituciones líderes de mejoramiento de maíz es el CIMMYT. A diferencia de las compañías comerciales, el CIMMYT trabaja para el ámbito público y se enfoca más bien a los agricultores de los lugares más marginados. Su página web contiene información sobre los diversos proyectos que se desarrollan tanto en México como en África y otras partes

del mundo. En los vínculos también se puede encontrar un formulario para pedir semilla del banco de germoplasma del CIMMYT.

<http://www.cimmyt.org>



6.3.5. IRRI

La institución líder para el mejoramiento de arroz es el IRRI, que esta en las Filipinas.

<http://www.irri.org>

Un sitio con información interesante sobre mejoramiento genético es el Cereal Knowledgebank, que es una base de datos conjunta entre el CIMMYT y el IRRI. Incluye información sobre los 3 cereales más importantes del mundo: maíz, trigo y arroz.



Tarea 6-2

Visite la siguiente página y averigüe más sobre las metodologías de mejoramiento que se usan en diferentes cereales.
<http://www.knowledgebank.irri.org/>

6.4. Paquetes de Software

6.4.1. Fieldbook para maíz

Fieldbook, es un programa que desarrollaron los investigadores del CIMMYT para el manejo de inventarios, ensayos y viveros de maíz. En esencia, es una serie de archivos y documentos maestros de Excel que se manipulan con algunos macros escritos en *visual basic*. Se definen archivos maestros de Excel (templates) que se usan; para la preparación de la semilla (*SeedPrep*), la impresión de los sobres, y para la preparación de los libros de campo (*trials or nurseries*). También se pueden hacer resúmenes de ciclos (*SeasonSummary*), así como envíos de semillas (*SeedShipment*). En los archivos Excel se introducen los datos fenotípicos, tanto de los ensayos como de los viveros. Con ayuda del programa se pueden hacer los cálculos de rendimiento y los análisis estadísticos para compensar la variabilidad de campo. Para ello se exportan los datos para ser analizados por un programa estadístico llamado REML que a su vez regresa los

resultados en el mismo formato de Excel. En resumen, es un paquete de macros y archivos de Excel, que le ayudan a un investigador a tener un programa de fitomejoramiento organizado y coherente.

El programa puede ser descargado gratuitamente en los siguientes dos sitios:

<http://www.cimmyt.org/dtma/mFinder/fb8.4.7.zip>

<ftp://ftp.cgiar.org/cimmyt/FIELDBOOK/FIELDBOOK>

6.4.2. Alternativas gratuitas de software

Muchos estudiantes no tienen recursos para comprar software comercial, o no quieren usarlos para ello, y es por eso que recurren al copiado ilegal de programas. Las personas honestas, respetuosas y comprometidas con la innovación científica no deberían desacreditar el valor de la propiedad intelectual. Las empresas invierten mucho dinero en investigación y desarrollo merecen una retribución económica por ello. Piratear programas de software con *copyright* restringido es como robarle patentes a otras empresas, o copiarle la tesis a un compañero. ¿Por qué hacerlo si no hay necesidad? Hoy en día existen alternativas muy atractivas de software gratuito bajo un esquema de licenciamiento GPL.

6.4.2.1. Sistema operativo Ubuntu

Uno de los sistemas operativos con más crecimiento en los últimos años es Ubuntu. Además de que es gratuito y de código libre, corre mucho más rápido que Windows Vista, y por si fuera poco, no tiene problemas de virus o de *spyware*. Instale Ubuntu en su computadora y Pruébalo. El sistema Ubuntu se puede descargar del siguiente sitio:

<http://www.ubuntu.com/>



Cuando aprenda más de los nuevos sistemas Linux, verá que existe un sinnúmero de programas disponibles y que puede hacer todo lo que hacía antes en Windows, con programas gratuitos, flexibles y poderosos. Hay más herramientas bioinformáticas disponibles en Linux que en los otros sistemas operativos (Windows y Mac). Si quiere incursionar más a fondo en la bioinformática y el cómputo biológico avanzado, no hay mejor forma que aprendiendo a manejar Ubuntu y usar los programas tipo Linux.

6.4.2.2. Open Office

¿Cuanto cuesta la licencia para tener instalado de manera legal el paquete de Microsoft Office en su computadora? Los que no tienen dinero para

pagarla, en lugar de usar una copia pirata, pueden probar una alternativa diferente. Una de las mejores es el OpenOffice. Tanto en Ubuntu como en Windows el paquete de Open Office puede ser usado para manejar archivos Word, Excel y PowerPoint.

Descargue el Open Office de forma gratuita del siguiente link:



<http://es.openoffice.org/>

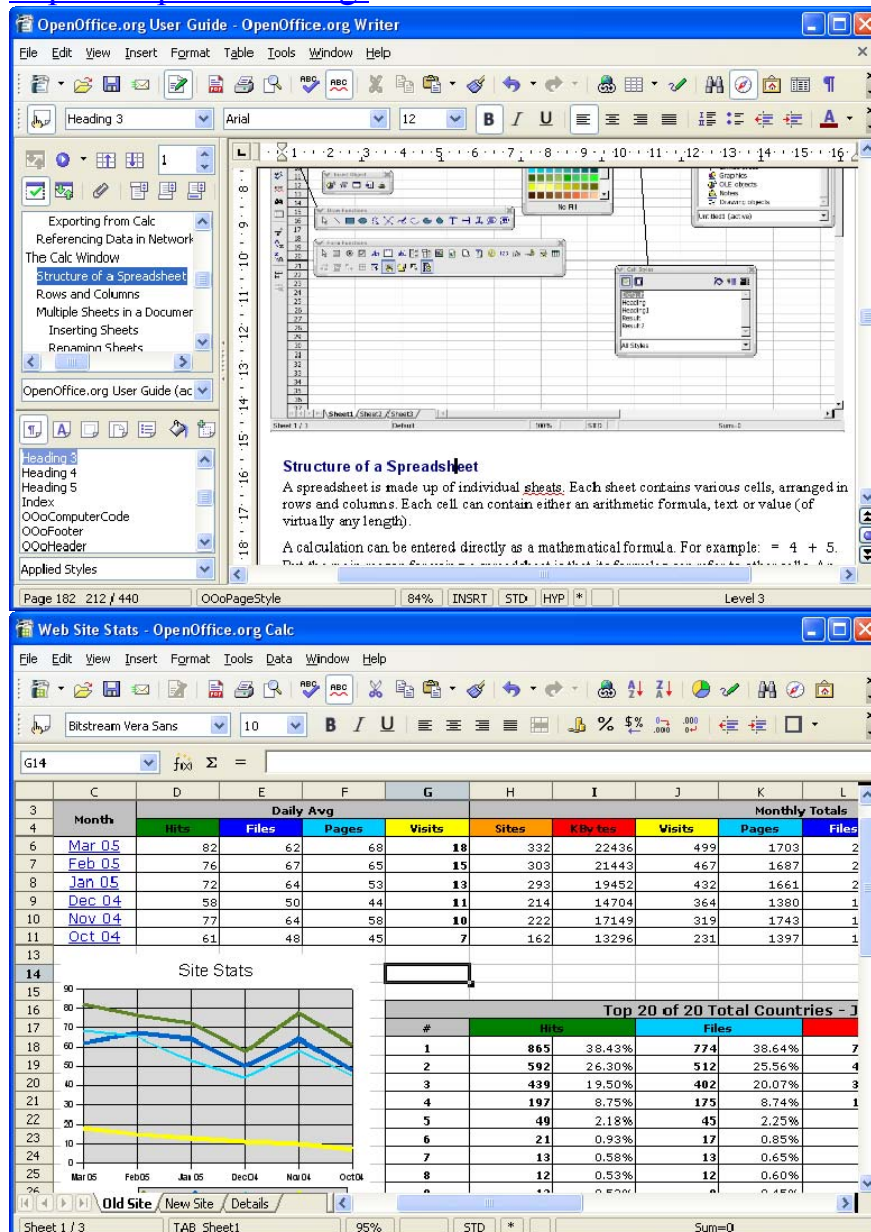


Figura 6-5. Ventanas de programas de OpenOffice

6.4.2.3. Emboss

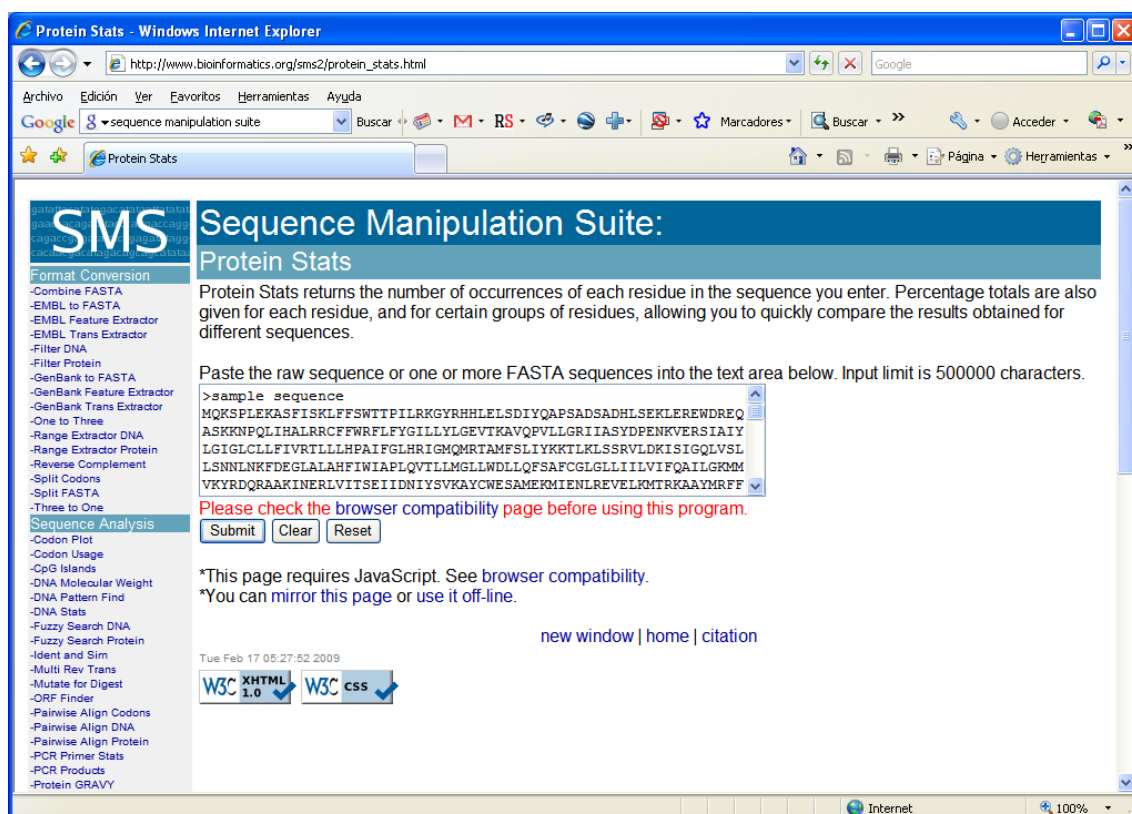
EMBOSS is "The European Molecular Biology Open Software Suite". EMBOSS is a free Open Source software analysis package specially developed for the needs of the molecular biology user community. The software automatically copes with data in a variety of formats and even allows transparent retrieval of sequence data from the web. Extensive libraries are provided with the package. It is a platform that allows other scientists to develop and release software in true open source spirit. EMBOSS also integrates a range of sequence analysis tools into a seamless whole. EMBOSS breaks the historical trend towards commercial software packages.

<http://emboss.sourceforge.net/>

6.4.2.4. Sequence Manipulation Suite

El Sequence Manipulation Suite (SMS) es una colección de programas de JavaScript para generar, formatear y analizar secuencias de DNA y proteínas. Es un paquete gratuito basado en una serie de páginas web que puede correr de manera local en cualquier buscador (navegador) como Explorer o Firefox. Para probar y descargar el SMS consulte la página oficial:

<http://www.bioinformatics.org/sms2/>



El SMS tiene una serie de funciones. La siguiente lista muestra algunos ejemplos de las funciones de este paquete que es gratuito y fácil de usar.

Format Conversion:

- **GenBank to FASTA** - accepts one or more GenBank files as input and returns the entire DNA sequence from each in FASTA format. Use this program when you wish to quickly remove all of the non-DNA sequence information from a GenBank file.
- **One to Three** - converts single letter translations to three letter translations.
- **Reverse Complement** - converts a DNA sequence into its reverse, complement, or reverse-complement counterpart. The entire IUPAC DNA alphabet is supported, and the case of each input sequence character is maintained. You may want to work with the reverse-complement of a sequence if it contains an ORF on the reverse strand.

Sequence Analysis:

- **Codon Plot** - accepts a DNA sequence and generates a graphical plot consisting of a horizontal bar for each codon. The length of the bar is proportional to the frequency of the codon in the codon frequency table you enter. Use Codon Plot to find portions of DNA sequence that may be poorly expressed, or to view a graphic representation of a codon usage table (by using a DNA sequence consisting of one of each codon type).
- **DNA Molecular Weight** - accepts one or more DNA sequences and calculates molecular weight. Sequences can be treated as double-stranded or single-stranded, and as linear or circular. Use DNA Molecular Weight when calculating molecule copy number.
- **DNA Pattern Find** - accepts one or more sequences along with a search pattern and returns the number and positions of sites that match the pattern. The search pattern is written as a JavaScript regular expression, which resembles the regular expressions written in other programming languages, such as Perl.
- **DNA Stats** - returns the number of occurrences of each residue in the sequence you enter. Percentage totals are also given for each residue, and for certain groups of residues, allowing you to quickly compare the results obtained for different sequences.
- **Fuzzy Search DNA** - accepts a DNA sequence along with a query sequence and returns sites that are identical or similar to the query. You can use this program, for example, to find sequences that can be easily mutated into a useful restriction site.
- **Mutate for Digest** - accepts a DNA sequence as input and searches for regions that can easily be mutated to create a restriction site of interest. The program also reports protein translations so that you can see which reading frames are altered by the proposed mutations. Use Mutate for Digest to find sequences that can be converted to a useful restriction site using PCR or site-directed mutagenesis.
- **Multi Rev Trans** - accepts a protein alignment and uses a codon usage table to generate a degenerate DNA coding sequence. The program also returns a graph that can be used to find regions of minimal degeneracy at the nucleotide level. Use Multi Rev Trans when designing PCR primers to anneal to an unsequenced coding sequence from a related species.
- **ORF Finder** - searches for open reading frames (ORFs) in the DNA sequence you enter. The program returns the range of each ORF, along with its protein translation. ORF Finder supports the entire IUPAC alphabet and several genetic codes. Use ORF Finder to search newly sequenced DNA for potential protein encoding segments.
- **Pairwise Align DNA** - accepts two DNA sequences and determines the optimal global alignment. Use Pairwise Align DNA to look for conserved sequence regions.
- **PCR Primer Stats** - accepts a list of PCR primer sequences and returns a report describing the properties of each primer, including melting temperature, percent GC content, and PCR suitability. Use PCR Primer Stats to evaluate potential PCR primers.
- **PCR Products** - accepts one or more DNA sequence templates and two primer sequences. The program searches for perfectly matching primer annealing sites that can generate a PCR product. Any resulting products are sorted by size, and they are given a title specifying their length, their position in the original sequence, and the primers that produced them. You can use linear or circular molecules as the template. Use PCR Products to determine the product sizes you can expect to see when you perform PCR in the lab.
- **Protein GRAVY** - Protein GRAVY returns the GRAVY (grand average of hydropathy) value for the protein sequences you enter. The GRAVY value is calculated by adding the hydropathy value for each residue and dividing by the length of the sequence ([Kyte and Doolittle; 1982](#)).
- **Protein Isoelectric Point** - calculates the theoretical pI (isoelectric point) for the protein sequence you enter. Use Protein Isoelectric Point when you want to know approximately where on a 2-D gel a particular protein will be found.
- **Protein Molecular Weight** - accepts one or more protein sequences and calculates molecular weight. You can append copies of commonly used epitopes and fusion proteins using the supplied list. Use Protein Molecular Weight when you wish to predict the location of a protein of interest on a gel in relation to a set of protein standards.
- **Protein Stats** - returns the number of occurrences of each residue in the sequence you enter. Percentage totals are also given for each residue, and for certain groups of residues, allowing you to quickly compare the results obtained for different sequences.
- **Restriction Digest** - cleaves a DNA sequence in a virtual restriction digest, with one, two, or three restriction enzymes. The resulting fragments are sorted by size, and they are given a title specifying their length, their position in the original sequence, and the enzyme sites that produced them. You can digest linear or circular molecules, and even a mixture of molecules (by entering more than one sequence in FASTA format). Use Restriction Digest to determine the fragment sizes you will see when you perform a digest in the lab.
- **Restriction Summary** - accepts a DNA sequence and returns the number and positions of commonly used restriction endonuclease cut sites. Use this program if you wish to quickly determine whether or not an enzyme cuts a particular segment of DNA.
- **Reverse Translate** - accepts a protein sequence as input and uses a codon usage table to generate a DNA sequence representing the most likely non-degenerate coding sequence. A consensus sequence derived from all the possible codons for each amino acid is also returned. Use Reverse Translate when designing PCR primers to anneal to an unsequenced coding sequence from a related species.
- **Translate** - accepts a DNA sequence and converts it into a protein in the reading frame you specify. Translate supports the entire IUPAC alphabet and several genetic codes.

7. Herramientas estadísticas

7.1. Conceptos básicos de estadística

7.1.1. ¿Para que sirve la estadística?

Para poder calcular algo en estadística, lo más importante es tener suficientes repeticiones. El mantra de los experimentos biológicos es: **siempre incluir repeticiones y controles**. De preferencia, hacer las más repeticiones posibles. El principal problema de los experimentos agronómicos para mejoramiento genético, es que no se hacen suficientes repeticiones para tener mayor certeza de los resultados. En un mundo ideal para un fitomejorador, sería deseable tener 5 repeticiones de un mismo genotipo sembradas en 50 diferentes localidades. En el mundo real muchas veces no se pueden hacer tantas repeticiones por cuestiones de espacio, costo, esfuerzo o de tiempo. Por ejemplo, a veces solo se hacen 2 repeticiones por localidad, y muchas veces solo se siembran menos de 5 localidades diferentes. Algunas veces no se tiene suficiente semilla para evaluarla en tantos ensayos, otras veces simplemente no se tienen las facilidades o el dinero.

La estadística se puede usar para resolver este problema de limitación de recursos. Se podría decir que **es un conjunto de herramientas matemáticas que permiten sacar el mayor provecho con el menor esfuerzo experimental posible**. Es decir, la bioestadística permite llegar a una conclusión lo más robusta posible, con el número limitado de repeticiones que tiene uno disponibles. La ventaja de la estadística es que calcula un resultado y al mismo tiempo da un valor sobre la confiabilidad de ese resultado en términos probabilísticos. A esto se le llama niveles de confianza, y se les clasifica en errores del tipo alfa (α) y beta (β). Para entender más a fondo la estadística, es necesario repasar algunos de sus fundamentos y conceptos básicos.

7.1.2. Medidas centrales

Con una serie de datos relativamente grande uno puede calcular las medidas centrales como el promedio, la mediana y la moda. En un conjunto de valores, por lo regular, el valor promedio es el más importante. Al **promedio** también se le llama **media aritmética**. Esta medida no se debe confundir con la **mediana** que es el valor central de un conjunto de números. La mediana es el número que se encuentra en medio, es decir, la mitad de los números es mayor que la mediana y la otra mitad es menor. La **moda** es el valor que se repite con mayor frecuencia.

Ninguna de estas medidas de tendencia central tomada individualmente proporciona una imagen completa de los datos. Supongamos que los datos

están agrupados en tres áreas, en una hay un valor bajo que se repite mucho; en la otra, hay muchos valores promedios y en el otro tercio hay valores más extremos, pero cada vez menos frecuentes. Tanto PROMEDIO como MEDIANA devolverán un valor situado en una zona central, y MODA devolverá el valor bajo dominante. (Ver Figura 7-1)

Lo que hay que recordar es que cuando la distribución de los datos es compleja, los valores de media, mediana y moda darán valores muy diferentes (ver Figura 7-1), mientras que con una distribución normal, los valores serán muy parecidos (ver Figura 7-2).

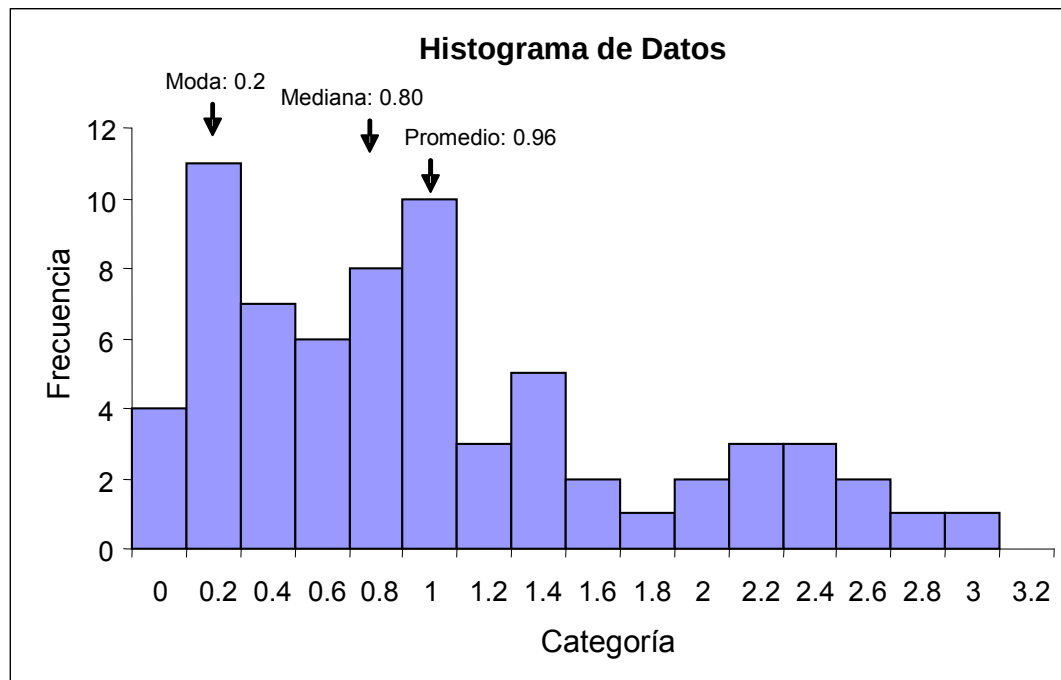


Figura 7-1. Ejemplo de histograma de datos con una distribución compleja. Se puede ver, que los valores del promedio, la mediana y la moda son diferentes uno del otro.

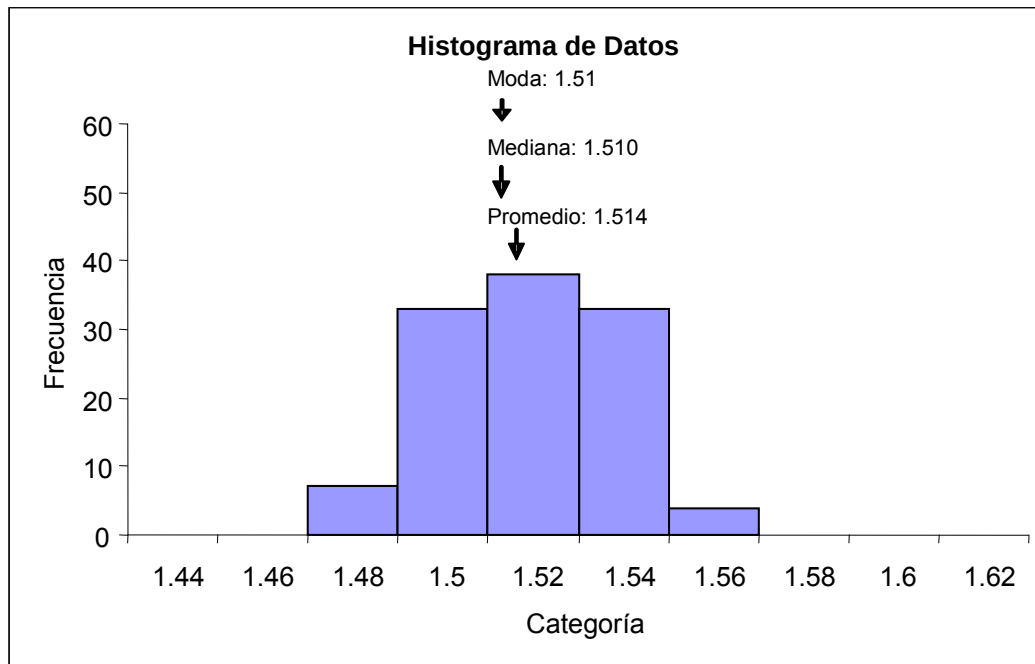


Figura 7-2. Ejemplo de histograma de datos con una distribución seminormal. En este caso los valores del promedio, la mediana y la moda son muy parecidos.

7.1.3. Desviación estándar y distribución normal

Cuando se tiene una serie de datos con sus respectivas repeticiones, primero se calcula el promedio. Esto se hace con la simple fórmula de sumar todos los valores y dividirlo entre el número de mediciones. Sin embargo, el promedio no dice nada sobre la repetibilidad y por eso es necesario hacer más cálculos para reducir la variabilidad de los datos. La desviación estándar (SD) es una medida de dispersión para variables numéricas (variables de intervalo). El concepto es de gran utilidad en la [estadística descriptiva](#). La desviación estándar (símbolo sigma, σ) informa sobre el promedio de las distancias que tienen los datos respecto de su media aritmética, expresada en las mismas unidades que la variable. El término desviación estándar fue incorporado a la estadística por Karl Pearson en 1894.

Figura 7-3. Distribución normal mostrando el porcentaje de valores que hay dentro de cada rango.

La distribución normal es una curva estandarizada que está basada en la fórmula de Gauss.

$$\varphi_{\mu, \sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sigma} \varphi\left(\frac{x-\mu}{\sigma}\right), \quad x \in \mathbb{R},$$

Para conocer con detalle un conjunto de datos, no basta con conocer las medidas de tendencia central, sino que necesitamos conocer también la desviación que representan los datos en su distribución respecto de la media aritmética de dicha distribución. El objetivo es tener una visión de los mismos más acorde con la realidad a la hora de describirlos e interpretarlos para la toma de decisiones. A esto se le llama prueba de hipótesis.

La [varianza](#) representa la media aritmética de las desviaciones con respecto a la [media](#), elevadas al cuadrado. Si atendemos a la colección completa de datos (la población en su totalidad) podemos obtener la varianza poblacional; y si por el contrario prestamos atención sólo a una muestra de la población, obtenemos en su lugar la **varianza muestral**. En la rutina de los investigadores casi siempre se usan las medias y varianzas muestrales, casi nunca las poblacionales. La varianza y la desviación estándar están estrechamente relacionadas. La desviación estándar es la raíz cuadrada positiva de la [varianza](#).

Expresión de la varianza muestral:

$$S_X^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1} = \frac{\sum X_i^2}{n - 1} - \bar{X}^2$$

Expresión de la desviación estándar muestral:

$$\sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

7.1.4. Tipos de errores

En las mediciones experimentales podemos distinguir entre dos tipos de errores. El error aleatorio debido al azar, y el error sistemático debido a algún desperfecto del instrumento, o problemas con la referencia o la curva de calibración. Por lo regular se trata de obtener la mejor **precisión** posible, con el más alto grado de **repetibilidad** (Figura 7-4).

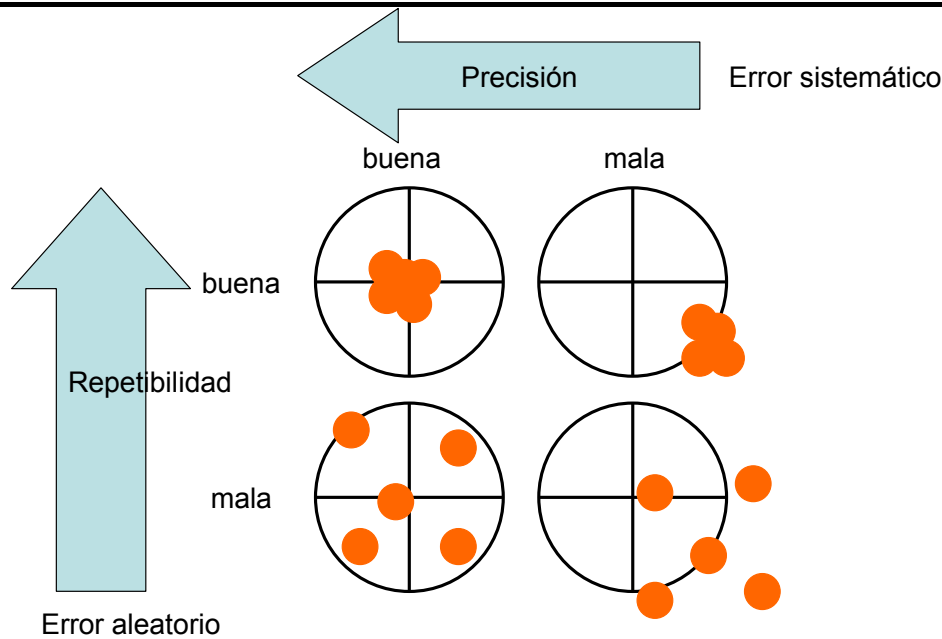


Figura 7-4. Tipos de errores y su relación con precisión y repetibilidad.

7.2. Diseño experimental

Los pasos para la realización de un experimento son:

- 1) **Estudio del arte previo.** Antes de hacer un experimento se debe tener un panorama de lo que ya se sabe (dominio del conocimiento general) y las interrelaciones con otras ramas científicas y del conocimiento. Después de eso se debe hacer la pregunta: ¿Qué es lo que todavía no se sabe? Solamente así se pueden plantear preguntas novedosas que contribuyan al avance de nuestro conocimiento.
- 2) **Planteamiento del problema** de investigación. Esto incluye la definición de **una pregunta concreta y relevante**. En base a esa pregunta específica se define una respuesta teórica que llamamos **hipótesis**. Una hipótesis es la respuesta favorita a una pregunta biológica pero que todavía no tenemos datos suficientes para sustentarla de manera rigurosa. Cuando se trata de una pregunta estadística, por ejemplo, el efecto de un factor sobre un resultado, se define una hipótesis nula (H_0) y una hipótesis alterna:

La hipótesis nula: Por lo regular predice que el azar es la causa principal del resultado que obtuvimos. Solo cuando tenemos evidencia experimental (suficiente repetibilidad) de que los resultados no se dieron por azar, es cuando rechazamos la hipótesis nula y decimos con cierto grado de certeza (nivel de significancia) que la hipótesis alterna es válida.

En estadística, una hipótesis nula es una hipótesis construida para anular o refutar. La hipótesis nula se presume verdadera hasta que los datos (prueba de hipótesis) indiquen lo contrario.

Ejemplo de Hipótesis nula para Chi cuadrado: "Si este material genético segrega en proporciones mendelianas, no habrá diferencias entre las frecuencias observadas (O) y las frecuencias esperadas (E)"

Ejemplo de hipótesis nula para student: "Si la humedad relativa no influye sobre el número de huevos por desove, no habrá diferencias entre las medias de esta variable entre las regiones húmedas o secas."

- 3) **Modelo y metodología:** En base a la pregunta planteada, se busca el mejor método para contestarla. Por ejemplo, si queremos saber cuál de todos las variedades de maíz es mejor, lo que debemos hacer es un ensayo de campo para ver cuál tiene mayor rendimiento de grano. El diseño experimental incluye la elección de los genotipos, los controles, el procedimiento a seguir, es decir: el diseño, las repeticiones, las localidades, el o los tratamientos a aplicar, las variables consideradas: variables independientes, variables dependientes, variables codependientes, variables extrañas, etc.

Tipos de variables:

* Variables independientes: Son las variables que el experimentador puede controlar y define *a priori* en el experimento. Por ejemplo, los genotipos, la irrigación, los tratamientos, etc. Se supone que estas variables determinan los resultados del experimento.

* Variables dependientes: Son características de la realidad que se ven determinadas ya sea por el azar o por los factores que se definieron o controlaron en el experimento. La clasificación "dependiente" indica conocimiento *a priori* de que esta variable es efecto y no causa. Es decir, depende del tratamiento y no al revés. El valor de la variable dependiente puede ser resultado del azar (hipótesis nula), o ser efecto directo o indirecto de otros fenómenos o variables desconocidas. Sólo cuando existe suficiente asociación con la variable independiente se dice que aceptamos la hipótesis alterna.

* Variable continua: Se presenta cuando el fenómeno que se mide puede tomar valores cuantitativos, por ejemplo el rendimiento, ya que esta variable puede asumir valores de rango continuo: 1, 2, 3, o 15 toneladas por hectárea.

* Variables discretas: Son aquellas que establecen categorías en términos no cuantitativos. Por ejemplo cuando se clasifican las localidades o los ambientes en: normal, sequía, bajo en nitrógeno, etc., o cuando se clasifica el germoplasma por color: blanco, amarillo, azul, etc.

El número de variables que se incluyan en un experimento, depende del investigador y del fenómeno que estudie. Mientras más variables independientes se agreguen, quizá obtenga una mayor explicación de los cambios en su variable dependiente. Por ejemplo: si quiere explicar el rendimiento de grano puede controlar variables como el genotipo, los alelos, los QTLs, la localidad, los riegos, la fertilidad del suelo, etc. También se pueden incluir variables independientes, pero que no son controlables a campo abierto, como intensidad de luz, temperatura, vientos, lluvias, etc. Además de la variable dependiente principal que es

el rendimiento de grano, también se pueden analizar variables codependientes como: el número de mazorcas, el tamaño de grano, días de floración, el intervalo de floración femenina-masculina (ASI), la senescencia, la altura de planta, etc. Si sólo se utiliza una variable, será difícil creer que el rendimiento depende sólo de un factor. La relación causa-efecto no siempre es tan evidente como quiere el investigador. Entre más variables se incluyan en el análisis mayor será la certeza de detectar efectos significativos y confiables.

- 4) **Resultados:** los datos obtenidos se deben organizar de manera sistemática y congruente. Usando estos datos se hace un análisis estadístico para describir cuáles fueron las relaciones observadas entre las variables. Se debe contestar si la o las variables independientes realmente influyeron significativamente sobre los valores de las variables dependientes. También se debe analizar si hubo tantas variables conocidas o extrañas como se pensaba, o si surgieron otras. Los resultados por lo general se presentan a manera de tablas o gráficas (de barras, de pastel, etc.). Todas las formas de representación de resultados deben incluir indicadores de variabilidad experimental con grados de confiabilidad (desviación estándar, barras de error, repetibilidad, resultados de pruebas de T, etc.)
- 5) **Conclusiones:** En la investigación y la experimentación, las conclusiones son determinaciones hechas mediante el estudio metódico de los resultados del trabajo precedente dentro de una cierta metodología. Las conclusiones toman a menudo la forma de teorías (explicación general fenomenológica). La conclusión es típicamente el resultado de una discusión de los datos considerando las premisas. Sin una discusión de las premisas, no hay conclusión, si son sólo aseveraciones sin evidencia, es una alegación. Naturalmente, la precisión de una conclusión dada, es dependiente de la verdad de la conclusión elegida.
- 6) **Difusión y divulgación.** A partir de las conclusiones ya es posible pensar en la elaboración de un informe que puede ser oral o escrito, y tiene la finalidad de comunicar los resultados a la sociedad y a la comunidad científica. La publicación del experimento y sus conclusiones se puede presentar en un congreso en forma de una charla, o hacerse a través de un artículo en una revista nacional o internacional.

Como muchas cosas en la ciencia, el proceso experimental es cíclico e iterativo, de forma que una pregunta o experimento genera una serie de datos cuya interpretación da origen a muchas preguntas más, que a su vez pueden ser contestadas al repetir los experimentos variando las

condiciones, eligiendo nuevos componentes o alterando las premisas (Figura 7-5).

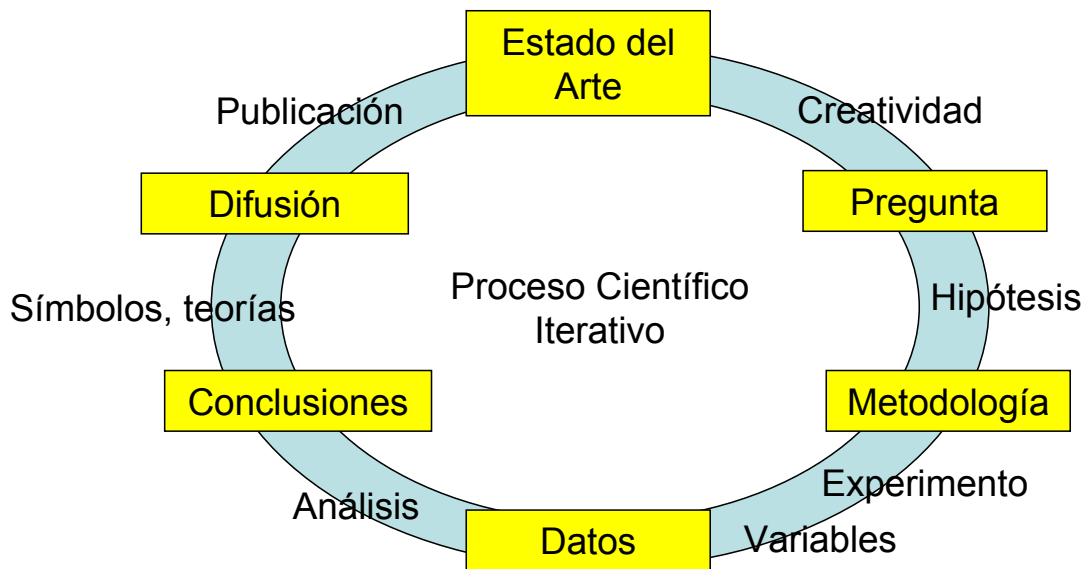


Figura 7-5. Proceso iterativo científico.

7.3. Análisis estadísticos

7.3.1. Ejemplo de análisis usando pruebas de T

Ahora que hemos repasado un poco las bases de la estadística descriptiva, trataremos de profundizar más con algunos ejemplos. Trataremos de ver como la estadística cuantitativa permite generar conclusiones con relevancia biológica (inferencia estadística). Supongamos que queremos saber si dos productos (un agroquímico y un fertilizante orgánico) incrementan el rendimiento de maíz. Para ello se diseña un experimento en el cual se usan los productos solos o de manera combinada.

Producto		Fertilizante	
Químico	Tratamiento XA	Tratamiento XB	
		Fertilizante:	
	Fertilizante:	Fertilizante:	
		0 kg	
	Químico:	Químico:	
		0 lt	
	Tratamiento YA	Tratamiento YB	
		Fertilizante:	
	Fertilizante:	Fertilizante:	
		2 kg	
	Químico:	Químico:	
		0 lt	

El tratamiento XA corresponde al control (sin químico y sin fertilizante), mientras que el tratamiento YB incluye los dos productos de manera combinada (con químico y con fertilizante). Para cada condición (nivel de tratamiento) se hacen 10 repeticiones que se manejan dentro de un campo experimental. En total se siembran 40 parcelas al azar. A continuación se muestra una tabla con los datos que se obtuvieron del experimento.

Rendimiento de maíz en kilos por parcela				
	Tratamientos			
	XA	XB	YA	YB
rep1	6	9	13	8
rep2	7	12	12	9
rep3	5	12	12	8
rep4	4	12	12	11
rep5	6	11	11	12
rep6	7	13	11	10
rep7	8	11	11	12
rep8	6	10	10	14
rep9	9	10	10	12
rep10	9	9	14	11

La tarea ahora consiste en hacer el análisis estadístico para saber si los diferentes tratamientos tienen un efecto significativo sobre el rendimiento de maíz. Explicaremos con estos datos algunos tipos de análisis estadísticos, por ejemplo, la prueba T y el análisis de varianza (ANOVA). Usaremos las funciones de Excel (o Spreadsheet de OpenOffice) para hacerlo.

Lo primero que se debe calcular es el número de observaciones, la suma, el promedio, la desviación estándar y el error estándar para cada tratamiento por separado. (Si no sabe usar Excel consulte primero la sección 7.4.1 de este capítulo)

	XA	XB	YA	YB	
NumObs	10	10	10	10	=CONTAR(B4:B15)
Suma	67	109	116	107	=SUMA(B4:B15)
Promedio	6.70	10.90	11.60	10.70	=PROMEDIO(B4:B15)
SD	1.636	1.370	1.265	1.947	=DESVEST(B4:B15)
SE	0.517	0.433	0.400	0.616	=B19/RAIZ(B16)

Podemos ver que los promedios tienen valores diferentes para cada tratamiento. La pregunta ahora es saber si las diferencias son significativas. La hipótesis nula en este caso sería: las diferencias entre los tratamientos se

deben al azar y no a los productos químicos que se aplicaron. Para visualizar mejor los resultados hacemos una grafica en Excel usando los valores promedio incluyendo barras de error.

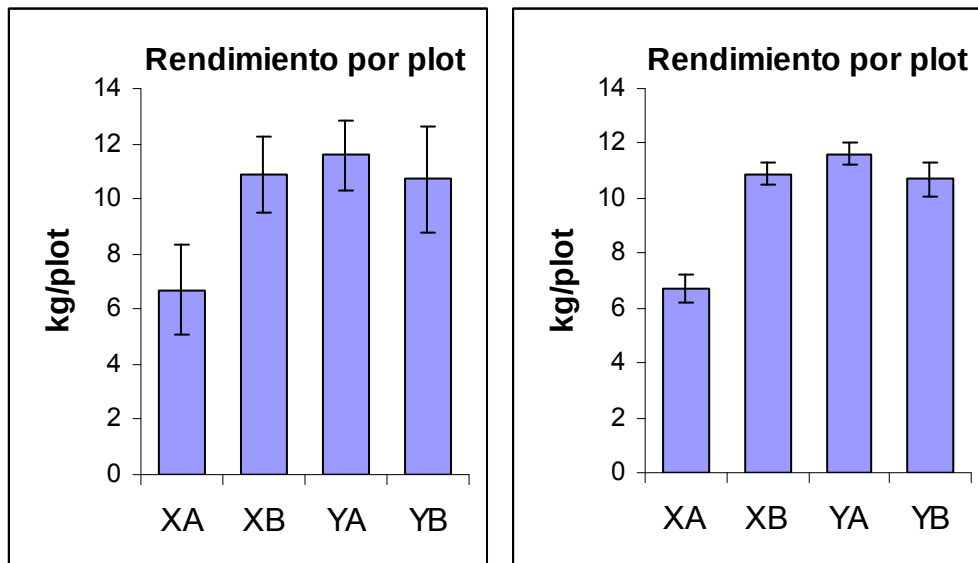


Figura 7-6. Resultados del experimento de rendimiento. En la gráfica de la izquierda se uso el valor SD para las barras de error y en la derecha el valor SE.

Las barras de error son mas pequeñas si se usa el valor SE (*standard error*, error estándar) que el valor SD (*standard deviation*, desviación estándar). La gráfica de la derecha se ve más bonita por tener errores más pequeños. Eso podría ser razón suficiente para poner siempre el valor SE en cualquiera de las graficas que hagamos. Afortunadamente, no solo hay argumentos estéticos sino también técnicos para usar los valores SE en nuestras graficas. Las barras de error son más informativas si se usa el error estándar (SE) que si se usa la desviación estándar (SD), debido a que el valor SE incluye una información importante, como es el número de repeticiones hechas (valor n) que no esta reflejada en el valor SD pero si en el valor SE.

$$SE = \frac{SD}{\sqrt{n}}$$

Con el valor SE uno puede ver intuitivamente que tratamientos fueron diferentes de otros. Una regla empírica dice que si las barras de error (SE) se tocan, entonces las diferencias no son significativas ya que el valor P será mayor a 5% ($P > 0.05$). Sólo cuando las barras de error no se tocan es cuando las diferencias podrían ser significativas ($P < 0.05$). En realidad siempre es necesario hacer pruebas para obtener un valor de P exacto. Solo con un análisis de varianza se puede tener certeza estadística.

Una de las pruebas más usadas es la prueba T. En este caso, usamos la función de Excel PRUEBA.T con la siguiente sintaxis (*matriz, matriz, 1, 3*) que significa una comparación entre dos matrices de datos, con la opción de análisis de una cola con muestras de varianza distinta. La prueba T permite comparar sólo dos grupos entre si, por lo que tenemos que hacer varias de ellas. Para ello hacemos una matriz de valores P en donde comparamos todos los tratamientos contra todos.

Valor de P de significancia estadística				
		4.11E-		
XB		06		=PRUEBA.T(C4:C13,D4:D13,1,3)
		4.53E-		
YA		07	0.125	=PRUEBA.T(C4:C13,E4:E13,1,3)
		5.34E-		
YB		05	0.397	0.119
Prueba.T	XA	XB	YA	

Como podemos ver, sí hay algunas diferencias significativas entre los tratamientos, tanto de los valores promedio como los valores P. El par significativamente más diferente, es la comparación entre el tratamiento XA con el YA. Por otro lado, el tratamiento XB no es significativamente diferente del tratamiento YA, ni del YB tampoco ($P > 0.05$).

Tomando esto en cuenta, podemos poner estrellitas en nuestra grafica para indicar los valores que son significativamente diferentes del control XA.

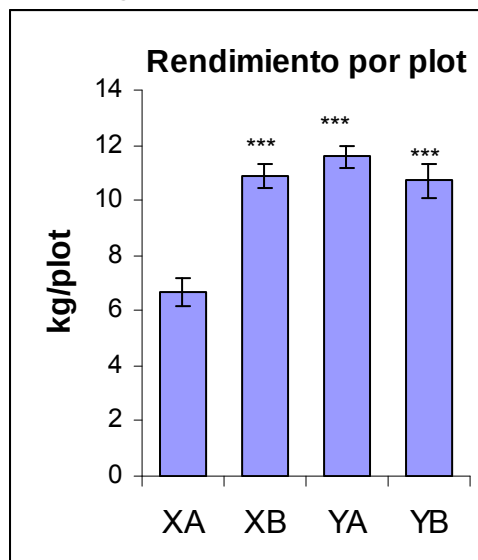


Figura 7-7. Gráfica con significancia estadística. Las estrellitas indican los valores P en la comparación con el tratamiento control XA. Una estrellita significa $P < 0.05$, Dos estrellitas indican $P < 0.01$, mientras que tres estrellitas indican $P < 0.001$.

La interpretación del anterior resultado es que ambos productos si incrementan significativamente el rendimiento. Tanto el fertilizante como el químico tienen un efecto positivo. Sin embargo, el efecto de los productos no es aditivo, ya que cuando se usan de manera combinada, el rendimiento no incrementa más. Si el efecto fuera aditivo, lo que se esperaría, es que el tratamiento YB tuviera un rendimiento más alto que XB o YA.

La razón T.

Antes, cuando no había los programas como Excel, los estadistas calculaban todo a mano y con lápiz. De hecho, al principio ni siquiera se tenían calculadoras de bolsillo. Para obtener sus resultados, se usaban formulas matemáticas y se consultaban tablas de estadística en donde se indicaban los valores de las proporciones T y F para un determinado valor P y grados de libertad (df).

Si bien hoy en día tenemos disponibles las computadoras con programas como Excel y R, no esta de más repasar la manera en que se obtienen algunos valores estadísticos.

	$x_i - \bar{X}$	$(x_i - \bar{X})^2$		$x_i - \bar{X}$	$(x_i - \bar{X})^2$
X1	X		X2	X	
18	3.53	12.48	13	2.67	7.11
17	2.53	6.42	12	1.67	2.78
16	1.53	2.35	12	1.67	2.78
16	1.53	2.35	12	1.67	2.78
16	1.53	2.35	11	0.67	0.44
15	0.53	0.28	11	0.67	0.44
15	0.53	0.28	11	0.67	0.44
	-			-	
15	0.53	0.28	10	0.33	0.11
	-			-	
15	0.53	0.28	10	0.33	0.11
	-			-	
14	0.47	0.22	10	0.33	0.11
	-			-	
14	0.47	0.22	10	0.33	0.11
	-			-	
13	1.47	2.15	9	1.33	1.78
	-			-	
12	2.47	6.08	9	1.33	1.78
	-			-	
11	3.47	12.02	8	2.33	5.44
	-			-	
10	4.47	19.95	7	3.33	11.11

NumObser						
(n)	15			15		
Suma Σ	217	0	67.73	155	0.00	37.33
Promedio						
(X)	14.47			10.33		

$$df=n-1 \quad \text{Varianza} = \frac{\sum (x_i - \bar{X})^2}{(n-1)} \quad SD = \sqrt{\text{Varianza}} \quad SE = \frac{SD}{\sqrt{n}}$$

Teniendo la suma de las diferencias cuadradas y los grados de libertad (df = 14)

Entonces podemos calcular la varianza, la desviación estándar y el error como se muestra a continuación:

	X1		X2	
Varianza	4.838	=67.73/14	2.667	=37.33/14
SD	2.200	= $\sqrt{4.838}$	1.633	= $\sqrt{2.667}$
		=2.2 /		=1.63 /
SE	0.568	$\sqrt{15}$	0.422	$\sqrt{15}$

Ahora, si comparamos los valores de 14.47 y 10.33 vemos que hay una diferencia de 4.13 entre los promedios de los dos grupos de variables X1 y X2. ¿Es significativa esta diferencia? Para responder esta pregunta, hay que calcular la razón T.

Primero hay que calcular el error estándar de la diferencia entre dos medias usando la siguiente formula:

$$S_{x1-x2} = \sqrt{\frac{\sum x_1^2 + \sum x_2^2}{n_1 + n_2 - 2} * \left(\frac{2}{n}\right)}$$

Se procede a sustituir las variables de la fórmula con los valores:

$$S_{x1-x2} = \sqrt{\frac{67.73 + 37.33}{15 + 15 - 2} * \left(\frac{2}{15}\right)} = \sqrt{\frac{105.06}{28} * \left(\frac{2}{15}\right)} = \sqrt{0.50286} = 0.7091$$

$$\text{A continuación de calcula la razón T} = \frac{X1 - X2}{0.7091} = \frac{4.13}{0.7091} = 5.82$$

El valor 0.7091 es la diferencia esperada entre dos grupos si se extraen aleatoriamente de una población común con la varianza calculada para X1 y X2. El valor de la razón T indica cuantas veces mayor es la diferencia de lo que se esperaría si la hipótesis cero fuera cierta. La diferencia observada es 5.82 veces mayor que la esperada según la hipótesis de nulidad. ¿Es lo suficientemente grande para rechazar la hipótesis cero con un nivel de confianza de 0.05? Para contestar necesitamos consultar una tabla t con los grados de libertad indicados (df=N-2=28). En este caso, el valor de 5.82 sí

es mayor al valor crítico, y por consiguiente, rechazamos la hipótesis cero. Concluimos que las diferencias de los promedios sí son significativas. Es decir, la diferencia de los dos grupos no se debe al azar, sino a algún factor externo que hace que sus distribuciones no sean de una misma población.

Sería un desperdicio de tiempo hacer que los estudiantes de postgrado calculen siempre sus estadísticas a mano como en el ejemplo anterior. Si las computadoras pueden hacer el trabajo, ¿para que desperdiciar el potencial intelectual de los jóvenes con una tarea tan de rutina? Lo que se requiere de los estudiantes es que piensen. Esto implica que diseñen experimentos, hagan las pruebas, y que interpreten los resultados y presenten sus conclusiones a una audiencia científica. Los cálculos aburridos los puede hacer la computadora. Sin embargo, es necesario saber como funcionan las fórmulas. El presente libro tiene el título de “Fundamentos”, y es por ello que incluimos muchas de estas nociones básicas para el mejoramiento genético. Ahora que hemos aprendido como lo hacían los estadistas de antaño, y cuando ya entendamos como lo hacen las computadoras modernas, los estudiantes de ahora —que serán científicos del futuro— nos podemos enfocar a la interpretación de los resultados.

7.3.2. Ejemplo de análisis usando ANOVA

Ahora haremos un análisis de varianza (*analysis of variance*, ANOVA) con los mismos datos de la sección anterior. Al igual que la prueba T, la ANOVA se utiliza para determinar una razón de las diferencias observadas —nivel/margen de error— para comprobar hipótesis. Esta razón denominada como **razón F**, determina la varianza de las medias de los grupos como una medida de las diferencias observadas entre ellos. La **razón T** sólo se utiliza para comprobar la diferencia entre dos medias. Es por ello que tuvimos que hacer varias pruebas T dentro de una matriz para comparar todos los tratamientos.

La ANOVA nos permite verificar la diferencia entre dos o más medias. Lo particular de ANOVA es que la misma considera la varianza total de todos los sujetos en una muestra incluyendo la varianza entre los grupos y la varianza dentro de los grupos. Se puede usar para analizar experimentos más complejos que incluyen combinaciones de tratamientos. Por lo tanto, es una técnica más amplia y de mayor uso que la razón T.

Una prueba T es equivalente a una ANOVA de un solo factor, mientras que para comparaciones entre dos o más factores ya no se debe usar la prueba T, sino que es mejor usar una ANOVA.

Grados de libertad: Con la expresión de grados de libertad se designa un número de observaciones que pueden variar alrededor de un parámetro constante. Por ejemplo: si se le pide que escoja cinco números y usted elige los que quiera, entonces tiene cinco grados de libertad. Sin embargo, si se le pide que escoja cinco números cuya media de 20 (la suma de 100), entonces puede

escoger cuatro números libremente, pero para el quinto tendrá que restringirse a un número específico para poder llegar a 100. Solo así obtendrá un promedio de 20. Si suma $20 + 30 + 29 + 11 \dots$ el quinto número tiene que ser 10. Por lo tanto tiene sólo cuatro grados de libertad. Usualmente, para determinar los grados de libertad (df), se resta uno del total de la muestra ($df = N - 1$). Cuando se hace un análisis de correlación entre dos variables (regresión lineal) entonces se le restan dos al número total de observaciones ($df = N - 2$).

Suma de los cuadrados: En ANOVA se calcula la suma de las diferencias cuadradas. ¿Porque es tan importante esta suma de cuadrados? La suma de los cuadrados dividido entre los grados de libertad (df) equivale a la varianza (var). Por su parte, la raíz cuadrada de la varianza equivale a la desviación estándar (SD). El error estándar (SE), se calcula a partir del SD dividido entre la raíz cuadrada del número de observaciones (n). Como se puede ver, la suma de los cuadrados está directamente relacionado con un gran número de parámetros relevantes para la estadística.

Cuando se habla de la suma de los cuadrados, la diferencia se refiere a la diferencia entre el valor x y el valor promedio de todo el ensayo. Es decir, no se refiere a los valores promedio dentro de los grupos. En el ejemplo de los datos que hemos usado, el promedio global es 9.975, mientras que el promedio de los grupos individuales es: 6.7, 10.9, 11.6, 10.7 para los tratamientos XA, XB, YA, YB respectivamente.

Para hacer una ANOVA, es necesario organizar los datos de acuerdo al programa que se va a usar para el análisis. La diferencia entre el formato requerido para Excel y el formato para el programa R radica en la forma en que se acomodan los grupos y los tratamientos. en Excel se especifican dos columnas de resultados según la primer variable (Grupo de fertilizante X o Y), mientras que la segunda variable de tratamiento se especifica hacia abajo (grupo de químico A o B).

Formato de datos para Excel

	Grupo X	Grupo Y
A	75	58
	68	56
	71	61
B	75	60
	66	62
	70	60

Formato de datos para R

Variable1	Variable2	Resultado
X	A	75
X	A	68
X	A	71
X	B	75
X	B	66
X	B	70
Y	A	58
Y	A	56
Y	A	61
Y	B	60
Y	B	62
Y	B	60

En cambio, en R, los tratamientos se especifican con columnas extras que indican los factores (variables dependientes por tipo de químico o fertilizante) y los niveles de tratamiento por cada factor. El resultado de rendimiento se pone en una sola columna (Resultado). La ventaja del formato de R es que permite agregar más columnas de resultados hacia la derecha, mientras que en Excel se tiene que hacer una tabla aparte para cada resultado (parámetro) que se quiera analizar.

Tomando todo esto en cuenta, usaremos nuestros datos como ejemplo. En Excel debemos acomodar los tratamientos y los datos así:

Quim	RendX	RendY
A	6	13
A	7	12
A	5	12
A	4	12
A	6	11
A	7	11
A	8	11
A	6	10
A	9	10
A	9	14
B	9	8
B	12	9
B	12	8
B	12	11
B	11	12
B	13	10
B	11	12
B	10	14
B	10	12
B	9	11

Para el análisis ANOVA en Excel, debemos seleccionar todo el rango de celdas -que se muestran arriba- como rango de entrada. En el menú herramientas de Excel seleccionar “*análisis de datos*” y después escoger: “*análisis de varianza de dos factores con varias muestras por grupo*”. En la ventana de dialogo de especificar el valor de filas por muestra (el número de repeticiones, en nuestro caso son 10 por tratamiento).

Para hacer el análisis ANOVA con el programa R, es mejor acomodar los datos de esta forma:

Fert	Quim	Rend
X	A	6
X	A	7
X	A	5

X	A	4
X	A	6
X	A	7
X	A	8
X	A	6
X	A	9
X	A	9
X	B	9
X	B	12
X	B	12
X	B	12
X	B	11
X	B	13
X	B	11
X	B	10
X	B	10
X	B	9
Y	A	13
Y	A	12
Y	A	12
Y	A	12
Y	A	11
Y	A	11
Y	A	11
Y	A	10
Y	A	10
Y	A	14
Y	B	8
Y	B	9
Y	B	8
Y	B	11
Y	B	12
Y	B	10
Y	B	12
Y	B	14
Y	B	12
Y	B	11

Para hacer el análisis ANOVA en R, copiar las celdas arriba en el *clipboard* (portapapeles) y escribir en la consola de R los siguientes 3 comandos en azul:

```
data1=read.table(file="clipboard", header=T, sep="\t")
attach(data1)
summary(aov(Rend~Quim*Fert))
```

El resultado que obtendremos será el siguiente:

```

              Df Sum Sq Mean Sq F value    Pr(>F)
Quim          1  27.225   27.225   10.951  0.002132 **
Fert          1  55.225   55.225   22.213  3.598e-05 ***
***
Quim:Fert      1  65.025   65.025   26.155  1.059e-05 ***
Residuals    36  89.500    2.486
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Ahora trataremos de interpretar el resultado. Primero comparemos las tablas ANOVA que nos dio Excel o R. El resultado ANOVA que obtendremos en Excel se vera así:

Análisis de varianza de dos factores con varias muestras por grupo

RESUMEN	RendX	RendY	Total
<i>A</i>			
Cuenta	10	10	20
Suma	67	116	183
Promedio	6.7	11.6	9.15
Varianza	2.68	1.60	8.34
<i>B</i>			
Cuenta	10	10	20
Suma	109	107	216
Promedio	10.9	10.7	10.8
Varianza	1.88	3.79	2.69
<i>Total</i>			
Cuenta	20	20	
Suma	176	223	
Promedio	8.8	11.15	
Varianza	6.8	2.77	

ANÁLISIS DE VARIANZA

Origen de las variaciones	Suma de cuadrados	Grados de libertad	Promedio de los cuadrados	F	Probabilidad	Valor crítico para F
Químico	27.23	1	27.23	10.95	0.002132	4.11
Fertilizante	55.23	1	55.23	22.21	0.000036	4.11
Interacción	65.02	1	65.02	26.16	0.000011	4.11
Dentro del gru.	89.50	36	2.49			
Total	236.98	39				

Vallamos paso por paso. Las primeras dos tablas se refieren a los datos de los tratamientos A o B por separado. En las columnas se muestran los tratamientos X o Y. En esas tablas se muestran algunos valores que ya habíamos calculado antes, como el número de observaciones por tratamiento, la suma, el promedio y la varianza, que corresponde al cuadrado de la desviación estándar.

El verdadero resultado ANOVA se muestra en la última tabla. Es casi el mismo resultado que nos arrojó R. En la primera columna se muestran los nombres de los tratamientos como origen de la varianza. La segunda columna muestra la suma de los cuadrados, separada por tratamiento, así

como la suma total de las diferencias cuadradas del experimento. En la segunda columna se puede ver que el tratamiento con fertilizante genera más suma de cuadrados que el tratamiento con el químico. La tercera columna muestra los grados de libertad para cada tratamiento. Más adelante explicaremos con más detalle la suma de los cuadrados y los grados de libertad.

La cuarta columna muestra los promedios de las sumas de los cuadrados. Equivale a la suma de los cuadrados dividido entre los grados de libertad (df). La quinta columna muestra la razón F para cada variable del tratamiento (Químico, Fertilizante e interacción entre los dos). La razón F se calcula al dividir el valor de los tratamientos (promedio de la suma de los cuadrados) entre el valor del residual (varianza dentro de los grupos=2.49)

$$27.23/2.49 = 10.95$$

$$55.23/2.49 = 22.21$$

$$65.02/2.49 = 26.16$$

En otras palabras, la razón F se refiere a la variación de los datos —por efecto de los tratamientos— que tantas veces más grande es en comparación con la varianza dentro de los grupos. La varianza entre los grupos es la varianza residual no explicada por los tratamientos. Entre mayor sea el valor F, menor será la probabilidad de que esos efectos se deban solo al azar (error tipo I). La sexta columna muestra precisamente ese valor P (probabilidad de un error tipo alfa). Si no tenemos Excel, en las tablas de estadística podemos consultar la probabilidad P, para determinado valor F, tomando en cuenta los grados de libertad (df). En nuestro caso, todos los tratamientos, así como la interacción entre ellos fue significativa a un nivel $P < 0.001$.

Cuando la hipótesis de nulidad se rechaza una vez llevado a cabo el análisis de varianza, lo único que puede afirmarse es que los resultados obtenidos de los grupos difieren y que las diferencias son mayores de lo que cabría suponer en función de la mera casualidad. Nos dice que el efecto del fertilizante o del químico fue significativo. El tercer renglón se refiere a la interacción entre los dos tratamientos. El resultado nos dice que la interacción fertilizante: químico es significativamente diferente de lo que se podría esperar de un efecto aditivo simple (suma de los efectos de cada tratamiento).

Una razón F alta no quiere decir que todos los grupos dentro de los tratamientos difieran significativamente de los demás. En nuestro caso solo tenemos dos niveles por tipo de tratamiento (fertilizante 0kg y fertilizante 2kg). En nuestro caso el resultado ANOVA nos dice que esos dos tratamientos sí difieren entre si.

Sin embargo, podríamos hacer un experimento con varios niveles de tratamiento, por ejemplo, cinco niveles de fertilizante: 0kg, 1kg, 2kg, 3kg. En ese caso, el resultado de ANOVA, aunque sea significativo, no nos va a decir cual nivel difiere de otro. Una razón F significativa puede ser el resultado que solo un subgrupo difiera de los otros subgrupos. Por ejemplo, que el tratamiento 0kg fuera significativamente distinto de los tratamientos con 1kg y 2kg. Pero un valor P significativo de ANOVA no dice si el tratamiento 2kg y el de 3kg son significativamente diferentes. Para contestar esto, se deben hacer pruebas de comparación múltiple adicionales. Una de ellas es la prueba de Tukey. En la sección X se explica como hacer esas pruebas adicionales usando el programa R. Si no sabe usar el programa R, consulte la sección X de este libro. En resumen se puede decir que primero se debe crear un objeto `aov( )` en R, para después hacer las pruebas de comparación con TukeyHSD.

Tarea 7-1.

De la siguiente lista de datos calcule todos los parámetros estadísticos: n, suma, promedio, suma de los cuadrados, varianza, SD, SE y df. Hágalo con MS Excel, o con el programa spreadsheet de OpenOffice.

X1	X2
18	13
17	12
16	12
16	12
16	11
15	11
15	11
15	10
15	10
14	10
14	10
13	9
12	9
11	8
10	7

7.3.3. Ejemplo: varianza, heredabilidad y repetibilidad

Una de las tareas más importantes de los mejoradores es distinguir entre los efectos debidos al genotipo y los debidos al ambiente y el azar. Esto es importante en todos los ensayos agronómicos, sobre todo los de mejoramiento, tanto los viveros de líneas como los ensayos de variedades. Las diferencias genotípicas entre líneas en un vivero o ensayo, son la señal

que el mejorador está intentando detectar. Sin embargo, la señal no es pura. Hay gradientes de campo que pueden causar grandes variaciones. Los efectos de las unidades experimentales (parcelas, *plots*) están llenas de ruido. La heredabilidad (repetibilidad) es la proporción de señal:ruido (*signal to noise ratio*). A continuación vamos a repasar algunas formulas estadísticas para discriminar la señal del ruido. También haremos un ejercicio sobre el cálculo de los componentes de varianza y repetibilidad.

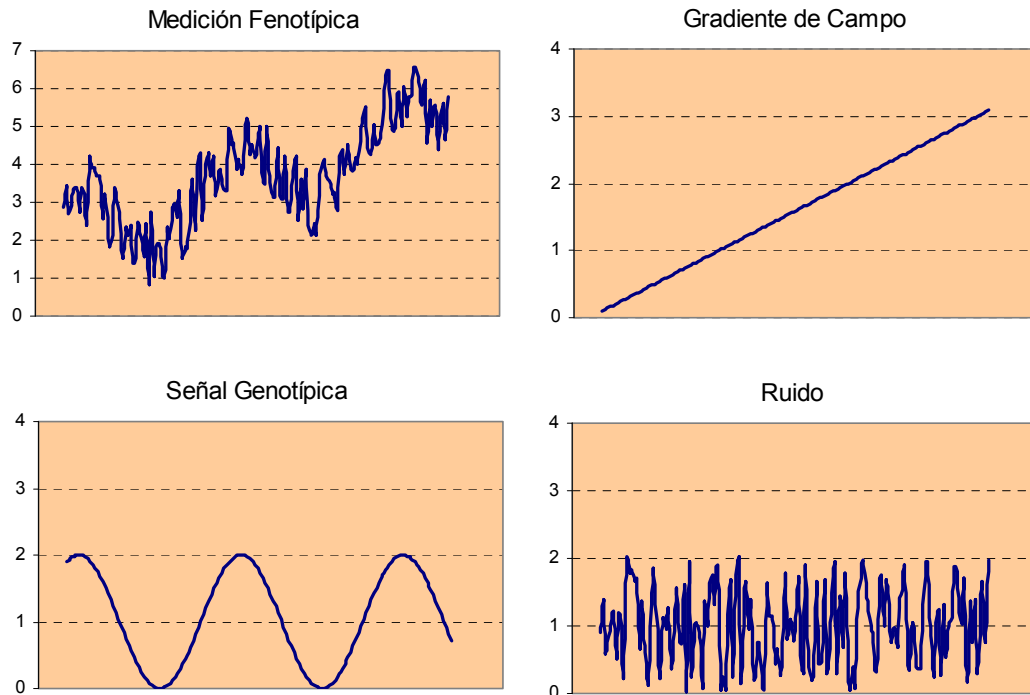


Figura 7-8. Ejemplo de mediciones fenotípicas y su descomposición en sus componentes de señal, gradiente de campo y ruido residual. Esta figura nos ayuda a imaginarnos mejor la magnitud del ruido en comparación con la señal. La descomposición en sus diversos componentes se hace por medio de un análisis de varianza.

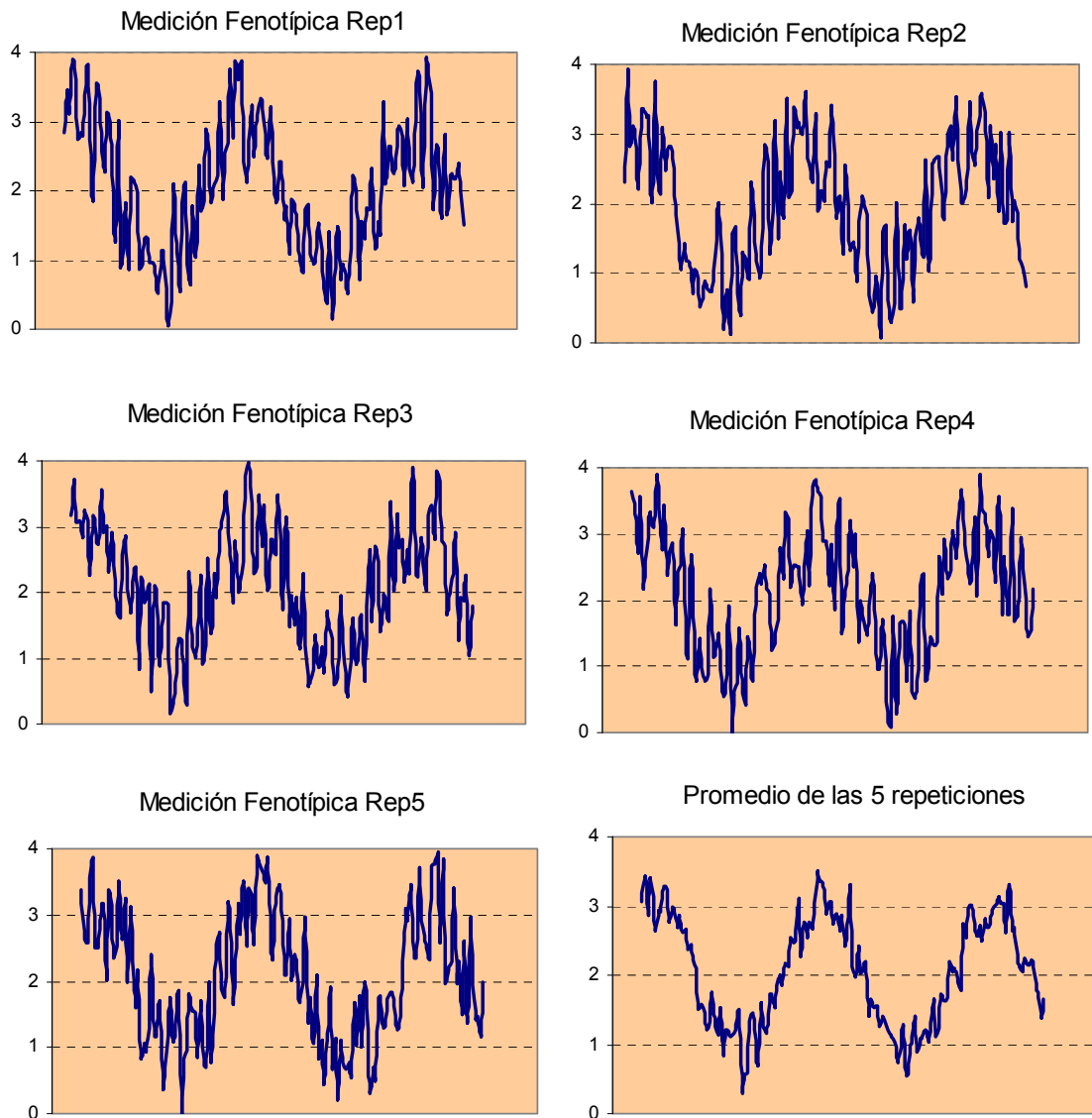


Figura 7-9. Mediciones fenotípicas con varias repeticiones. Entre más repeticiones se hacen, el ruido por azar cada mes tiene un menor efecto. El promedio de las 5 repeticiones tiene menos ruido que las repeticiones individuales, y por consiguiente, la señal genotípica es más clara y confiable.

Los objetivos de esta sección son:

- Repasar los modelos lineales para mediciones en las parcelas en ensayos de variedad y viveros, así como los análisis estadísticos derivados de estas.
- Entender el propósito de la replicación en programas de mejoramiento.
- Repasar el concepto de repetibilidad experimental (Heredabilidad H).

- Modelar la relación entre replicas, error estándar del promedio de un cultivar (SEM), la diferencia mínima significativa (LSD) entre los promedios de dos cultivares y la repetibilidad.

El modelo lineal para mediciones en parcelas

Los efectos genotípicos, ambientales y el ruido son parte del **valor fenotípico**. Estos valores se obtienen de las mediciones en las parcelas. Pueden ser valores como altura de planta, número de mazorcas o rendimiento de grano. Los efectos del genotipo y del ambiente se incorporan en los **modelos lineales**. Estos modelos son los mas utilizados para analizar resultados de ensayos agronómicos. Primero se requiere estar familiarizado con estos modelos, para después discutir mas a fondo sobre como incrementar la precisión de los ensayos, maximizar la respuesta de selección y utilizar eficientemente los recursos limitados de los mejoradores. Así que ahora, vamos a examinar el efecto de la variabilidad del campo, así como la repetibilidad en la precisión con la que los efectos de los genotipos son estimados.

Las mediciones de la parcela son analizadas utilizando un modelo lineal que permite la separación de los efectos genotípicos, del efecto ambiental y el error debido al azar. Una parte del ruido se debe a las diferencias en la calidad el suelo, disponibilidad de agua y nutrientes en las parcelas, y también a la variabilidad inherente de los organismos vivos. En un **diseño completamente al azar** (*completely randomized desing*, CRD) o en un diseño sin bloques, la medición de una parcela (P) en un vivero o ensayo de variedad, puede ser descompuesta en los siguientes componentes:

$$P_{ij} = \mu + G_i + e_j$$

Ecuación 7-1

Donde:

P_{ii} = valor fenotípico de una parcela que se obtuvo por una medición cuantitativa

μ = promedio de todas las parcelas

G_i = efecto en el enésimo genotipo, expresado como una desviación del promedio de todas las parcelas

e_j = el efecto “residual” de la enésima parcela, expresado como la desviación del promedio de todas las parcelas

En un vivero sin réplicas, las mediciones de parcelas sencillas se mezclan completamente, los efectos genotípicos y los de la parcela. En este caso, no se puede separar entre los diferentes efectos.

En ensayos con réplicas, el promedio de la variedad j sobre la parcela ($\sum Y_{ij}/r$), tiene el valor esperado $\mu + G_i$. En la práctica el promedio de una variedad estimada de un ensayo es:

$$\bar{Y}_{i.} = \mu + G_i + \Sigma e_j/r$$

Ecuación 7-2

Dónde:

Σe_j = es la suma de los errores de ese genotipo en las diferentes repeticiones.

r = es el número de réplicas.

Como todos los agrónomos saben, algunas áreas en el campo son más productivas que otras. Si asignamos un genotipo **al azar**, en varias parcelas diferentes (Por ejemplo, si tenemos varias réplicas), los efectos positivos y negativos de la parcela tienden a cancelarse entre ellos y Σe_j tiende a cero. Sin embargo, cuando tenemos pocas réplicas ($r < 3$), los residuos ($\Sigma e_j/r$) pueden hacer que el promedio que estimamos de las parcelas, difiera de manera importante del verdadero promedio del genotipo, $\mu + G_{i.}$

La razón por la que hacemos réplicas de los ensayos de mejoramiento, es para minimizar el efecto de los errores residuales de la parcela en el estimado del promedio del genotipo. Las réplicas no se hacen principalmente para generar un término de error para probar si las diferencias entre genotipos son significativas (y tampoco es para compensar por parcelas pérdidas).

La varianza del promedio de un genotipo es:

$$\sigma^2_Y = \sigma^2_e / r$$

Ecuación 7-3

El **error estándar de un promedio** (standard error of mean, SEM) es la raíz cuadrada de σ^2_Y .

Los mejoradores seguidos están interesados en la **varianza de la diferencia** entre dos variedades. La diferencia (D) debe ser expresada como sigue:

$$\bar{D} = \bar{Y}_1 - \bar{Y}_2$$

Ecuación 7-4

Para dos variables independientes y no correlacionadas, la varianza de la suma es igual a la suma de las dos varianzas. Nota que si $Y = c X$, donde c es una constante, entonces $\sigma^2_Y = c^2 \sigma^2_X$, por lo tanto,

$$\begin{aligned}\sigma_D^2 &= \sigma_{Y_1}^2 + (-1)^2 \sigma_{Y_2}^2 \\ &= \sigma_e^2 / r + \sigma_e^2 / r \\ &= 2 \sigma_e^2 / r\end{aligned}$$

Ecuación 7-5

El **error estándar de la diferencia entre dos promedios** (*Standard error of a difference between 2 means*, SED) es la raíz cuadrada de σ_D^2 .

La **mínima diferencia significativa** (*least significant difference*, LSD) es el valor crítico de la **prueba de t** (t-test) de la diferencia entre dos promedios. Esta se calcula como sigue:

$$\begin{aligned}\text{LSD} &= t_{\alpha/2, \text{edf}} \times \text{SED} \\ &= t_{\alpha/2, \text{edf}} \times \sqrt{(2 \sigma_e^2 / r)}\end{aligned}$$

Ecuación 7-6

Donde $t_{\alpha/2, \text{edf}}$ es el valor t para un nivel de significancia de $\alpha/2$, y el número de grados libertad en el término de error (edf) del análisis de varianza del ensayo para $\alpha/2$, $t_{\alpha/2, \text{edf}}$ es siempre aproximadamente 2. Como una aproximación,

$$\text{LSD}_{0.05} \approx 3 \text{ SEM}$$

Ecuación 7-7

Es importante que el concepto de LSD quede claro: **El valor $\text{LSD}_{0.05}$ es la diferencia entre estimados de los promedios de dos líneas con el mismo valor genotípico, que se espera que sea excedido por mero azar con un 5% de las repeticiones de un ensayo de variedad.**

Visto de otra manera, suponga que toma dos muestras de semilla de una variedad cultivada en la misma parcela. Cada submuestra las pone en diferente sobre, dándole diferente nombre. Después, toma estas dos muestras y las compara en un ensayo con réplicas en diferentes sectores de la parcela. Por puro azar, la diferencia en el promedio del rendimiento entre las muestras de material idéntico podría exceder el valor $\text{LSD}_{0.05}$ solo el 5% de las ocasiones.

Los valores de SEM, SED y LSD son mediciones importantes de la precisión de un ensayo de variedad, o de la habilidad de este para detectar diferencias.

Varianza genotípica y fenotípica

La varianza genética en un ensayo, es la varianza de los efectos del cultivar, el valor G_i 's en la ecuación 5.1, y se denota como σ^2_G .

La varianza fenotípica en un ensayo de cultivar, es la varianza de los promedios del cultivar a través de las réplicas del ensayo. Se denota como σ^2_P . Ya que los promedios están basados en las mediciones de la parcela, las cuales contienen tanto a G_i 's como a e_j 's, σ^2_P contiene la varianza genética y una porción de la varianza residual.

$$\sigma^2_P = \sigma^2_G + \sigma^2_e/r$$

Ecuación 7-8

Donde σ^2_e es la varianza de parcela residual o de error del ANOVA, y r es el número de réplicas. σ^2_G y σ^2_e son estimadas del ANOVA como sigue:

Fuente de variación	Grados libertad (Df)	Promedio de la suma de cuadrados (Mean square)	Valor correspondiente a
Réplicas	r-1		
Genotipos	g-1	MS_G	$r\sigma^2_G + \sigma^2_e$
Residual	(r-1)(g-1)	MS_E	σ^2_e

Entonces para estimar σ^2_G :

$$\sigma^2_G = (MS_G - MS_E)/r$$

Ecuación 7-9

σ^2_e se estima directamente como la varianza del error residual del experimento.

La repetibilidad en los ensayos

La repetibilidad en un ensayo de variedades es la proporción de la variación dentro de los promedios de una línea, que es debida a la variación en los efectos genotípicos. También se le llama **heredabilidad en sentido amplio** (*broad-sense heritability*, H) y es calculada de la siguiente forma:

$$H = \sigma^2_G / \sigma^2_P = \sigma^2_G / [\sigma^2_G + \sigma^2_e/r]$$

Ecuación 7-10

H es una medida importante de la precisión (fiabilidad) de un experimento. Es un estimado de la correlación esperada entre diferentes repeticiones o corridas del mismo experimento. Es decir, es la correlación que esperaríamos entre promedios estimados de la misma serie de cultivares,

estimado dos veces en diferentes parcelas del mismo campo de cultivo. Los valores que toma H se encuentran entre 0 y 1. Si observamos la ecuación 5.10, nos daremos cuenta de que H no es una constante. Mientras mayor sea el número de réplicas (r), o menor sea el error residual ($\sigma_e^2=0$), H se aproxima más a 1. Entonces puede “comprar” H s mayores para un experimento, invirtiendo en más réplicas.

Note que el valor de H pertenece sólo al ensayo donde fue estimado. Los valores de H estimados para ensayos con menos de 30 variedades son menos confiables que los ensayos de 100 variedades. También tenga en cuenta **que H no nos dice nada acerca de la transmisión mendeliana** de una característica de padres a hijos. H es solamente una medida estadística de la **repetibilidad** de una característica en un experimento en particular. Esto a menudo **causa confusión**, debido a que en biología molecular, la heredabilidad tiene una connotación de herencia y de transmisión de ácidos nucleicos (genes). El termino heredabilidad se acuñó cuando todavía no se sabía nada del DNA, de la época clásica de Hallauer en los años 30 cuando se fundaron los principios de la genética cuantitativa. La mayoría de los términos que se usan en el mejoramiento están basados en estadística y no tanto en el DNA. Hoy en día convendría cambiar el término de heredabilidad por el de repetibilidad para no crear confusión entre los biólogos moleculares y los mejoradores. Sin embargo, ahora que sabemos los alcances de esos términos, eso nos ayudara a ampliar los canales de comunicación entre científicos de diferentes disciplinas.

Modelar el efecto de la replicación en la precisión y repetibilidad de los ensayos de variedades

Todos sabemos que hacer muchas réplicas es caro y laborioso. Podemos usar la Ecuación 7-6 y Ecuación 7-10 para modelar el efecto al cambiar el número de réplicas en un ensayo. Por lo tanto, estas ecuaciones son útiles para aquellos mejoradores que estén planeando o estructurando sus ensayos. Es importante saber si los ensayos están sub o sobre-replicados, para saber si los recursos económicos están siendo bien invertidos para lograr un avance genético deseado.

Ejercicio de modelación:

En la siguiente tabla se presenta el análisis de varianza para el rendimiento de grano de un ensayo de variedad con arroz de temporal de tierras bajas, llevado a cabo en Champassak Laos, durante la temporada de lluvias del 2004.

Fuente de variación	Grados libertad (Df)	Promedio de la suma de cuadrados (Mean square)	Valor correspondiente a
Réplicas	3	688426	
Genotipos	21	759729	$r\sigma_G^2 + \sigma_e^2$
Residual	66	159692	σ_e^2

Ahora, con esos datos:

- Estime σ_G^2 y σ_e^2
- Calcule el valor predicho de SEM, SED, $LSD_{0.05}$ y H para ensayos con 1-8 réplicas. El valor apropiado de t es 1.997

Número de réplicas	σ_G^2	σ_e^2	SEM	SED	$LSD_{0.05}$	Reducción en LSD debido a la adición de réplicas	H
1							
2							
3							
4							
5							
6							

- Si dos ensayos de variedad con 3 réplicas como el descrito anteriormente fueran llevados a cabo uno a lado del otro en el mismo campo y en la misma temporada, ¿cuál sería la correlación esperada en los promedios de las variedades entre los dos ensayos?
- ¿Cuáles son las dos maneras posibles de incrementar la precisión y repetibilidad de un ensayo de variedad?

Resumen

- En ensayos de campo y viveros, los efectos del genotipo y de la parcela siempre se confunden. Los efectos de la parcela son “ruido” que enturbian el estimado del valor genotípico en los experimentos de mejoramiento.
- En los programas de mejoramiento, el propósito de la replicación, es separar los efectos genotípicos de los debidos a la parcela, reduciendo la confusión, mejorando la precisión de los ensayos e incrementando nuestra habilidad de identificar genotipos superiores.

- Las varianzas genotípicas o de error estimadas de los análisis de varianza pueden ser usadas para estimar el LSD y la repetibilidad o *heredabilidad en sentido amplio*, H
- Incrementar el número de replicas reducirá el LSD e incrementará H así como la precisión y repetibilidad del ensayo.
- Los beneficios obtenidos en la precisión disminuyen rápidamente con cada réplica adicional. Raramente es económicamente efectivo utilizar más de 4 réplicas.
- Las ecuaciones de LSD y H pueden ser utilizadas para modelar el efecto de incrementar o disminuir el número de réplicas en la precisión y repetibilidad del ensayo.

7.3.4. Ejemplo de gráfica

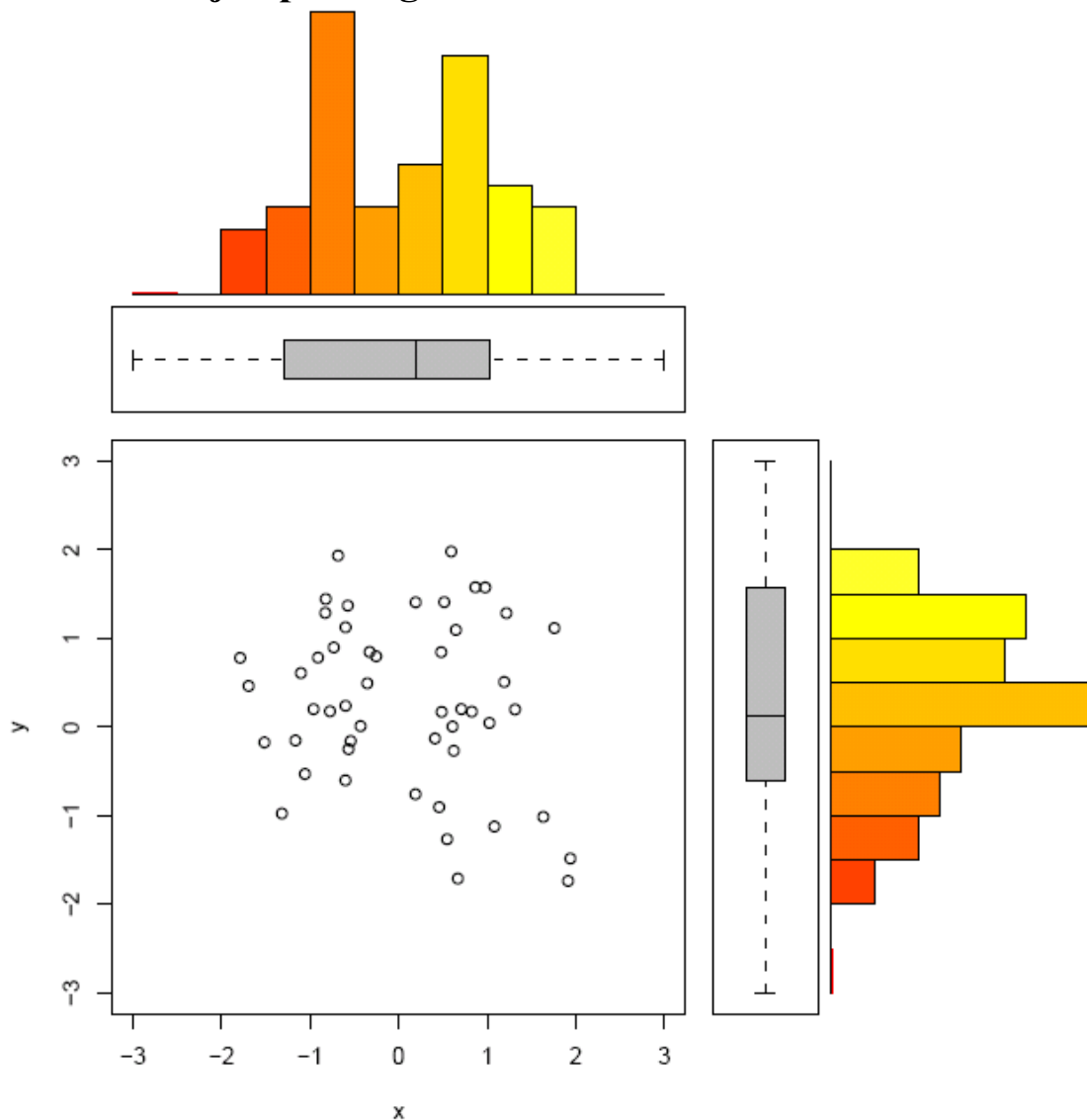


Figura 7-10. Scatterplot xy con histogramas y boxplots marginales.

7.3.5. Ejemplo de análisis de bloques incompletos

Tarea 7-2

Refuerce sus conocimientos de estadística. Descargue el programa FieldBook del CGIAR, instalelo y haga un análisis de un ensayo de campo con los bloques de un diseño alfa-lattice.

7.3.6. Ejemplo de análisis espacial

Tarea 7-3

Haga un análisis espacial usando las coordenadas de campo de columnas y filas.

7.3.7. Ejemplo de análisis con modelos lineales

Tarea 7-4

Importe los datos a R y haga un análisis con las funciones aov y lm.

7.4. Software para estadística

7.4.1. Uso de Excel

Excel es uno de los programas que más usan los fitomejoradores y los bioquímicos analíticos. Las hojas de Excel se pueden usar tanto para almacenar y organizar datos, como para hacer cálculos, gráficas y análisis estadísticos. En el capítulo 6.4.1 (ver página 204) se menciona el paquete de Fieldbook basado en macros de Excel que los mejoradores de maíz usan para manejar sus ensayos de rendimiento. Excel tiene una amplia gama de funciones integradas para estadística. Entre las más simples se pueden mencionar el promedio, la media, moda, máxima, mínima, desviación estándar y número de valores de un determinado rango. También hay funciones para calcular la pendiente (slope), la intersección (intercept) con el eje-y y el coeficiente de correlación a partir de una serie de datos xy. Algo muy conveniente del programa Excel son las funciones de búsqueda vertical, horizontal y las tablas dinámicas (*pivot tables*). Entre las funciones de análisis se pueden mencionar los histogramas, las tablas de correlación y el análisis de varianza. Una de las funciones más avanzadas de Excel es el solver. Se puede usar para resolver problemas lineales y no lineales. También cuenta con algunas funciones para hacer análisis estadísticos con modelos lineales simples. Según la versión de Excel que se tenga instalada, las funciones pueden estar en inglés o en español. A continuación se muestran algunos ejemplos de funciones sencillas.

Tabla 7-1. Formulas y funciones de Excel

Función	Versión inglesa	Versión española	Formula

Promedio	=AVERAGE(x)	=PROMEDIO(x)	$\frac{\sum x}{n}$
Mediana	=MEDIAN(x)	=MEDIANA(x)	
Moda	=MODE(x)	=MODA(x)	
Mínima	=MIN(x)	=MIN(x)	
Máxima	=MAX(x)	=MAX(x)	
Numero de mediciones (n)	=COUNT(x)	=CONTAR(x)	n
Grados libertad (DF)	=COUNT(x)-1	=CONTAR(x)-1	$n - 1$
Desviación estándar (SD)	=STD(x)	=DESVEST(x)	$\sqrt{\frac{\sum (x - \bar{x})^2}{(n - 1)}}$
Error Estándar (SE)	=STD(x)/SQRT(n)	=DESVEST(x)/RAIZ(n)	$SE = \frac{SD}{\sqrt{n}}$
Pendiente (valor m)	=SLOPE(y, x)	=PENDIENTE(y, x)	$m = \frac{n \sum (xy) - \sum x \sum y}{n \sum (x^2) - (\sum x)^2}$
Intersección (valor t)	=INTERCEPT(y, x)	=INTERSECCION.EJE(y, x)	
Coefficiente de correlación R	=CORREL(y, x)	=PEARSON(y, x)	$r = \frac{n \sum (xy) - \sum x \sum y}{\sqrt{[n \sum (x^2) - (\sum x)^2][n \sum (y^2) - (\sum y)^2]}}$
Coefficiente de correlación R ²	=RSQ(y, x)	=COEFICIENTE.R2(y, x)	$r^2 = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$

Las fórmulas de la tabla anterior parecen un poco complicadas. Sin embargo, al inspeccionarlas con calma, nos podemos dar cuenta que sus componentes se basan en simples manipulaciones algebraicas de los datos primarios. Se basan en algunas sumas simples y cuadradas. La Figura 7-11 muestra un ejemplo de los cálculos que se hacen para una regresión lineal.

A11 = =COUNT(B3:B8)										
	A	B	C	D	E	F	G	H	I	J
1										
2		x	y	xy	x ²	y ²				
3		1.0	2.6	2.6	1.0	6.8				
4		2.3	2.8	6.44	5.3	7.8				
5		3.1	3.1	9.61	9.6	9.6				
6		4.8	4.7	22.56	23.0	22.09				
7		5.6	5.1	28.56	31.4	26.0				
8		6.3	5.3	33.39	39.7	28.1				
9										
10	n	Σ x	Σ y	Σ (xy)	Σ (x ²)	Σ (y ²)				
11	6	23.1	23.6	103.16	110.0	100.4				
12										
13		(Σ x) ²	(Σ y) ²							
14		533.61	556.96							
15										
16	slope, m =	0.5842					=(A11*D11-B11*C11)/(A11*E11-B14)			
17	y-int, b =	1.6842					=(C11-C16*B11)/A11			
18	r =	0.9741					=(A11*D11-B11*C11)/SQRT((A11*E11-B14)*(A11*F11-C14))			

Figura 7-11. Ejemplo de cálculos de regresión lineal en Excel. A partir de los datos x,y se calculan las sumas de las multiplicaciones y de los cuadrados, para después calcular los valores m, b y R.

El valor R^2 por si solo no es suficiente para decir si una correlación es significativa o no. Por ejemplo, un valor de $R^2=0.9$ no es significativo si solo se tienen 3 puntos. Mientras que un valor de $R^2=0.5$ es altamente significativo si se tienen cientos de datos. Es necesario tomar en cuenta los grados de libertad para una determinada comparación. En el caso de las correlaciones entre dos variables, los grados de libertad son $(n-2)$. La Figura 7-12 muestra la forma de calcular el valor P para un grupo de datos xy.

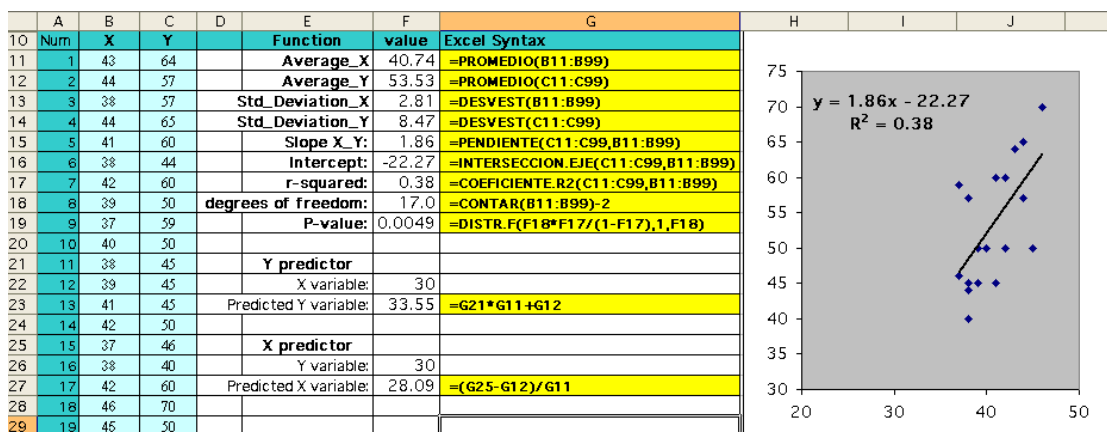


Figura 7-12. Ejemplo de funciones en Excel. Para calcular el valor P se usa el valor R^2 corregido con los grados de libertad usando la función de

distribución F. Un valor de $R^2=0.38$ con una lista de 19 observaciones si es significativo ya que el valor $P=0.0049$.

7.4.2. R para estadística

El R es un programa para analizar datos. El programa puede hacer diversos tipos de análisis estadísticos, así como producir muchos tipos de gráficas. La ventaja de R, en comparación de otros como SAS y SPSS, es que R es un programa gratuito y de código abierto.

R is a software package especially suitable for matrix data analysis and graphical representation. R can be used as a statistical tool but also as a programming language itself, making it very flexible and highly customizable. Excellent graphical tools make R an ideal environment for exploratory data analysis and for preparing publication ready graphs (exportable as .ps or .jpg files). All work is done in command style text functions and therefore it is different from other windows style programs (like SPSS) that use menus with select and click options for predefined statistical procedures. It takes considerable time to learn to use R, but once you have passed the first burden, it is quite convenient to handle. As a programming language it is not for beginners, but rather for advanced users for whom the statistical functions of Microsoft Excel are no longer enough (for example if you want to do Principal Component Analysis). In comparison to SAS and SPSS which are very expensive commercial programs for doing statistics, R is **free software** distributed under the GNU and GPL license terms.

Descargue R de la página oficial:

<http://cran.r-project.org/>

Revise y descargue estos otros programas complementarios:

Editor de texto para editar scripts de R: <http://www.sciviews.org/Tinn-R/>

Análisis de datos de microarreglos: <http://bioconductor.org/>

Después de instalar R en su computadora de manera local, el programa inicia con la ventana de la consola:

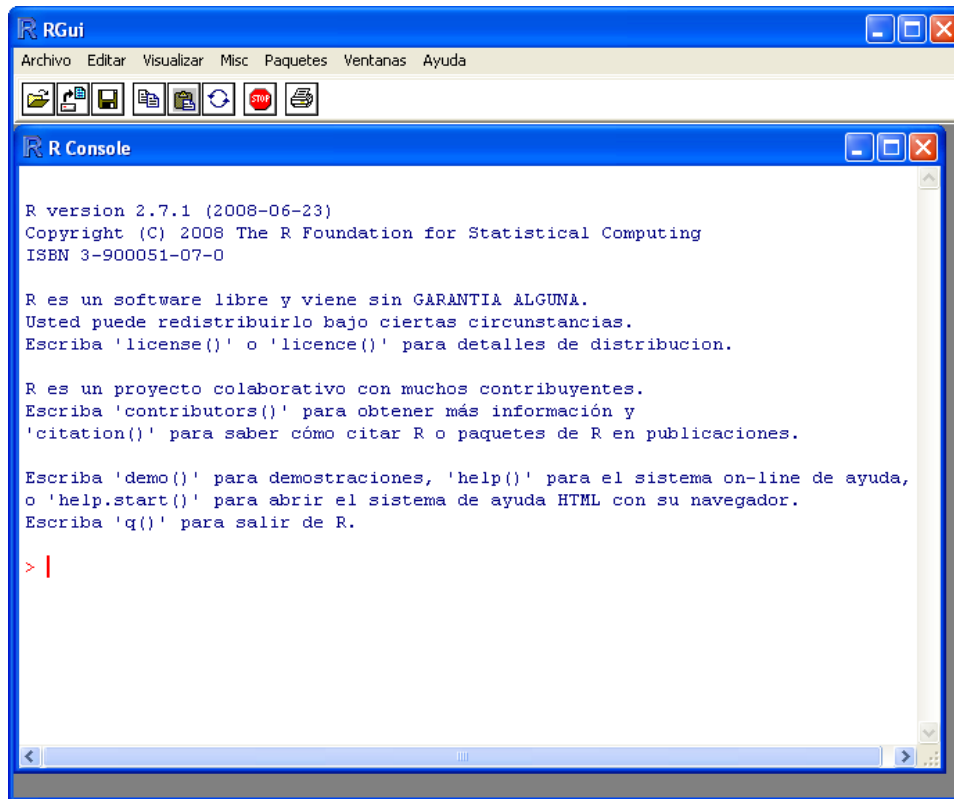


Figura 7-13. Ventana inicial de la consola de R

Tabla 7-2: Consola de R

Introducción a R (Uso básico)

El programa R utiliza **objetos** y **funciones**. Los objetos contienen datos, mientras las funciones utilizan los datos dentro de los objetos, para llevar a cabo cálculos y obtener soluciones. Si quieres ver todas las funciones de R, escribe:

`help()` # Abrirá una ventana con información general (enlista todas las funciones)

Para obtener **ayuda específica sobre una función dada**, escribe su nombre dentro del paréntesis:

`help(var)` # Ayuda específica sobre la función varianza (variance)
`help(sd)` # Ayuda específica sobre la función desviación estándar (standard deviation)
`?sd` # Alternativa para abrir ayuda sobre la función desviación estándar

En muchas de las funciones de R se necesita abrir y cerrar paréntesis para introducir variables o atributos de la función respectiva. Aunque también en algunas ocasiones, los paréntesis se dejan vacíos.

R distingue mayúsculas de minúsculas en todo lo que se escribe. Como en los comandos unix, en R puedes **volver a llamar los comandos previamente escritos**, utilizando las teclas de las flechas hacia arriba o hacia abajo.

Definiendo y editando objetos

Para trabajar en R, necesitas **definir objetos**. Un objeto puede ser una variable, un vector o una matriz, dependiendo de la cantidad de datos y su estructura. Escoge un nombre para el objeto, que no sea el nombre de una función. Algunos nombres posibles son a, b, c, x, y, z, data1, data2, matrix1, model1, etc.

La siguiente sección serán en inglés para practicar un poco más esté idioma. Para hacer un posgrado en biotecnología vegetal, es muy importante poder entender instrucciones técnicas en inglés. Las palabras en azul se refieren a los comandos que se usan en la consola de R. Haga copy de esos comandos y péguelos en la consola para que vea el resultado en R, y de esta forma aprenda a usar este programa para hacer análisis estadísticos (Tabla 7-2).

#The # sign denotes comments that are not interpreted by the console. I use them here to explain some features of the functions or commands.

#To work with R you need **first to define variables**, vectors or matrix objects. This is done either manually or by importing data from external files. First I will explain the **manual entry of data** directly from the console. This can be done with the combine function:

```
x = c(1,3,2,7,9) #create a vector x with 5 components
y = c(a=1,a=3,b=2,c=7,c=9,d=2,d=3) #define y with variable names and
values
z=matrix(c(10,20,30,40,50,60),ncol=2) #define matrix with 2 columns.
Values are equally distributed into the two columns. The first column is
filled first and downwards.
x; y; z #show the contents of objects x, y and z
```

#The definition of the variables can be done with =, but also with the symbols of <- or also -> depending on the side in which you put the variable.

```
x = c(1,3,2,7,9) #my preferred option to create a vector x with 5
components
```

```
x <- c(1,3,2,7,9)  #alternative option to create a vector x with 5
                    #components
c(1,3,2,7,9) -> x   #alternative option to create a vector x with 5
                    #components
```

#To show the contents of any object simply type its name

```
x                # list contents of object x
```

```
x=edit(x)        # edit contents of object x in a data.frame window
```

#Now that you have defined an object x, you can do things like:

```
summary(x)       # shows a summary of x (min, max, mean, median,
                 # quartiles)
```

```
sd(x)            # gives the standard deviation of data in x
```

```
sd(x)/sqrt(length(x))  #gives the standard error of data in x
```

#The object x and the results of those functions are stored in the RAM memory of R. When using R no file is written to the harddisk unless you save it specifically. When you quit R program you can save the workspace so that all objects created during the current session are available next time you work with R.

#If you need to do further calculations with the results of a given function, you need to define a new object to store the output. For example:

```
a=length(x)      # stores the results of length(x) in new object a
```

```
ls()             #to list all objects stored in the current workspace
```

#Remember that everything you type in R is case sensitive. As in unix command line, in R you can **recall previously typed commands** using the up and down arrow keys.

#A convenient way to enter matrix data manually into R is by using the command `edit(data.frame())`

```
data1 = edit(data.frame())  # opens a window to define data1 manually
```

Después de que hayas introducido los datos y cerrado la ventana del editor, tus datos se guardaran como el objeto `data1`. Dando un “click” en las marcas de las columnas, puedes definir si los valores serán numéricos o categóricos. Puedes nombrar tus variables, o dejar los nombres predeterminados (`var1`, `var2`, `var3`, etc.). Si cambias los nombres de las variables, utiliza uno corto y fácil de recordar, de tal manera que tengas fácil acceso a él (función `attach`). No utilices espacios o caracteres extraños para el nombre de las variables. Para ver que esta guardado en una matriz determinada, simplemente escribe el nombre del objeto. Aquí yo utilice “`data1`”, pero tú puedes definirlo con otro nombre similar

```
data1          # to list the content of the data1 object in the console
window
data1=edit(data1) # to re-edit the contents of object data1 and saving it
edit(data1)      # It prints editions on screen but it does not save changes in
memory
attach(data1)     #to make the variables defined in data1 accesible to the
console
var1 # to show the contents of var1 (defined inside attached data1)
detach(data1) #to detach the variables defined in data1
data1$var1 # to show the contents of var1 from data1 without the attach
function
```

La primera barrera para usar R es introducir una gran cantidad de datos en un objeto. La entrada manual de datos descrita anteriormente, no es práctica ni conveniente en determinadas ocasiones. Hay mejores opciones para introducir datos a R, desde otros archivos. Se pueden importar datos desde el **portapapeles de Windows**. Esto es apropiado para conjuntos de datos pequeños y medianos. Para conjuntos de datos muy grandes, puedes importar los datos desde **archivos de texto externos**.

Subiendo datos desde el portapapeles

Más que escribir manualmente los datos con la función *edit data frame*, es preferible importar los datos desde Excel o Word, vía el portapapeles de Windows. Para hacer esto, sólo abre tu archivo de Excel o Word, y copia la tabla de datos, incluyendo o no, los nombres de las filas y los títulos de las columnas. Después ve al programa R y escribe el comando:

```
data1=read.table(file="clipboard", sep="\t") #Para importar sin  
nombres de filas y títulos de columnas.
```

Las variables de las columnas serán nombradas como V1, V2, V3, consecutivamente. Si la primera celda está vacía, se podrá reconocer de manera automática los nombres y títulos de las filas. Este comando crea el objeto "data1" con el contenido del portapapeles de Windows. Transfiere los datos desde la memoria del portapapeles hacia la memoria de R. Lo anterior funciona bien para Excel, Word, Power Point y otros programas de Windows. Para mí, esta es la forma más conveniente de importar datos a R, pues no necesito crear archivos de texto, carpetas o definir directorios de trabajo.

```
data1=read.table(file="clipboard") #for importing without row names  
and column headings. Column variables will be named V1, V2, V3,  
consecutively. If the first cell is empty then this allows automatic  
recognition of row names and headings.
```

```
data1=read.table(file="clipboard", header=T) #for forcing import with  
row names and column headings. This option works even if the first cell  
is not empty.
```

```
data1=read.table(file="clipboard", header=T, sep="\t")
```

```
data1=read.table(file="clipboard", header=T, sep="\t",
na.strings="")
```

Breve nota sobre importar datos

Algunos tutoriales te recomiendan exportar los datos de hojas de Excel a archivos de texto, y desde estos archivos importarlas a R. Esto es **totalmente innecesario**, puesto que se puede utilizar directamente el portapapeles para importar los datos. Sin embargo, algunas veces se tiene una gran cantidad de datos que no cabe en el portapapeles. Por ejemplo, cuando se hacen microarreglos de diez mil genes. En estos casos, debes de usar la función *read table*, para importar tus archivos de texto como se describe a continuación.

Subiendo datos desde archivos de texto

El programa R necesita definir una carpeta para subir y guardar archivos. Para cambiar el directorio en el cual quieres trabajar, ve al menú ->file->. Cambia el directorio y selecciona una carpeta del directorio que elegiste. También lo puedes hacer manualmente en la consola:

```
setwd("C:/Axel StatisticsR")    # sets the working directory manually
getwd()                        #displays the path for the working
    directory
dir()                           # lists the files in the current work
    directory
```

Para importar datos desde un archivo de texto (*tab delimited*), tienes varias opciones:

```
data1=read.table("file.txt", header=T)
data1=read.table("file.txt", header=T, sep="\t")
data1=read.table("file.txt", header=T, sep="\t", row.names = 1)
```

La opción `sep="\t"` te permite definir el símbolo de separación entre valores, en la mayoría de los casos, estos están separados por un tabulador. La opción `sep=","` no funciona para formatos *csv* (valores separados por comas). En este caso es preciso usar la opción `sep=","`. La opción *row names*, te permite especificar el nombre de las filas.

El archivo de texto debe estar en el directorio predeterminado de R. Utiliza *setwd* para definir el directorio correcto. Una manera más conveniente de subir archivos, es usando el comando

`file.choose()`. Con esta función no es necesario recordar o escribir el nombre del archivo manualmente. También puedes **navegar la carpeta con ventanas** (browse the folder with windows). Para importar datos desde un archivo de texto delimitado por tabulaciones estándar, utiliza el comando:

```
data1=read.table(file.choose(),header=T) # selecciona el archivo
con una ventana emergente
```

Si tu archivo de datos no tiene el nombre de las variables en la primera columna, entonces utiliza la opción `header=F`. Recuerda que en todo lo que escribes, R reconoce mayúsculas de minúsculas y que puedes llamar los comandos previos para corregir cualquier error, utilizando las teclas de flechas hacia arriba o hacia abajo.

Análisis de objetos

Si tienes definido un objeto `x`, puedes hacer cosas simples como:

```
x                # shows data stored in the object x on screen
sort(x)          # list data increasing order
length(x)        # number of items in x
```

Con objetos de datos guardados en el espacio de trabajo, se pueden realizar muchas funciones individuales como: `sum(x)`, `max(x)`, `min(x)`, `median(x)`, `mean(x)`, pero es mejor utilizar la función `summary`, que integra las funciones anteriores:

```
summary(data1)    # shows a summary of x (min, max, mean, median, quartiles)
sd(data1)         # standard deviation of elements in data1
sd(data1)/sqrt(length(data1)) # gives the standard error of data in data1
var(data1)        # variance value, or covariance matrix of elements in data1.
```

Breve nota de la función `var()`. Si se da un objeto vectorial de una dimensión, el resultado es la varianza total del conjunto de datos de dicho objeto. Si el objeto es una matriz con más de una dimensión, entonces la función `var()` calcula una matriz de covarianza, equivalente a la función `cov()`. En una matriz de covarianza, los elementos en diagonal, son varianzas dentro de la variable dada, y los otros elementos son covarianzas entre diferentes variables. La covarianza es una medida de la manera en que las variables se relacionan, es decir que tan independientes o que tan correlacionadas están. Una matriz de covarianza, se puede escalar a una matriz de correlación con la función `cov2cor()`.

```
plot(data1)       # matrix correlation plots with all elements in data1.
cor(data1, use = "pairwise.complete.obs") # correlation coefficient matrix of all elements in
data1
round(cor(data1, use = "pairwise.complete.obs"),2) # correlation matrix rounded to 2 decimals
round(cor(data.frame(data1$var1, data1$var2, data1$var3)),2) # correlation matrix rounded to 2
decimals with some selected variables only
```

Definiendo nuevos objetos

Las funciones escritas como tal, solo mostrarán los resultados en la pantalla. Si necesitas hacer cálculos con los resultados de una función dada, necesitas usar el comando `object=` para definir un nuevo objeto que guardará los resultados de dicha función en la memoria de trabajo. Por ejemplo:

```
Matrix1=var(data1) # stores the covariance matrix in new object called Matrix1
Matrix2=cor(data1) # store the correlation coefficient matrix of elements in data1
Matrix1; Matrix2    #shows the contents of the two newly created objects. The semicolon
                    allows separating two functions given in the same line
```

Todos los objetos se guardan en la memoria RAM de R. Cuando se utiliza R ningún archivo se escribe en el disco duro, a menos que se guarde específicamente. Cuando cierras el programa R, puedes guardar el espacio de trabajo, de tal manera, que todos los objetos que se crearon en la sesión, estén disponibles la siguiente vez que se trabaje con R.

```
ls()                #to list all objects stored in the current workspace
Matrix1             # to list the content of the Matrix1 object in the
                    console window
edit(data1)         # It lets you print editions on screen but it does not save
                    changes in the object memory.
```

Para editar y guardar los cambios, utiliza el siguiente comando:

```
data1=edit(data1)   # to re-edit the contents of object data1 and
                    saving it
rm(x)               #to remove object x from workspace
```

Trabajando con matrices

Realiza algunas manipulaciones con matrices como:

```
data3=2*data1       # define a new matrix with scalar matrix multiplication by factor 2
data4=data1+data3    # matrix addition
data5=solve(data1)   # define a new matrix using the inverse matrix of data1
data2=t(data1)       # define a new matrix using the transposed matrix from data1

data6=data.frame(data1,data3) # use data.frame for adding elements to objects
```

Extraer un componente de una matriz involucra uno o dos índices
[Fila,columna]

```
data1[2,3]          #element of data1 at the second row, third column
z=data1[2,3]         #define new object with a selected data entry of data1
```

#conditional recoding of values within an object

```
data1= x = rnorm(1000)  #create a set of 1000 random values
x[x > 1.2] = NA        # to recode all values of x > 1.2 to NA
```

```
data1= data.frame(a = rnorm(1000), b= rnorm(1000) , d=rnorm(1000), entry=c(1:10))
data1$a[data1$b > 1] = 5      # to recode all values of b > 1 to 5
```

La función *attach* se utiliza para tener acceso a varias variables dentro del objeto. Utiliza *detach* para eliminar el acceso a esas variables.

```
attach(data1) # allows direct access to the variables defined inside an object
hist(var1)    # plots an histogram of variable var1 inside attached object data1
detach(data1) # detaches variables
hist(data1$var1) # plots an histogram of var1 from data1 without attaching
```

Valores propios y vectores propios (eigenvalues, eigenvectors, respectivamente) de una matriz se manejan con la función *eigen()*:

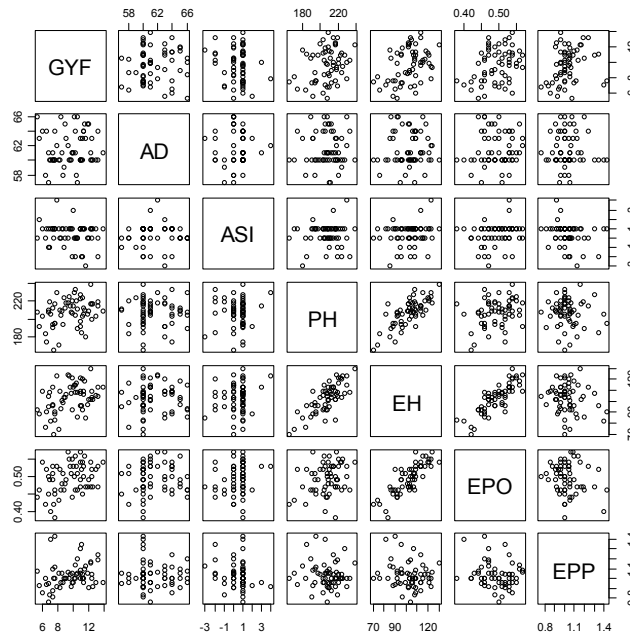
```
m2=matrix(c(10,20,30,40),ncol=2)
eigen(m2)
```

Creando gráficos

Hay muchas maneras sencillas para graficar utilizando los datos en los objetos.

`plot(data1)` #Salida gráfica de los arreglos de gráficas con los elementos en data1.

Esta característica de gráficos es particularmente notable, pues grafica todas las variables en la matriz, contra todas las variables. Con ella, puedes hacer análisis exploratorios y descubrir algunas correlaciones escondidas. Si deseas hacer esto en Excel, te tomará mucho más tiempo.



`boxplot(data1)` #graphical output with boxplots of all elements in data1

`attach(data1)` #to make the variables defined in data1 accesible to the console

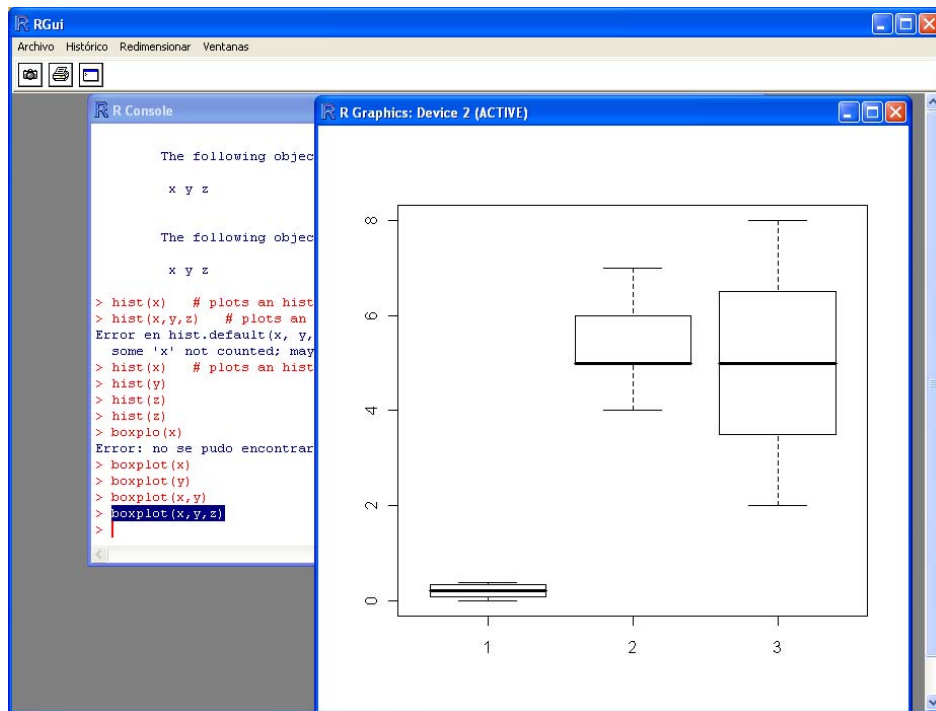
`var1` # to show the contents of var1 (defined inside attached data1)

`boxplot(var1, var2)` #produces a boxplot of only the variables 1, 2

`plot(data.frame(var1, var2, var3))` #plot arrays of some selected elements only.

`detach(data1)` #to detach the variables defined in data1

`data1$var1` # to show the contents of var1 from data1 without the attach function



Si quieres obtener algunas estadísticas involucradas en los diagramas de caja, utiliza los siguientes comandos:

`b=boxplot(data1); b$stats` # gives the value of the lower end of the whisker, the first quartile (25th percentile), second quartile (median=50th percentile), third quartile (75th percentile), and the upper end of the whisker.

`plot()` is a general graphic command with numerous options.

`plot(x)` # plots the data x

`plot(var1, var2)` # plots var1 in dependence of var2

`abline(line(var1,var2))` # add a regression line

`plot(y,z, main="Enter Title Here")` # scatterplot with variables y and z

`fit=lm(y~z)` #A fitted straight line is shown in the plot by executing two more commands

`abline(fit)`

Exportando gráficas

Hacer click derecho en cualquier lugar de la ventana activa de la gráfica, mostrará un menú sensible al contexto, que te permite guardar la gráfica como metafile(EMF) o con formato postscript (PS). Las opciones *Copy as metafile* o *Copy as bitmap* (BMP) coloca la gráfica en el portapapeles. Inmediatamente necesitas pegarla en alguna aplicación como MS Word. Hay mas formatos de gráficos disponibles en el menú principal. Mientras que la ventana de la gráfica este activa, selecciona el menú *File* y luego en *Save As*. Esto enlistará seis formatos de archivo (metafile, postscript, PDF, PNG, BMP y JPG en tres niveles de calidad), así que tienes muchas opciones.

¿Cuál es la mejor opción de formato de gráficas? En general el formato **"metaarchivo MS mejorado"**, conserva la calidad de la gráfica, aún cuando se redimensiona en la aplicación que se utilice después. Por otro lado, JPG es la elección popular, ya que el tamaño del archivo es compatible con internet. Excepto por muy raras circunstancias, no se recomienda el mapa de bits (bitmap), puesto que es muy grande y tiene muy poca calidad de imagen cuando se redimensiona. El formato de archivos postscript es útil cuando se incluye el archivo del gráfico en otro archivo del mismo formato, o cuando la impresora es compatible. La calidad de la imagen no se deteriora cuando se redimensiona.

Utiliza el formato metafile para aplicaciones de MS office (por ejemplo Word, Powerpoint). Dentro de la ventana del gráfico haz click derecho y selecciona *copy as metafile*. Esto transferirá la figura al portapapeles de windows. Ve a Powerpoint y pégalo como metafile. Para publicaciones, guarda los gráficos con formato jpg.

Exportando datos

```
write.table(x, file = "dataOut.txt", append = FALSE, col.names = NA,
sep = " ")
# with the above commands a textfile will be created in in the
working directory. To change the working directory use the following
getwd()                                #displays the path for the working
    directory
dir()                                  # lists the files in the current work
    directory
setwd("C:/Axel StatisticsR")           # sets the working directory manually
# open the text files in excel using the text import wizard.
```

Manejo de los paquetes de R

Hay varias páginas en internet dedicadas a temas estadísticos específicos. Por ejemplo, para el análisis de los datos de un microarreglo utilizando R, puedes consultar: <http://bioconductor.org/>

Descarga los paquetes de R desde diversos sitios de Internet y guárdalos en carpetas como archivos zip. Dentro del programa R, usa la opción *install from local zip files*. Instala los paquetes directamente desde R, utilizando los siguientes comandos:

```
install.packages("corrgram")
install.packages("agricolae")
install.packages("gplots")
install.packages("UsingR")
install.packages("lattice")
install.packages("maanova")
```

Once you have downloaded the packages, activate them with the library function. Then you can use the extended functions in those packages.

Example (agricolae)

```
data1=read.table(file="clipboard", header=T, sep="\t") # import data
attach(data1)      #to make the variables defined in data1
accessible to the console
```

```
library(agricolae)      #load and activate library
model1=aov(Yield~Gen) #define a model and change variables
df=df.residual(model1)
MSerror=deviance(model1)/df
comparison = HSD.test(Yield,Gen,df,MSerror, group=TRUE)
bar.err(comparison,std=TRUE) #for plotting a graph
```

#Make an additional LSD test with:

```
comparison = LSD.test(Yield,Gen,df,MSerror, p.adj="bonferroni", group=FALSE)
bar.err(comparison,std=TRUE) #for plotting a graph
#For adjusting the graph you can modify some options like:
bar.err(comparison,std=TRUE,ylim=c(0,45),density=4,border="blue")
```

#AMMIS biplots:

```
model= AMMI(Env, Entry, Rep, Yield, graph="biplot",number=FALSE)
#For adjusting the graph you can modify some options like:
model= AMMI(Env, Entry, Rep, Yield, xlim=c(-4,4),ylim=c(-4,4), graph="biplot",number=FALSE)
```

Example (Pairwise.T.test)

```
data1=read.table(file="clipboard", header=T, sep="\t") # import data
attach(data1)      #to make the variables defined in data1
accessible to the console
```

```
pairwise.t.test(Var1, Var2) # make t.tests of Var1 grouped by category Var2
```

if you want to export the data you need to use following commands:

```
x=pairwise.t.test(Var1, Var2) # create an object x for exporting
write.table(x$p.value, file = "dataOutPairwise.txt", append = FALSE,
col.names = NA, sep = " ")
```

a textfile will be created in in the working directory.

```
getwd()          #displays the path for the working
                directory
```

```
dir()            # lists the files in the current work
                directory
```

open the text file in excel using the text import wizard.

Another option to make multiple comparisons.

```
model1=aov(var1~var2)
TukeyHSD(model1, "var1")
```

#Example PCA analysis (bioconductor)

#first you need to define the source to download the package

```
source("http://bioconductor.org/biocLite.R")
```

```
biocLite("pcaMethods") #downloading needs only to be done the
                        first time
```

```
library(pcaMethods) #load the library. Necessary to do it every time
you run R
```

```
# For importing data you have 3 options
```

```
## 1) load the matrix with data. Considering that row names are
present
```

```
data=read.table(file.choose(), header=T, sep="\t", row.names=1)
```

```
## or 2) not considering the ID of metabolites or genes
```

```
data=read.table(file.choose(), header=T, sep="\t")
```

```
## or 3) importing data from clipboard
```

```
data=read.table("clipboard", header=T, sep="\t")
```

```
# depends on the type of data, you can use:
```

```
## 1) variables in the rows and samples in the columns
```

```
md = prep(t(data), scale = "none", center = TRUE)
```

```
## or 2) variables in the columns and samples in the rows
```

```
md = prep((data), scale = "none", center = TRUE)
```

```
## then do the analysis. In this case for the first 5 principal
components.
```

```
resPPCA = pca(md, method = "ppca", center = FALSE, nPcs = 5)
```

```
slplot(resPPCA) # to plot the results
```

```
resPPCA@scores # to see all the principal components (treatment)
```

```
resPPCA@loadings # to see all the loadings
```

#Example PCA analysis (native R)

```
data=read.table("clipboard", header=T, sep="\t") #importing data
from clipboard
```

```
# depends on the dataframe you have:
```

```
## 1) variables in the rows and samples in the columns
```

```
md = prep(t(data), scale = "none", center = TRUE)
```

```
## or 2) variables in the columns and samples in the rows
```

```
md = prep((data), scale = "none", center = TRUE)
```

```
md=data.frame(md[,2],md[,3],md[,4],md[,5],md[,6],md[,7])
```

```
## Analysis for 2 principal components.
```

```
pc=princomp(md, scale=TRUE)
```

```
loadings(pc)
```

```
biplot(pc)
```

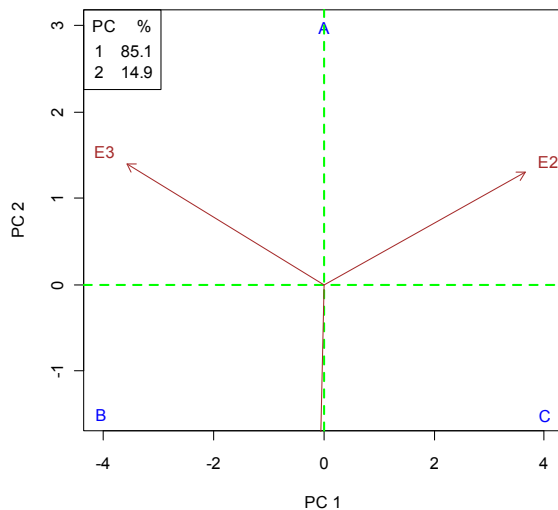
```
pc
```

```
summary(pc)
```

#Example PCA analysis (agricolaeAMMIS)

Rep	Plot	Entry	Env	Yield
1	1	A	E1	36.3
2	4	A	E1	37.9
1	2	B	E1	49.9
2	5	B	E1	50.3
1	3	C	E1	80.1
2	6	C	E1	82.3
1	1	A	E2	44.4
2	4	A	E2	46.3
1	2	B	E2	22.3
2	5	B	E2	20.8
1	3	C	E2	85.4
2	6	C	E2	86.7
1	1	A	E3	82.4
2	4	A	E3	86.3
1	2	B	E3	92.3
2	5	B	E3	92.8
1	3	C	E3	90.3
2	6	C	E3	94.6

```
data5=read.table(file="clipboard", header=T, sep="\t") # import data
attach(data5) #to make the variables defined in data1
accessible to the console
library(agricolae) #load and activate library
model= AMMI(Env, Entry, Rep, Yield, graph="biplot",number=FALSE)
```



```
#For changing the range of the axes use xlim and ylim:
model= AMMI(Env, Entry, Rep, Yield, xlim=c(-3,3),ylim=c(-4,4), graph="biplot",number=FALSE)
```

#Example: Conditional Averaging

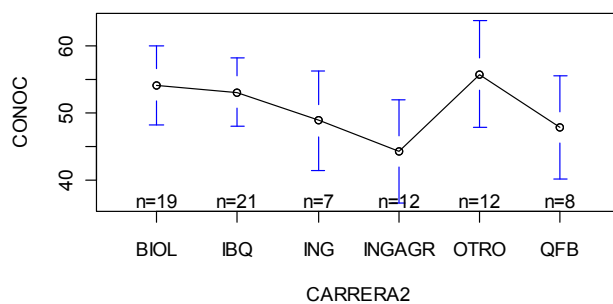
```
# use the same datatable as above (data5)
data5=read.table(file="clipboard", header=T, sep="\t") # import data
```

```
data5$avRep=ave(data5$Yield,data5$Rep) # create new column
with averages among Reps
data5$avEntry=ave(data5$Yield,data5$Entry) # create new column
with average among Entries
```

#Making special graphics

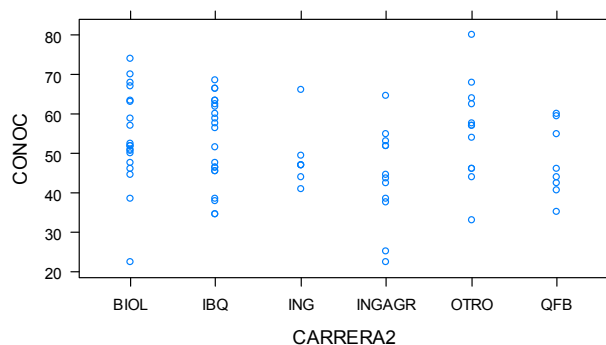
Example (gplots)

```
data1=read.table(file="clipboard", header=T, sep="\t") # import
data
attach(data1) #to make the variables defined in data1
accessible to the console
library(gplots) #load extended plot library
plotmeans(var1~var2) #plot means with error bars
```

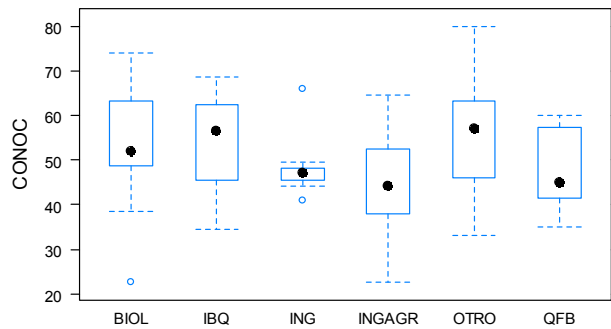


Example (lattice)

```
x=read.table(file="clipboard", header=T, sep="\t") # import data
attach(x) #make variables accessible
library(lattice) #load library
xyplot(var1~ var2)
```

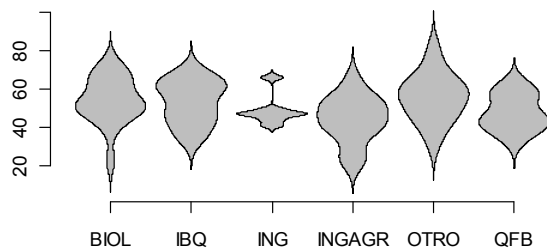


```
dotplot(var1~ var2)
barchart(var1~ var2)
stripplot(var1~ var2)
bwplot(var1~ var2)
```



Example (Simple)

```
install.packages("UsingR")
require(UsingR)
data1=read.table("clipboard", header=T, sep="\t")
attach(data1)
simple.violinplot(var1 ~ var2, col=8)
```

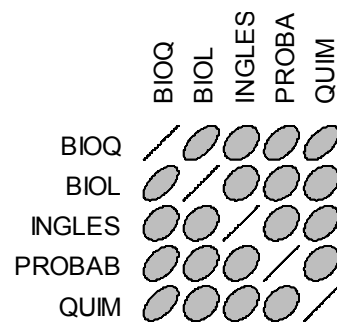


`par(new=TRUE)` #for adding to an existing graph. Next function will add.

`plot(var1 ~ var2, add = TRUE)` # so that a new plot is not started

Example (ellipse)

```
data1=read.table("clipboard", header=T, sep="\t")
library(ellipse)
plotcorr(cor(data1, use = "pairwise.complete.obs"))
```



```
library(corrgram)
corrgram(data1)
```

#Trellis graphs

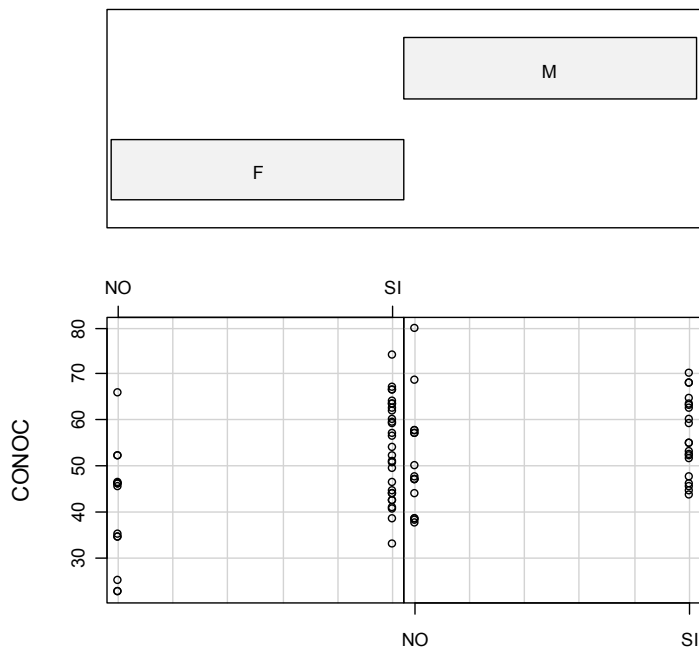
#In order to produce Trellis plots you must load the "Lattice" library and start a "trellis aware" device.

```
library(lattice)
trellis.device()
```

#Make conditional plots with the coplot function

```
coplot(var1 ~ var2 | var3) # plot var1 against var2, subdivided in
categories var3
```

Given : SEXO

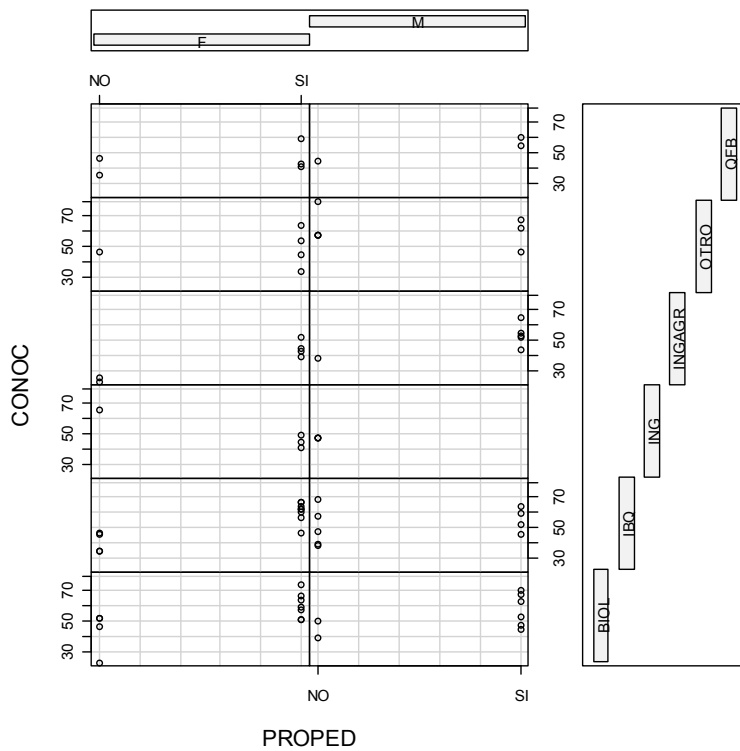


PROPED

It is possible to use two conditional variables in the form of var3*var4

`coplot(var1 ~ var2 | var3*var4)` # plot var1 against var2, subdivided in categories var3 and var4

Given : SEXO



Given : CARRERA2

#Options for the coplot function:

```
coplot(formula, data, given.values, panel = points, rows, columns,
show.given = TRUE, col = par("fg"), pch = par("pch"), xlab =
paste("Given :", a.name), ylab = paste("Given :", b.name), number =
6, overlap = 0.5, ...) co.intervals(x, number = 6, overlap = 0.5)
```

```
# formula
```

Add a trendline

#Use the functions lm() and abline() to add a trendline to an existing plot.

```
data1=read.table("clipboard", header=T, sep="\t")
```

```
attach(data1)
```

```
plot(var1, var2)
```

```
model1=lm(var1~var2)
```

```
abline(model1,lwd=2)
```

#The lwd specifies the line width

```
model1
```

#To show the slope and intercept of the

```
trendline
```

To add text to a graphic

```
text (x, ...) text.default (x, y = NULL, labels = seq(along = x), adj =
NULL, pos = NULL, offset = 0.5, vfont = NULL, cex = 1, col = NULL,
font = NULL, xpd = NULL)
```

```
mtext(text, side = 3, line = 0, outer = FALSE, at = NULL, adj = NA,
cex = NA, col = NA, font = NA, vfont = NULL)
```

```
title(main="Name")
```

Multiple plots per page

#You can combine multiple plots into one overall graph, using either the

par() or layout() function. With the par() function, the option mfrow=c(nrows, ncols) creates a matrix of nrows x ncols plots that are filled in by row. mfcrow=c(nrows, ncols) fills in the matrix by columns.

```
par(mfrow=c(2,2))
```

```
plot(var1,var2, main="Title plot1")
```

```
plot(var1,var3, main="Title plot2")
```

```
hist(var1, main="Title Histogram")
```

```
boxplot(var1, main="Title Boxplot", xlab="label1", ylab="label2")
```

The layout() function has the form layout(mat) where mat is a matrix object specifying the location of the N figures to plot.

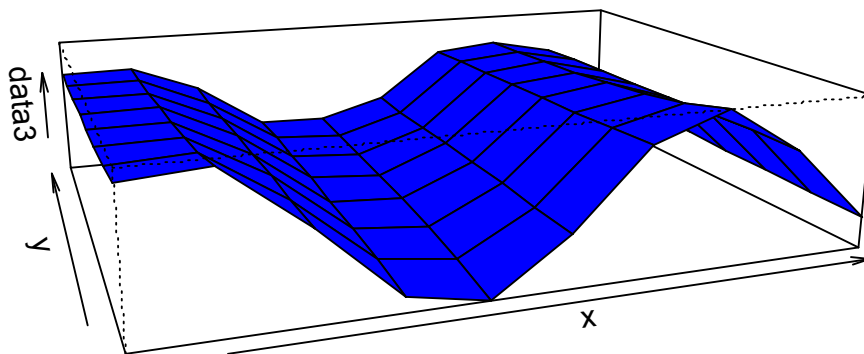
Example of one figure in row 1 and two figures in row 2


```
layout(matrix(c(1,1,2,3), 2, 2, byrow = TRUE))
hist(wt)
hist(mpg)
hist(displ)
```

3D Perspective Plot

0.841	0.746	0.678	0.628	0.587	0.554	0.526	0.502
0.909	0.789	0.710	0.652	0.606	0.570	0.540	0.514
0.141	0.141	0.140	0.140	0.139	0.139	0.138	0.138
-0.757	-0.687	-0.634	-0.592	-0.558	-0.530	-0.505	-0.484
-0.959	-0.819	-0.730	-0.667	-0.619	-0.580	-0.548	-0.521
-0.279	-0.276	-0.272	-0.269	-0.266	-0.263	-0.260	-0.257
0.657	0.611	0.573	0.543	0.516	0.494	0.474	0.456
0.989	0.836	0.742	0.676	0.625	0.585	0.553	0.525
0.412	0.401	0.390	0.380	0.371	0.363	0.355	0.347
-0.544	-0.518	-0.495	-0.475	-0.457	-0.441	-0.427	-0.414

```
data3=read.table(file="clipboard", header=F, sep="\t") # import data
without header
data3=data.matrix(data3) # convert table to a data matrix
x=c(1:10) # define x variable. In this case 10 values
y=c(1:8) # define y variable. In this case 8 values
par(mfrow=c(2,2)) # for plotting 4 graphs in one page
contour(x, y, data3, col="blue") # plot contour
image(x, y, data3) # plot image
persp(x, y, data3, col="blue", theta = -20, phi = 15, scale = FALSE,
      axes = TRUE)
```

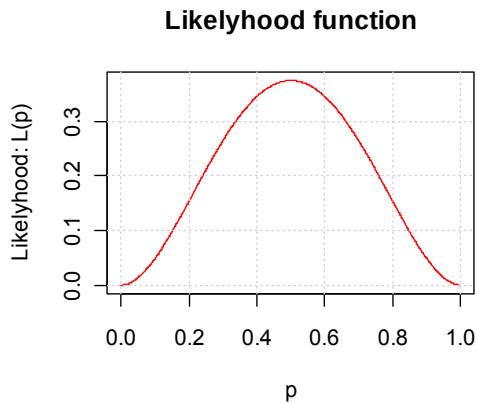


Firsts Exercises

Example of Graph:

```
p <- c(0:1000)/1000
x <- 6*(p^2)*((1-p)^2)
plot(p,x,col="red",type="l")
```

```
title(main="Likelihood function", ylab="Likelihood: L(p)")
grid()
```



Example of data import and analysis: To perform statistical tests, first copy the data into the clipboard and then use the following commands in R:

```
data1=read.table(file="clipboard", header=T, sep="\t")
summary(data1)      #lists a summary of the data
attach(data1)       #to make the variables defined in data1
                    #accessible to the console
t.test(var1,var2)    #T.test between data contained in two variables. Use the names of the
                    #variables inside the object. In my case they were named as var1, and var2.
t.test(var1~var2)    #T.test within data of variable1 grouped by categories specified in
                    #variable 2. Grouping factor of var2 must have exactly 2 levels.
```

Note the important difference between the `t.test(var1,var2)` and `t.test(var1~var2)` since in the first case you are comparing all values of var1 against all values of var2. In the second case, you make subgroups of the values contained in var1 depending in the grouping levels defined in var2. If you have more than 2 grouping levels in var2, then it is better to use the `pairwise.t.test` function.

```
pairwise.t.test(var1, var2) # make t.tests of var1 grouped by categories
                             var2. The result is a pairwise comparison matrix. Note the comma
                             in (var1, var2)
```

```
ave(var1, var2) # make averages of var1 grouped by factor categories
                 var2. The result is a vector of the same length as var1.
```

#To make an analysis of variance:

```
a=aov(var1~var2)      #defines a new object a with the results of the analysis of variance
a                     #lists the content of the previously created object a
summary(a)            #lists the summary of the statistical results of object a
attributes(a)         #lists all attributes of new object
```

#To make a linear model:

```
model1=lm(var1 ~var2+var3+var4)
summary(model1)       # show summary of the linear model
attributes(model1)    # shows all attributes of the new object
```

#Examples & Exercises

Make	Cylinder	Weight	Mileage	Type
Volkswagen	V4	2330	26	Small
Ford	V4	2345	33	Small
Eagle	V4	2560	33	Small
Acura	V6	3265	20	Medium
Buick	V6	3325	23	Large
Chrysler	V6	3450	22	Medium

#Copy the above table from word and then import data using the command `read.table(file="clipboard" header=T)` in the R console as following:

```
CarData=read.table(file="clipboard", header=T)
```

Boxplots are very useful when comparing grouped data. For example, side-by-side boxplots of weights grouped by vehicle types are shown below:

```
attach(CarData)
```

```
boxplot(Weight ~Type)
```

```
# create boxplot of Weight grouped by
```

```
Type
```

```
title("Weight by Vehicle Types")
```

```
# put title to the graph
```

```
boxplot(Mileage
```

```
~Cylinder)
```

```
title("Mileage by Cylinder")
```

```
plot(Mileage, Weight)
```

```
# create a x,y
```

```
scatterplot
```

```
title("Mileage by weight")
```

```
# put title to the graph
```

#Example of grain yield trial data

REP	BLK	PLOT	ENTRY	RANGE	ROW	GYF	AD	ASI	PH	EH	EPO	EPP
1	1	1	6	1	1	11.00	60.0	-2.0	185.0	96.0	0.52	1.35
1	1	2	5	1	2	8.60	58.0	1.0	194.0	99.0	0.51	0.96
1	1	3	2	1	3	7.46	60.0	0.0	165.0	70.0	0.42	1.00
1	1	4	28	1	4	10.99	63.0	1.0	192.0	109.0	0.57	1.08
1	1	5	11	1	5	11.23	63.0	-2.0	224.0	120.0	0.54	1.04
1	2	6	12	1	6	11.32	64.0	1.0	205.0	109.0	0.53	1.00
1	2	7	1	1	7	8.20	60.0	0.0	210.0	109.0	0.52	0.91
1	2	8	19	1	8	10.01	61.0	3.0	215.0	113.0	0.53	0.96
1	2	9	27	1	9	12.46	63.0	1.0	217.0	109.0	0.50	1.00
1	2	10	26	1	10	13.19	64.0	0.0	215.0	113.0	0.53	1.24
1	3	11	25	1	11	10.23	64.0	-2.0	233.0	108.0	0.46	1.00
1	3	12	9	1	12	11.64	63.0	-3.0	180.0	85.0	0.47	1.13
1	3	13	4	1	13	7.08	60.0	1.0	205.0	82.0	0.40	1.40
1	3	14	18	1	14	5.32	66.0	0.0	208.0	92.0	0.44	1.07
1	3	15	13	1	15	8.96	65.0	0.0	213.0	103.0	0.48	1.00

1	4	16	3	1	16	9.61	60.0	1.0	215.0	113.0	0.53	1.04
1	4	17	14	1	17	11.20	60.0	1.0	203.0	92.0	0.45	1.00
1	4	18	30	1	18	11.98	65.0	0.0	209.0	102.0	0.49	0.91
1	4	19	20	1	19	5.98	64.0	1.0	203.0	109.0	0.54	0.85
1	4	20	7	1	20	6.53	64.0	1.0	174.0	90.0	0.52	0.94
1	5	21	23	2	20	7.55	64.0	1.0	195.0	88.0	0.45	1.12
1	5	22	10	2	19	11.30	64.0	1.0	213.0	106.0	0.50	0.96
1	5	23	17	2	18	9.00	63.0	1.0	218.0	106.0	0.49	1.00
1	5	24	8	2	17	8.41	66.0	0.0	204.0	93.0	0.46	0.96
1	5	25	24	2	16	10.06	66.0	0.0	230.0	115.0	0.50	1.00
1	6	26	29	2	15	11.52	65.0	1.0	210.0	118.0	0.56	1.09
1	6	27	21	2	14	6.82	61.0	1.0	208.0	104.0	0.50	0.76
1	6	28	16	2	13	12.93	60.0	1.0	203.0	105.0	0.52	0.96
1	6	29	15	2	12	7.36	61.0	1.0	212.0	113.0	0.53	0.87
1	6	30	22	2	11	9.66	60.0	1.0	208.0	107.0	0.51	0.92
2	1	31	28	2	10	12.25	60.0	1.0	239.0	130.0	0.54	1.16
2	1	32	25	2	9	9.88	61.0	1.0	220.0	121.0	0.55	0.96
2	1	33	15	2	8	7.91	62.0	4.0	230.0	123.0	0.53	0.92
2	1	34	14	2	7	10.53	57.0	0.0	211.0	108.0	0.51	1.05
2	1	35	1	2	6	6.80	57.0	-1.0	210.0	95.0	0.45	0.95
2	2	36	6	2	5	10.36	58.0	0.0	220.0	117.0	0.53	1.04
2	2	37	27	2	4	10.96	60.0	1.0	223.0	121.0	0.54	1.00
2	2	38	13	2	3	13.86	61.0	1.0	209.0	113.0	0.54	0.96
2	2	39	10	2	2	12.55	60.0	1.0	217.0	102.0	0.47	0.92
2	2	40	21	2	1	9.82	61.0	-2.0	207.0	92.0	0.44	1.05
2	3	41	18	3	1	12.29	61.0	1.0	217.0	102.0	0.47	1.17
2	3	42	22	3	2	9.26	62.0	0.0	218.0	124.0	0.57	1.06
2	3	43	11	3	3	13.22	60.0	-2.0	219.0	104.0	0.47	1.19
2	3	44	3	3	4	6.90	60.0	-1.0	219.0	108.0	0.49	0.83
2	3	45	8	3	5	9.48	60.0	1.0	225.0	123.0	0.55	1.00
2	4	46	12	3	6	12.37	60.0	0.0	200.0	103.0	0.52	1.04
2	4	47	20	3	7	8.72	59.0	-1.0	224.0	103.0	0.46	1.00
2	4	48	2	3	8	7.30	60.0	1.0	190.0	97.0	0.51	0.81
2	4	49	17	3	9	7.64	60.0	1.0	217.0	83.0	0.38	1.00
2	4	50	29	3	10	11.82	65.0	0.0	205.0	97.0	0.47	1.05
2	5	51	9	3	11	10.23	61.0	0.0	204.0	115.0	0.56	1.14
2	5	52	19	3	12	8.46	61.0	1.0	209.0	101.0	0.48	1.05
2	5	53	7	3	13	8.10	60.0	1.0	171.0	74.0	0.43	1.05
2	5	54	5	3	14	6.43	58.0	0.0	183.0	76.0	0.42	1.00
2	5	55	23	3	15	7.53	60.0	1.0	195.0	90.0	0.46	1.43
2	6	56	30	3	16	10.49	66.0	0.0	189.0	86.0	0.46	1.20
2	6	57	4	3	17	7.32	60.0	1.0	194.0	96.0	0.49	1.10
2	6	58	24	3	18	5.60	63.0	2.0	191.0	88.0	0.46	0.93
2	6	59	26	3	19	10.92	60.0	1.0	227.0	107.0	0.47	1.30
2	6	60	16	3	20	10.96	60.0	1.0	197.0	96.0	0.49	1.05

#Copy the above table and then type the command in the R Console:

```
TrialData=read.table(file="clipboard", header=T)
attach(TrialData)           #make variables accessible
```

```
boxplot(GYF ~ENTRY) # create boxplot of GYF grouped by category ENTRY
```

```
boxplot(GYF ~ENTRY, notch=T) # Same as above but with notch option set to TRUE. If the notches of two plots do not overlap this is 'strong evidence' that the two medians differ
```

```
boxplot(GYF ~REP) # create boxplot of GYF grouped by category REP
```

```
a=aov(GYF ~ENTRY) # make anova of GYF grouped by category ENTRY
```

```
summary(a) # show summary of anova
```

```
pairwise.t.test(GYF, BLK) # make t.tests of GYF grouped by category BLK
```

#The pairwise.t.test function is quite useful since it makes T-tests with all possible two groups and provides a matrix of P values. This allows identifying groups that are significantly different from each other. Unfortunately, it seems that R can only handle about less than 20 groups.

#Linear models let you find causal relationships of the dependent variables. A breeder, for example, would be interested in modeling grain yield in dependence of other trial parameters and secondary traits. Copy the table with grain yield data and then enter following command in R for modeling.

```
TrialData=read.table(file="clipboard", header=T, sep="\t")
```

```
attach(TrialData) #make variables accesible
```

```
model1=lm(GYF ~REP+BLK+PLOT+ENTRY+AD+ASI+PH+EH+EPO+EPP)
```

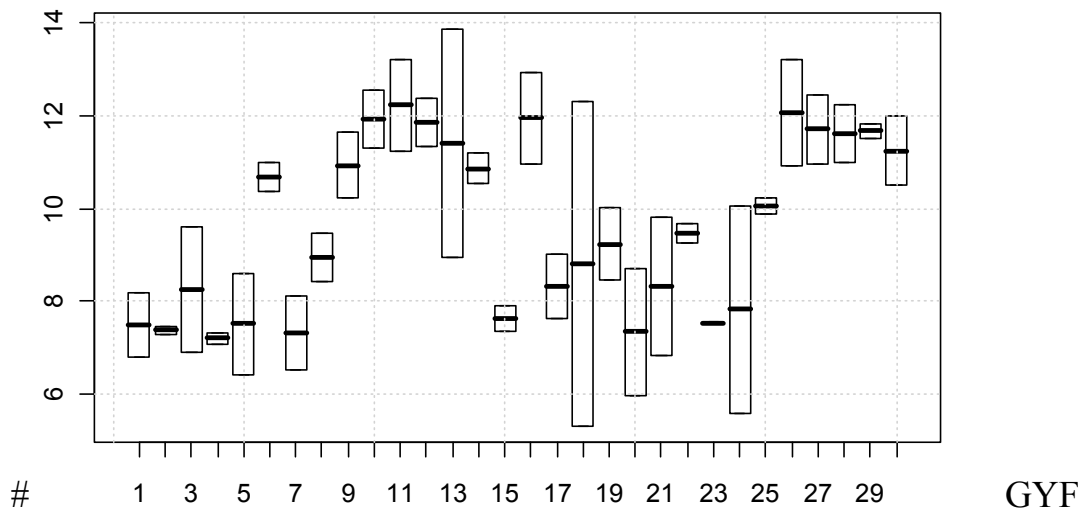
```
summary(model1) # show summary of the linear model
```

the lm coefficients let you know how much is the dependence from the other variable, whereas the p values indicate if the relationships are significant or not.

#Construct a new variable called SELINDEX taking into account the coefficient of those variables which were significant in the linear model. In this case use GYF, EPP and ASI

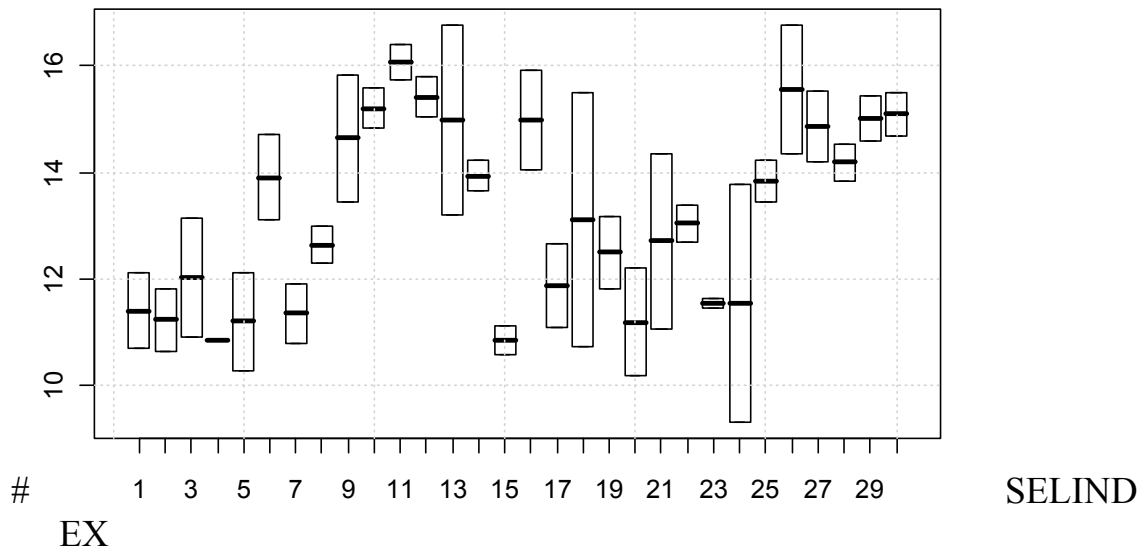
```
TrialData$SELINDEX=GYF+EPP*3.5-ASI*0.4 #Define a new variable called SELINDEX based on a linear combination of other variables GYF, EPP and ASI. The coefficients given by the linear model were used as weighting factors for this selection index.
```

```
boxplot(GYF ~ENTRY)
```



#From the above boxplot, decide which is the highest yielding entry? Is it entry 10, 11, 12, 13 or 26? Are the differences significant? We can plot the selection index to find an answer.

`boxplot(SELINDEX ~ENTRY)`



`title(main="Selection Index")` # to put a title to the graph

#From the above boxplot it can be seen that entry 11 is significantly better than entry 10, 12, 13 and 26. This was not significant if only looking at GYF parameter (previous boxplot). This highlights the usefulness of including secondary traits into the analysis.

Some interesting codes

#To add an extra column containing the rank to a ragged array indexed by more than one grouping factors:

```
rank.lists <-with( data1, tapply( yield, list( site=site, year=year ), rank
))
```

#To add an additional column ``rank" containing the rank of the ``yield" of

the different varieties in relation to the indices ``year" and ``site" to the data1 object. I achieved to calculate the ranks but I do not manage to merge this result to the original dataframe.

```
model1 <- lme(math ~ grade, mathdata, random=~grade|studentid)
```

#Ejemplo en español

```
data1 = read.table(file="data1.txt", header=T)
```

#El objeto "data1" contiene ahora los datos del archivo data1.txt.

Utilice help() para aprender mas de las funciones que va a utilizar para analizar sus datos. Mas que poner el nombre del archivo, es mejor usar la función file.choose() para escoger cualquier archivo de texto en cualquier directorio.

```
data1 = read.table(file.choose(), header=T)
```

```
summary(data1) # Le da un resumen de las estadísticas básicas
```

```
attributes(data1) # Le da los "atributos"
```

```
attributes(data1)$names # Le da los nombres de las variables
```

```
attach(data1) # Anexa los datos de manera que R sepa de la existencia de esas variables sin necesidad de referirnos al objeto.
```

```
summary(long) # Nos da el resumen de la variable "long"
```

```
summary(data1$long) # tambien da el resumen de la variable "long"
```

para acceder una variable sin el commando attach ponga el nombre del objeto seguid del simbolo de pesos \$

```
plot(longitud,c) # Obteniendo una gráfica (scatter plot o diagrama de dispersión) de longitud x c
```

#Glosary of frequent R functions

```

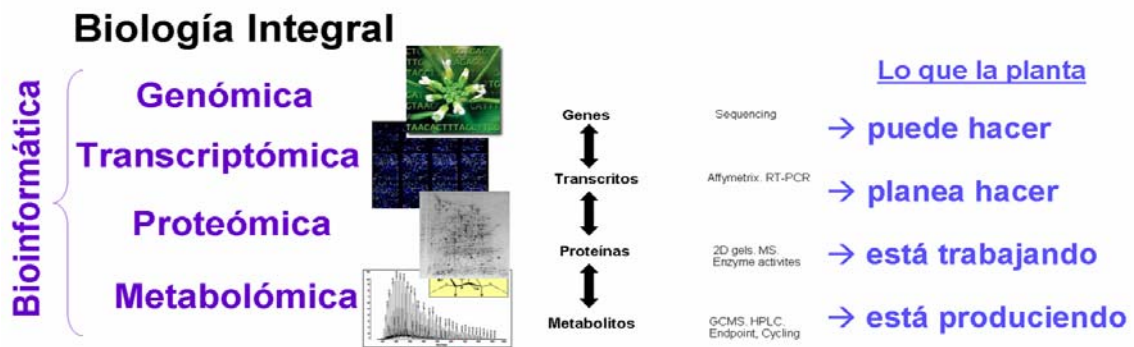
? name #does the same as help(name)
a=aov(var1~var2) #defines a new object a with the results of the
analysis of variance
attach(data1) #to make the variables defined in data1 accessible to the console
attach(object) #makes the elements of object available to be called directly
attributes(data1) #lists the attributes of data1: variable and row names
boxplot(data1) #boxplot
data1=edit(data.frame()) # opens a window to define data1 manually
data1=read.table(file.choose(),header=T) # select file with a popup window
data1=read.table(file="clipboard") #import data from clipboard. Works only if
there are no empty cells
data1=read.table(file="clipboard", header=T, sep="\t") #import data from clipboard.
Works even if empty cells are present
data1=read.table(file="clipboard", header=T, sep="\t", na.strings="") #import data from
clipboard. Works even if empty cells are present. Fills empty cells
with NA.
detach(data1) #to detach the variables defined in data1
help(name) #provides help about "name"
hist(data1) #simple histogram
ls() #shows the variables in the workspace
names(table) #what are the variables inside the table
pairs(data1) #splom plot
pairwise.t.test(var1, var2) # make t.tests of var1 grouped by categories var2. The result
is a pairwise comparison matrix.
plot(data1) #simple scatter plots
rm(list = ls()) #removes all variables from the work space
rm(variable) #removes that variable or object
round(cor(data1),2) #correlation matrix rounded to 2 decimals
summary(lm(var1~0+var2*var3)) #summary of the results of a linear
model with interactions
summary(lm(var1~var2)) #summary of the results of a linear model
summary(a) #summary of the contents of object a
summary(aov(var1~var2)) #summary of the analysis of variance
summary(data1) #lists a summary of the data
t.test(var1,var2) #T.test between two variables. Use the names of
the variables inside the object. In my case they were named as var1,
and var2.
t.test(var1~var2) #T.test within data of variable1 grouped by
categories specified in variable 2. Grouping factor of var2 must have
exactly 2 levels.

```

8. Biología de Sistemas

8.1. Introducción

La biología de sistemas, también llamada biología integral, incluye la genómica, transcriptómica, proteómica y metabolómica, todas ellas unidas por las herramientas bioinformáticas. Se trata de entender las funciones moleculares de un organismo en términos globales. Para lograrlo, primero hay que aprender el idioma del DNA y entender el lenguaje de las proteínas. La expresión de un gen, o la aparición de un fenotipo se pueden entender como un flujo de información. En algunos casos es un flujo monodireccional, mientras que en otros es bidireccional. Al inicio está el DNA y los genes, que representan el potencial de lo que una planta puede hacer; después está el RNA y los transcritos que es lo que la planta planea hacer para adaptarse al ambiente durante su desarrollo; posteriormente se forman las proteínas que definen lo que la célula está realizando en determinado momento; al final se transforman los metabolitos, que es lo que la planta produce e interesa cosechar. Por ejemplo; en el caso del maíz, interesa el almidón, las proteínas y los lípidos que se obtienen de los granos.



Genoma: Totalidad de genes que tiene un organismo. Se puede definir como el DNA que está dentro del núcleo y organelos de las células. Hay que recordar que algunos organelos tienen su propio genoma (cloroplastos y mitocondrias).

Metagenoma: Totalidad de genes que se encuentran en un determinado hábitat. Es la suma de los genomas de las especies individuales que viven en un ambiente.

Epigenoma: Conjunto de modificaciones que no alteran la secuencia de nucleótidos, pero que sí afectan la expresión de los genes, y además son heredables (metilación del ADN, acetilación de histonas, etc.).

Transcriptoma: Conjunto de RNAs que están presentes en un determinado momento en un tejido específico. Por lo regular se refiere a los mRNAs que codifican para proteínas.

Proteoma: Conjunto de proteínas que están presentes en un tejido. Se refiere tanto a la acumulación de proteínas estructurales, así como enzimas funcionales y factores regulatorios.

Metaboloma: Conjunto de metabolitos de un determinado tejido. Los metabolitos son todas aquellas moléculas orgánicas que participan en el metabolismo celular. Muchas veces, los metabolitos es lo que finalmente nos importa, porque es lo que cosechamos y tiene importancia económica.

8.1.1. Justificación de las ciencias bioquímicas

El objetivo último de la investigación en las ciencias biológicas es conocer y comprender cómo funcionan los organismos vivos. La biotecnología surge de la combinación de la tecnología con la biología para generar un beneficio para el ser humano y el medio ambiente. Se trata de aprovechar a los procesos y las entidades biológicas para derivar mejores productos y servicios.

La bioquímica se puede entender como la química de la vida. Es decir, es un tema fundamental sobre las células y sus moléculas, tanto de su estructura como el funcionamiento de las mismas. Si bien la bioquímica se considera como un área científica tradicional, en tiempos recientes han surgido tendencias que la enriquecen. Las ciencias bioquímicas (CBs), siendo una disciplina históricamente reduccionista, pueden ampliar su impacto con el enfoque holístico que permite las ciencias genómicas, por ejemplo.

Con el advenimiento de las llamadas ciencias *ómicas*¹, el principal reto que enfrentan es demostrar su valía en torno a contestar preguntas biológicas, validar o refutar hipótesis científicas y brindar soluciones a problemas biotecnológicos concretos. Sin embargo, la posibilidad de generar gran cantidad de datos, no incluye implícitamente que esa información sea aprovechable. Se necesita un esfuerzo intelectual para que esos datos se conviertan en información biológicamente relevante y pueda ser interpretada correctamente a nivel de sistema.

Un posgrado (doctorado) pretende formar recursos humanos del mas alto nivel científico, y es por ello que es primordial considerar las bases y fundamentos de la bioquímica tradicional, pero también incluir algunas herramientas para

¹ La definición de las ciencias *ómicas*, así como su naturaleza epistemológica, es un tema en debate; sin embargo, para fines prácticos éstas se pueden definir como aquellos abordajes conceptuales y experimentales (genómica, transcriptómica, proteómica, metabolómica, etc) que surgen de la posibilidad técnica de obtener una gran cantidad de datos de forma muy rápida (por ejemplo, secuenciar la totalidad del contenido genético de un organismo en una sola corrida), y que permiten el estudio -de una forma integrada y a nivel sistema- de todos los componentes moleculares de una célula.

que los estudiantes estén mejor preparados a los retos del futuro, incluyendo los conceptos que demandan las preguntas científicas de vanguardia.

En particular, se espera que los estudiantes contribuyan al análisis de los grandes bancos de información² que se generan en la actualidad de manera rutinaria y de prácticamente cualquier organismo. Así, es de esperarse que los jóvenes estudiantes e investigadores sean capaces de un proceso de “re-invencción” en el que las lecciones aprendidas de los estudios reduccionistas sean aprovechadas como pilares para el desarrollo de marcos teóricos y experimentales, que permitan el estudio de las funciones celulares de una forma integrada. Este proceso, de hecho, comienza a vislumbrarse, en tanto que si bien el entendimiento “no reduccionista” del metabolismo se ha visto favorecido significativamente por el advenimiento de las ciencias *ómicas*, muchas de estas herramientas descansan en principios teóricos y experimentales de la bioquímica, como por ejemplo la secuenciación por el método Sanger (electroforesis capilar) o el método del pirofosfato (luminiscencia enzimática en base a un sustrato limitante).

El reto biológico al que se enfrentan los investigadores hoy en día sigue vigente, y se pone de manifiesto en dos ejemplos concretos: por un lado, es común que la función biológica – sustentada en mecanismos bioquímicos – de hasta un 50% de los genes identificados por un proyecto genómico, sea completamente desconocida; por el otro lado, el contenido genómico de un cierto organismo, y las funciones biológicas esperadas en el mismo – muchas de éstas previamente descritas experimentalmente a través de ensayos bioquímicos – no concuerdan, o bien, la relación gene-función no resulta obvia.

Algunos elementos de los cursos de posgrado nacen de dichas exigencias biológicas, lo que justifica que, si bien se busque fortalecer los preceptos tradicionales de las CBs, principalmente se desarrollen nuevos conocimientos de vanguardia, una actitud crítica e inquisitiva, y las habilidades necesarias para llevar a cabo investigación biológica en la actualidad. Cabe notar que esta justificación está en concordancia con las actividades de investigación del común denominador de los laboratorios de CINVESTAV Irapuato.

Los objetivos particulares de un curso de doctorado podían ser en primer lugar: adquirir un conjunto de conocimientos bioquímicos centrales y actualizados. En segundo lugar: desarrollar habilidades para adquirir estos conocimientos a partir del análisis crítico de la información publicada. Otras habilidades que los estudiantes deben adquirir durante su formación, son por ejemplo: diseñar, realizar y analizar experimentos, y presentar apropiadamente sus resultados.

² La información en las bases de datos accesibles por internet (base de la “ómica”) forma parte de la información publicada que se debe tomar en cuenta tanto como (no más que) la información en los artículos científicos, de la cual deriva la información en los libros de texto. A los estudiantes se les debe hacer notar, que hay errores en las tres fuentes de información citadas (no solamente en los libros de texto), por lo que el análisis crítico de la misma debe fomentarse.

8.2. Genómica

Se denomina genómica al conjunto de ciencias y técnicas dedicadas al estudio integral del funcionamiento, la evolución y el origen de los genomas. El genoma es **la suma de todos los genes** de un determinado organismo. Es decir, es toda la información que está codificada en el DNA de una célula. La genómica usa conocimientos derivados de distintas ciencias como son: biología molecular, bioquímica, informática, estadística, matemáticas, física, etc. Muchas veces, la genómica es usada como sinónimo de otras áreas de estudio relacionadas, por ejemplo, la proteómica o la transcriptómica.

Las ciencias genómicas han tenido un importante auge en los últimos años, sobre todo gracias al desarrollo de las técnicas de secuenciación de DNA, a los avances en bioinformática, y a los procedimientos de ensamblaje y anotación cada vez más sofisticadas. La genética clásica parte de un fenotipo, generalmente mutante, y busca el o los genes responsables de dicho fenotipo. A diferencia, la estrategia de la genómica es en la dirección opuesta. Tiene como objetivo predecir la función de los genes a partir de la determinación previa de todas las secuencias. Es decir, primero se determina el genotipo (secuencia nucleotídica) y después se trata de averiguar cual es el fenotipo que causa ese gen (función de la proteína).

La genómica (principalmente potenciada por el proyecto del genoma humano) es un campo de investigación con un gran potencial e impacto. La aplicación de las nuevas tecnologías para la secuenciación del DNA a diversas áreas del conocimiento puede dar a lugar a nuevas disciplinas, como por ejemplo, la medicina genómica y la agro-biotecnología. La bioinformática, la proteómica, la metabolómica, y otras, eran antes consideradas como tecnologías exóticas. En la actualidad son muy comunes en los países desarrollados, pero cada vez están siendo más implementados en algunos en países latinoamericanos.

Roderick en 1986 introdujo el término genómica para discernir la secuencia de bases o moléculas que constituyen el genoma de los organismos. Los objetivos principales de la genómica son: conocer la totalidad de los genes de un ser vivo y la expresión de sus transcritos, en relación con sus procesos metabólicos, sus características estructurales y sus características funcionales; así como relacionar y aplicar dicha información en los procesos de enfermedad del organismo en estudio.

En este sentido, se calcula que la diferencia genética entre dos individuos de una sola especie es de 0,1 a 1% del genoma. Dentro de estas variaciones se encuentran los polimorfismos de nucleótido simple (SNP's), la forma más común de variación del DNA en muchas especies. El estudio de los

SNP's es de gran utilidad para entender la susceptibilidad o resistencia genética a una enfermedad en particular, y conocer la predisposición a determinadas patologías y variaciones genéticas del metabolismo. (ver referencia a proyecto HapMap).

En humanos, la investigación genómica acorta el camino en la investigación médica hacia sus aplicaciones en la prevención, diagnóstico y tratamiento de enfermedades transmisibles, genéticas, cardiovasculares, cáncer, diabetes, psicosis mental grave entre otras. La medicina genómica está en sus inicios en los países desarrollados y en corto plazo servirá de base para el diseño de nuevos mecanismos de atención a la salud. En el área agrícola, la genómica vegetal y microbiana puede servir de base para el desarrollo de nuevas variedades vegetales con mayor productividad y calidad nutricional.

La pregunta a plantear es ¿cómo podemos tener acceso a los beneficios de la genómica? Muchos de los conocimientos actuales de la genómica ya han sido publicados, pueden considerarse como conocimientos del público en general, aunque existen compañías que hacen uso de esta información para desarrollar productos y servicios de carácter privado. Lo que hace falta son recursos humanos altamente calificados que sirvan de puente entre la investigación, los mejoradores y los agricultores. Finalmente, el beneficio de la ciencia se debe de llevar del microtubo del laboratorio al campo, y del campo al plato del consumidor.

Tarea 8-1:

Con sus palabras describa las diferencias entre: genoma, epigenoma, transcriptoma, proteoma y metaboloma. De también un ejemplo de que tipos de equipos de laboratorio y metodologías se usan para cada una.

Tarea 8-2:

Describa las diferencias entre: suma de todos los genes y la suma de todas las secuencias de DNA en una célula. Considere la importancia de aquello que no se puede definir como un gen en el sentido estricto, como secuencias repetitivas, transposones, telómeros, centrómeros, etc.

8.3. Transcriptómica

Con el aumento de secuencias de genes en las bases de datos, se ha producido una explosión en la aplicación de herramientas a gran escala para analizar perfiles de expresión en los sistemas biológicos. La capacidad de analizar cualitativa y cuantitativamente poblaciones de mRNA (transcriptómica) y proteínas (proteómica) suscitan la tentadora posibilidad

de descifrar las rutas reguladoras y funcionales que representan el puente entre el genotipo y el fenotipo.

En términos prácticos, el perfil transcripcional de un organismo se puede determinar usando microarreglos de cDNA u oligonucleótidos. Las tecnologías de miniaturización y automatización han permitido masificar el estudio de todo el transcriptoma. Antes, con los *northers* solo se estudiaban unos pocos genes (menos de 10). La transcriptómica es más simple que la proteómica ya que el RNA es químicamente más homogéneo que las proteínas. El mRNA se puede extraer, manipular *in vitro*, amplificar y secuenciar. Esto permite analizar cuantitativamente miles de genes de forma simultánea.

Sin embargo, la técnica que es más confiable para la cuantificación de transcritos es el análisis cuantitativo de PCR en tiempo real (qRT-PCR).

Tarea 8-3

Averigüe por internet cuales son las metodologías que más se usan para detectar ácidos nucleicos (DNA o RNA).

Tarea 8-4

Haga una revisión de artículos científicos que utilizan el enfoque experimental de la transcriptómica. Escoja un artículo y discútalos con sus compañeros.

Tarea 8-5

Averigüe que distingue las metodologías que se enfocan a los ácidos nucleicos de las que analizan proteínas y metabolitos. Haga una tabla comparativa con ventajas y desventajas.

8.4. Proteómica

Un eslabón en la cadena de información entre el genotipo y el fenotipo es el proteoma. El término proteómica se está convirtiendo en algo ambiguo que engloba casi cualquier aspecto de la expresión de proteínas, el estudio de su estructura o su función. Sin embargo, en términos generales, debería definirse como el estudio del perfil del conjunto de proteínas de una célula o un tejido en un determinado momento. Sería análogo al perfil de transcritos en una situación dada. A diferencia de los estudios de transcriptómica en los que el RNA se extrae y manipula con facilidad, los estudios con proteínas presentan múltiples desafíos. Primero, la heterogeneidad fisicoquímica y complejidad estructural de las proteínas, lo que complica su extracción, solubilización, manejo, separación e identificación. Además, no existe ningún tipo de tecnología similar a la

PCR que permita amplificar aquellas proteínas que son poco abundantes. Algunos de estos problemas son particularmente críticos cuando se trabaja con tejidos vegetales. Hay ya algunas bases de datos disponibles que contienen datos sobre perfiles de proteínas. La segunda categoría en el análisis de proteómica es la llamada proteómica comparativa, donde el objetivo es caracterizar las diferencias entre distintas poblaciones de proteínas en dos situaciones biológicas diferentes. Este tipo de enfoque es en cierta medida análogo a los perfiles comparativos de genes con librerías de cDNA substractivas.

Gran parte del esfuerzo en proteómica se ha realizado para desarrollar métodos de alta resolución para separar e identificar proteínas. Uno de los pasos críticos es la extracción y preparación de muestras. Los tejidos vegetales son usualmente más complicados a la hora de preparar extractos de proteínas que otras fuentes biológicas. Tienen menor contenido de proteínas, usualmente contienen más proteasas y otros compuestos (polisacáridos, lípidos y compuestos fenólicos) que interfieren en la estabilidad de las proteínas, además de toda una colección de metabolitos secundarios que pueden interferir en los procesos de extracción, fraccionamiento y análisis.

A pesar de los grandes progresos que se han realizado en proteómica de plantas, hay que tener en cuenta una serie de limitaciones teóricas y prácticas. Los costos y los materiales necesarios para llevar a cabo este tipo de estudio son muy altos. Hay mucha esperanza de que en un futuro próximo será posible obtener datos de proteínas a gran escala. Sin embargo, la cantidad de datos que se puede esperar de la proteómica es mucho menor que la de otras tecnologías, como la genómica.

8.5. Metabolómica

La metabolómica aparece como una nueva herramienta funcional de la biología integral que contribuye a nuestra comprensión de las complejas interacciones moleculares en los sistemas vivos. Al ser una de las últimas «ómicas» aparecida, presenta todavía muchos campos metodológicos que se pueden y se deben mejorar antes de que aparezcan bases de datos que nos permitan integrar las concentraciones de metabolitos con el resto de recursos de la genómica. La metabolómica representa la progresión lógica desde los estudios a gran escala de micromatrices de cDNA y proteómica hasta el tratamiento en conjunto de un sistema biológico completo (biología de sistemas).

Los metabolitos de las plantas son importantes para muchas de las respuestas de resistencia al ataque de patógenos, en condiciones de estrés y

contribuyen al color, sabor y olor de flores y frutos. El fenotipo bioquímico de una planta es el resultado de la interacción entre el genotipo y el entorno. Por tanto, se necesita identificar y cuantificar los metabolitos para estudiar la dinámica del metaboloma y analizar los flujos metabólicos. El reto de la metabolómica reside en encontrar cambios en las cantidades de metabolitos que se puedan correlacionar con el estado fisiológico y de desarrollo de una célula, de un tejido o de un organismo.

La enorme diversidad bioquímica que existe en las plantas hace que los metabolitos distintos que pueden encontrarse en plantas se estimen en más de 200.000. Esto hace que una aproximación en gran escala al estudio de los metabolitos presentes en las plantas sea un tremendo reto. La necesidad de utilizar tecnologías diversas para caracterizar los metabolitos presentes en plantas refleja el grado de complejidad para abordar este tipo de estudio. La deducción de cómo es el contexto biológico de una planta dada a partir de datos obtenidos por la medida de las concentraciones de metabolitos requiere, por tanto, que la identificación y cuantificación de los mismos se realice con gran fiabilidad. El puente necesario que debe existir entre la metabolómica y las otras aproximaciones de la genética funcional: transcriptómica y proteómica, permitirá el desarrollo de bases de datos que almacenen, integren, relacionen y permitan establecer relaciones causales entre genes, transcritos, proteínas y metabolitos.

Tarea 8-6

Busque revisiones y artículos que hablen de las diferentes rutas metabólicas en plantas: almidón, sacarosa, antocianinas, flavonas, carotenos, alcaloides, terpenoides o compuestos fenólicos.

Para obtener un conocimiento global de cómo funciona un sistema biológico es esencial saber cómo responden los tres niveles de expresión. Sólo así será posible deducir asociaciones relevantes entre macromoléculas, identificar correlaciones funcionales con fenotipos y construir modelos que describan de forma cuantitativa la dinámica de un sistema biológico. Estos análisis fenotípicos a gran escala permitirán en un futuro predecir como será el comportamiento de una planta dada en un determinado contexto medioambiental.

La metabolómica es el análisis del metabolismo desde una perspectiva global. Es como ver un mapa de metro desde arriba (Fig. 9-1).



Figura 8-1. Ciudad de Berlín, Alemania, con sus líneas de metro.

Figura 8-2. Ejemplo de rutas metabólicas en hígado.

La metabolómica trata de medir todos los metabolitos que están presentes en un tejido, y de esta forma hacer inferencias simples sobre las propiedades complejas de un sistema de redes entrelazadas. Por ejemplo, de cómo están reguladas las distintas rutas metabólicas. La **metabolómica estructural** se concentra en elucidar todas las rutas metabólicas existentes, tomando en cuenta todos los pasos químicos, los procesos enzimáticos y los genes que están involucrados. La **metabolómica funcional** trata de entender como están coordinadas las diversas rutas bioquímicas por medio de la regulación individual o coordinada de enzimas clave.

Para hacer metabolómica, es muy importante entender todas las técnicas analíticas que se usan en los laboratorios para medir metabolitos cuantitativamente. En los siguientes capítulos se explican algunas metodologías y principios de manera ilustrativa y breve.

8.6. Del genotipo al fenotipo

Hoy en día, uno de los temas más importantes de las ciencias biológicas es determinar la **relación que existe entre los genes y sus funciones específicas**. Es decir, entender cómo se llega **del genotipo al fenotipo** y

viceversa. La forma de comprender este vínculo ha cambiado considerablemente en los últimos tres siglos.

En el siglo XIX, en la época de Gregor Mendel y Charles Darwin, el vínculo entre el genotipo y el fenotipo se postulaba sin saber nada sobre el DNA. Parecía un modelo muy simple. Lo que se veía era el fenotipo, pero ya se intuía que había algo más llamado **gen**. Los genes eran unas instancias ocultas que se heredaban a los descendientes. Se sabía que existían dos genes para cada característica, y que uno veía de la madre, y el otro del padre. Cada gen podía tener diferentes variantes que se llamaban **alelos**. También se sabía que unos alelos eran **dominantes** sobre otros que eran **recesivos**. Las leyes de herencia de Mendel explicaban como segregaban las características fenotípicas en las generaciones filiales F1 y F2.

Tarea 8-7

Repase las leyes de herencia de Mendel. Haga un cuadro donde explique la segregación Mendeliana en la generación F2.

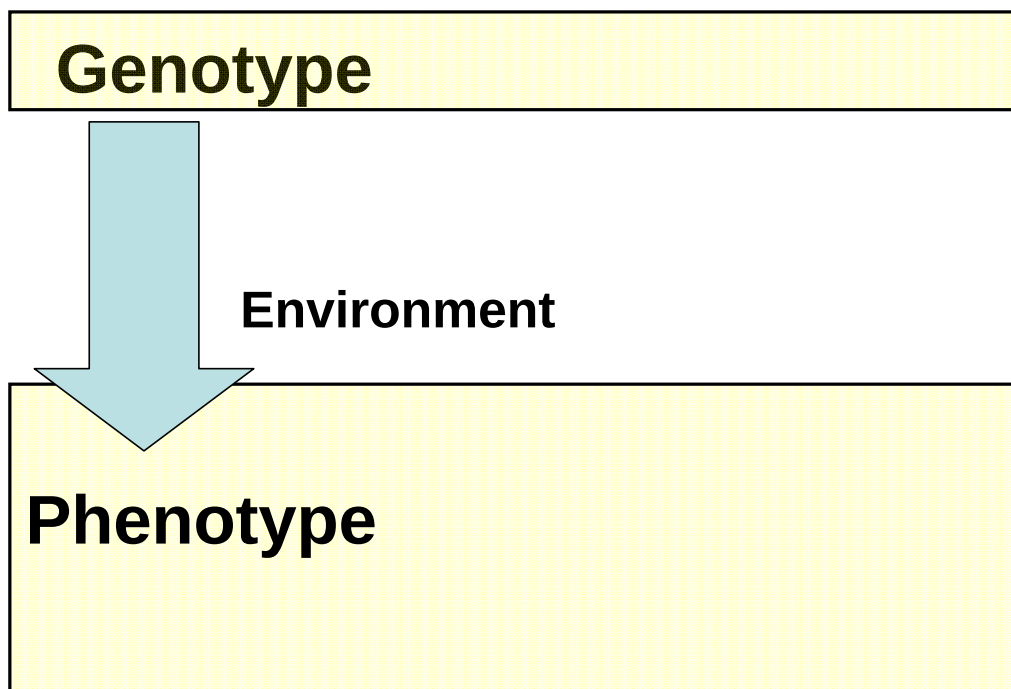


Figura 8-3. Vínculo entre el fenotipo y el genotipo como lo entendía Gregorio Mendel y otros científicos en el siglo XIX. Un fenotipo es causado por un gen. O bien, un gen se expresa en forma de un fenotipo determinado. La manera en que se heredaban esas características esta descrita por las leyes de la genética mendeliana. Esto aplicaba muy bien para el color de las flores o la forma de las semillas de chícharo (*Pisum sativum*). Se pensaba que el ambiente solo tenía una influencia en los casos de fenotipos más complejos.

Mas tarde, a mediados del siglo 20 se estudiaron las proteínas y el DNA. Se descubrió que el DNA era el portador de la información genética. También se encontró que el DNA se transcribía a RNA, y que este se traducía a una secuencia específica de amino ácidos según un **código genético universal**. Nació entonces el **dogma central de la biología molecular**, que postula que la información fluye en una sola dirección, del genotipo al fenotipo. Esto implicaba que no se pueden heredar las características adquiridas durante la vida. Otra forma de ver esto, es que el DNA es una molécula dictadora que determina todo lo que esta subordinado a nivel molecular.

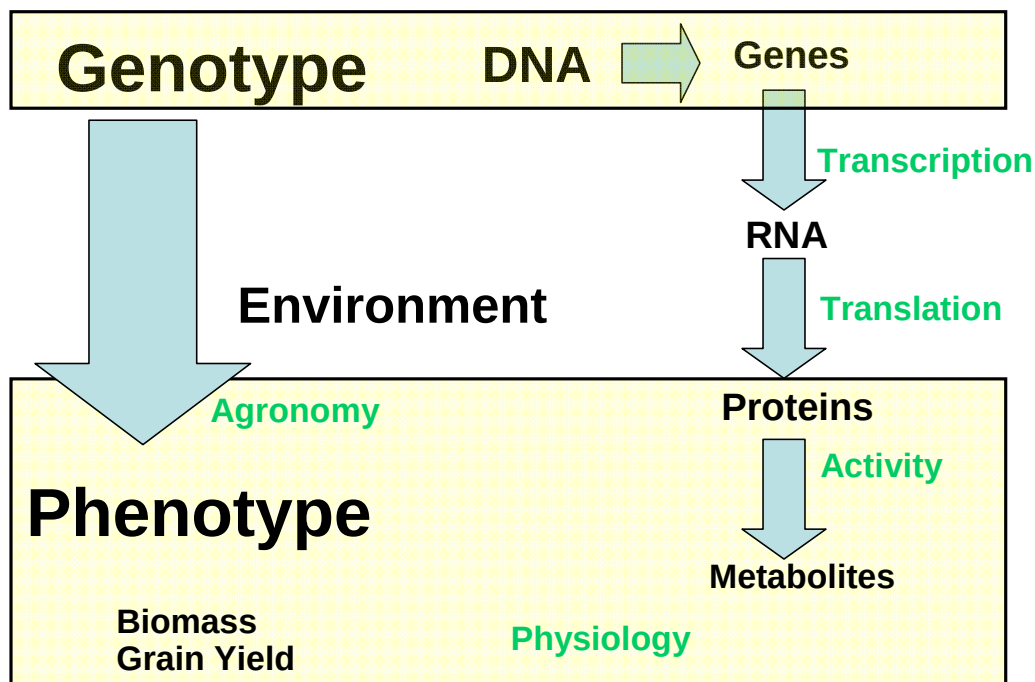


Figura 8-4. Vínculo entre el fenotipo y el genotipo como lo entendían los científicos en el siglo 20. El DNA se replica y se hereda a las siguientes generaciones. La información del DNA se transcribe primero al mRNA y después se traduce al idioma de las proteínas con 20 amino ácidos. Las proteínas pueden formar enzimas que catalizan reacciones bioquímicas para transformar metabolitos. Las enzimas funcionan dentro de rutas metabólicas que están compartimentalizadas en la célula vegetal. Por medio de la fotosíntesis se fija el dióxido de carbono y se forman carbohidratos como sacarosa y almidón. Lo que finalmente cosechamos en forma de biomasa vegetal es el resultado de la acumulación de metabolitos en ciertos órganos de la planta. El rendimiento depende entonces no solo de los genes, sino de la fisiología y la agronomía, y estas a su vez están influenciadas por el medio ambiente.

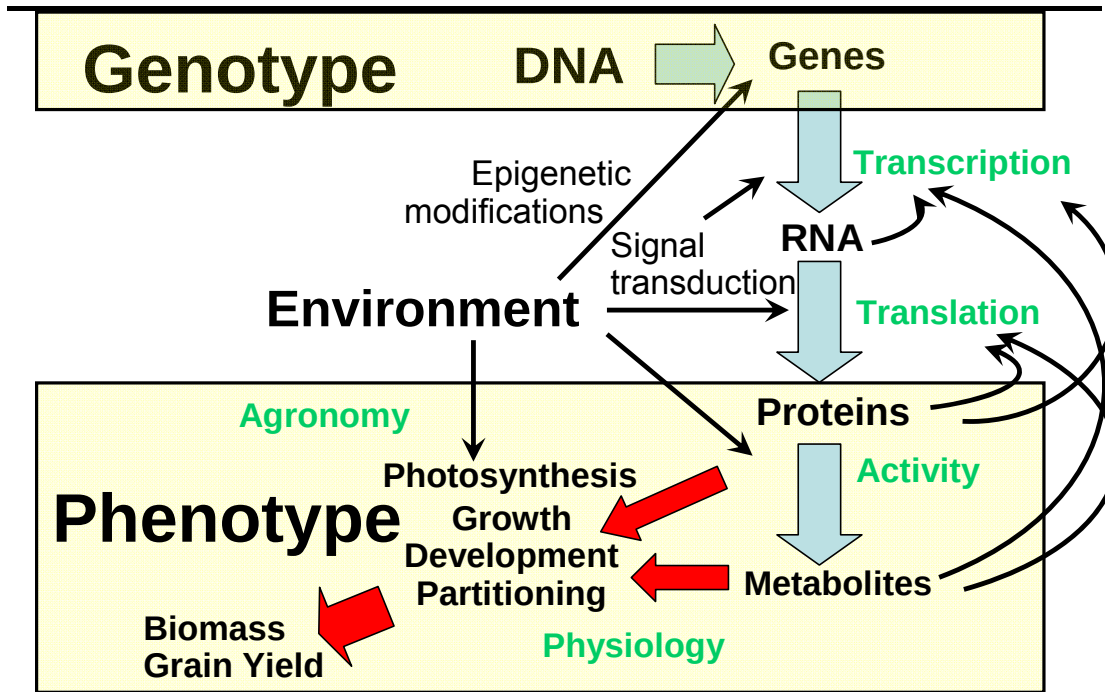


Figura 8-5. Vínculo entre el fenotipo y el genotipo como lo están empezando a entender los científicos del siglo 21. El modelo simplista se ha venido haciendo cada vez más complejo. Ya no solo se conoce un flujo de información en una sola dirección, sino que también hay mucha retroalimentación. Por ejemplo, la epigenética demuestra que el DNA puede ser modificado por medio el ambiente. Los microRNA pueden afectar la expresión de los genes de manera retrograda (*gene silencing*). El flujo de información es bidireccional, ya que algunos RNAs se pueden convertir en DNA por medio de la enzima transcriptasa reversa (RT). El ambiente puede afectar los procesos moleculares en todos los niveles, desde la transcripción, hasta la actividad enzimática. Todo se ha mas vuelto mas complicado. Con este último modelo se puede entender porque el rendimiento de grano del maíz no es un simple reflejo de los genes, y por consiguiente, el mejoramiento molecular no puede ser tan simple como solo secuenciar el DNA a nivel genómico.

De la **genómica estructural** se está pasando cada vez más a la **genómica funcional**. La secuenciación de genomas completos es sólo el principio de una tarea más amplia dentro de la **biología de sistemas**.

9. Herramientas bioquímicas

9.1. Fundamentos de enzimología

9.1.1. ¿Qué es una reacción química?

Una reacción química es todo proceso molecular en el que una o más sustancias (reactantes) sufren transformaciones para convertirse en otras moléculas (productos). Las reacciones químicas por lo regular consisten en un reacomodo de enlaces covalentes entre diferentes átomos. Es decir, se trata de un reacomodo de los orbitales electrónicos. Por ejemplo, en la reacción $2\text{H}_2 + \text{O}_2 \rightarrow 2 \text{H}_2\text{O}$ se forman dos moléculas de agua a partir de dos moléculas de hidrógeno y una molécula de oxígeno. En esta reacción se rompen enlaces y se distribuyen los electrones de una forma diferente. A las representaciones simbólicas de las reacciones se les llama ecuaciones químicas. Una ecuación química indica la forma y los pasos en que se puede llevar a cabo una posible transformación de las moléculas. Las reacciones se dan bajo ciertas premisas y leyes que pueden ser cualitativas o cuantitativas.

Tarea 9-1

Para entender las reacciones químicas es importante definir muy bien ciertos conceptos básicos. Investigue en un libro avanzado de química los siguientes conceptos y haga un breve resumen.

- La conservación de las masas
- La conservación de las cargas eléctricas
- El concepto de molaridad de *Avogadro*
- El concepto de equilibrio químico de *Le Chatelier*
- La reversibilidad de una reacción
- La entropía y la entalpía de una reacción

Tarea 9-2

¿Cómo aplican los conceptos anteriores para la biología vegetal? De ejemplos usando las reacciones de la fotosíntesis en las hojas de las plantas.

Las ecuaciones químicas se pueden clasificar en distintas categorías:

- Reacción de síntesis: Elementos o compuestos sencillos se unen para formar un compuesto más complejo. $\text{A} + \text{B} \rightarrow \text{AB}$
- Reacción de descomposición: Un compuesto se fragmenta en elementos o compuestos más sencillos. $\text{AB} \rightarrow \text{A} + \text{B}$
- Reacción de desplazamiento simple o sustitución simple: Un elemento reemplaza a otro en un compuesto. $\text{A} + \text{BC} \rightarrow \text{AC} + \text{B}$. En su explicación sencilla, es cuando alguien le roba la pareja a otro.

- **Reacción de doble desplazamiento o doble sustitución:** Los iones en un compuesto cambian lugares con los iones de otro compuesto para formar dos sustancias diferentes. $AB + CD \rightarrow AD + CB$. Para un entendimiento práctico es el intercambio de parejas.

Los dos tipos fundamentales de reacciones químicas son: las reacciones del tipo ácido-base (**neutralización y disociación**) y las reacciones del tipo redox (**oxidación o reducción**).

Reacciones ácido-base	Reacciones redox
Se presentan cuando se intercambian protones, sin cambios en los estados de oxidación en las moléculas. Por ejemplo la reacción $H_2O \rightarrow H^+ + OH^-$ consiste en un reacomodo de los protones, sin alterar el estado de oxidación de cada átomo. La reacción $HCl + NaOH \rightarrow NaCl + H_2O$ es otro ejemplo, y en este caso se llama reacción de neutralización. Las reacciones ácido-base pueden estar acompañadas por cambios en el pH, pero no necesariamente siempre es así. Por ejemplo, la formación de un enlace peptídico o la hidrólisis de un enlace glicosídico son reacciones del tipo ácido-base en las que no cambia el pH.	Se dan cuando hay cambios en los estados de oxidación de los átomos. Por ejemplo en la reacción $2H_2 + O_2 \rightarrow 2 H_2O$ el hidrogeno pierde un electrón (se oxida) mientras que los átomos de oxígeno ganan dos electrones (se reducen). Otro ejemplo de una reacción redox es la formación de óxido de hierro producida al reaccionar el oxígeno del aire con el hierro. Las reacciones más comunes de oxidorreducción dentro de una célula viva son las que involucran los cofactores NADH o NADPH.

Desde un punto de vista general se pueden postular solamente esos dos grandes modelos para las reacciones químicas, que corresponden a un reacomodo ya sea de **electrones** o de **protones**. Bajo el concepto de Lewis, las reacciones de ácido-base involucra el reacomodo de orbitales ocupados y libres (orbitales con y sin electrones). Mientras que la reacciones redox involucran el movimiento de uno o más electrones entre orbitales de diferentes átomos.

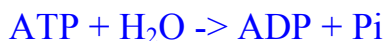
Las transformaciones químicas como la combustión, deshidratación, hidrólisis, neutralización, pueden ser clasificadas ya sea en reacciones ácido-base, reacciones redox o una combinación de ambas. Los procesos de solubilización, precipitación, evaporación, condensación más que ser reacciones químicas son procesos físicos.

Tarea 9-3

Clasifique las siguientes reacciones químicas:

**Tarea 9-4**

Dentro de una célula vegetal suceden algunas de las siguientes reacciones bioquímicas. Clasifíquelas según el tipo de reacción química:

**9.1.2. ¿Qué es un catalizador?**

Para que una reacción química se lleve a cabo, las moléculas necesitan pasar de un estado electrónico a otro. Este paso a veces requiere de un estado transitorio de más alta energía. Los catalizadores por definición son aquellas sustancias que de alguna manera intervienen en la formación de este estado transitorio. Los **catalizadores** son aquellas sustancias que estabilizan este estado transitorio y de esta forma aumentan la velocidad de reacción.

La estabilización del estado transitorio significa que se requiere menos energía de activación para llegar a ese estado. El modelo probabilístico de química (Boltzmann) dice que las moléculas tienen una amplia distribución de su energía cinética, por lo que a una determinada temperatura siempre hay un número de moléculas que tienen más o menos energía que el promedio. El promedio global de las moléculas depende directamente de la temperatura (constante de Boltzmann multiplicada por la temperatura = kT). Con una reducción en la energía de activación se logra que un gran número de moléculas tenga más energía que ese nivel mínimo. Esta es la razón por la que las reacciones se aceleran con un catalizador, ya que hay más moléculas de la población global que pueden reaccionar.

Los catalizadores se pueden clasificar en dos grandes grupos:

- **Catalizadores Inorgánicos:** son metales como el níquel y el platino. Se usan industrialmente y trabajan en condiciones como presión o temperatura extremas.
- **Catalizadores Biológicos:** son las enzimas formadas generalmente por proteínas. Son biomoléculas complejas que tienen diversos grupos funcionales y que están acomodados en una conformación tridimensional definida.

9.1.3. ¿Qué es una enzima?

Son los catalizadores biológicos por excelencia, proteínas que al igual que el resto se constiuyen con los 20 aminoácidos conocidos, pero presentan regiones características. También existen otros catalizadores biológicos que a diferencia de las enzimas están constituidos por ácidos ribonucleicos y por lo tanto se denominan ribozimas, como el fragmento 16S de los ribosomas en el que reside la actividad peptidil transferasa. Las principales características de ambos son:

- Disminuyen la energía de activación
- Se unen y modifican la estructura del sustrato
- Se requieren en pequeñas cantidades
- Son altamente específicas
- Son regulables. Es decir, son susceptibles de ser activadas o inhibidas
- Intercambian diferentes formas de energía (cinética, química, vibracional, etc)
- Se desnaturalizan bajo condiciones extremas como una temperatura de 60°C o un pH de 1.

Las enzimas pueden formar complejos con otras proteínas o factores. A esos complejos funcionales (enzima-cofactor) se les llama **holoenzimas**. Algunos de los principales componentes de las holoenzimas son:

- Coenzimas: Son moléculas orgánicas tales como CoA, NAD, NADP, ATP y ADP.
- Grupos prostéticos: son cofactores complejos que pueden unir metales como: Fe, Cu, Mg, Ca, Zn y Mo.

9.1.4. Características generales de una reacción enzimática

Como todos los catalizadores, las enzimas funcionan disminuyendo la energía de activación de una reacción y acelera sustancialmente la velocidad la velocidad de conversión química. Las enzimas no alteran el balance energético de las reacciones en que intervienen, ni modifican, por lo tanto, el equilibrio de la reacción (ΔG), pero consiguen acelerar el proceso, incluso millones de veces (ver **Figura 9-1**).

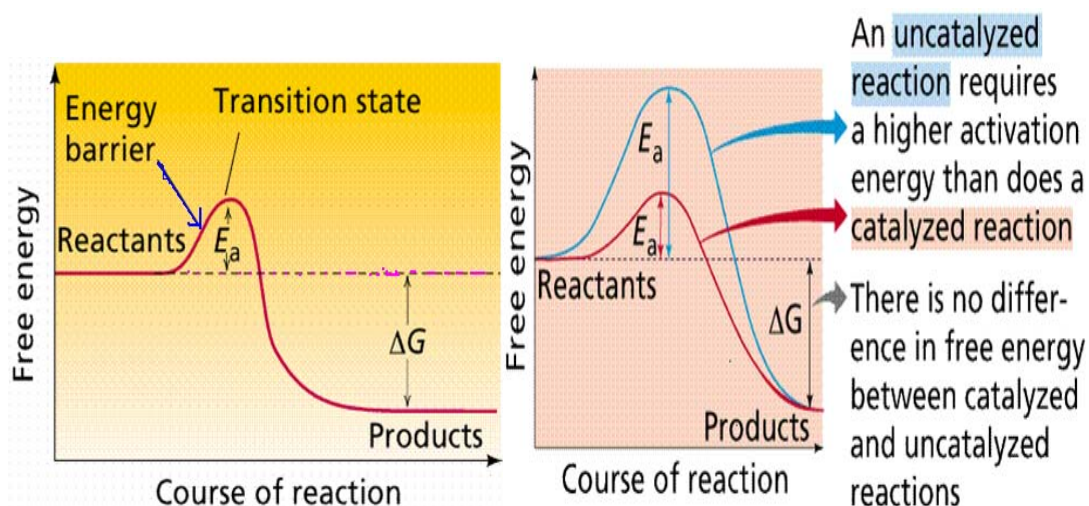


Figura 9-1. En la imagen de la izquierda se muestra como varía la energía libre en una reacción química cuando se pasa de reactantes a productos. En la imagen de la derecha se muestra la comparación de una reacción llevada a cabo por una enzima donde la diferencia de energía libre es menor, debido a que las enzimas disminuyen la energía de activación.

Una reacción que se produce bajo el control de una enzima, o de un catalizador en general, alcanza el equilibrio mucho más deprisa que la correspondiente reacción no catalizada. Por ejemplo, la descomposición del agua oxigenada (H_2O_2) para dar H_2O y O_2 puede ocurrir sin catalizador, con un catalizador inorgánico (platino), o con una enzima específica (catalasa). Las respectivas energías de activación (E_a) para cada proceso son 18, 12 y 6 K cal/mol. Así, se puede calcular que el platino acelera la reacción 20,000 veces, mientras que la catalasa la acelera 370,000 veces. Al igual que ocurre con otros catalizadores, las enzimas no son consumidas por las reacciones que ellas catalizan, ni alteran su equilibrio químico, sin embargo, difieren de otros catalizadores por ser más específicas. Las enzimas pueden catalizar alrededor de 4,000 reacciones bioquímicas distintas.

9.1.5. ¿Cómo funcionan las enzimas?

Prácticamente todas las reacciones químicas que tienen lugar en los seres vivos están catalizadas por enzimas. Estas llevan a cabo reacciones específicas y casi siempre actúan sobre un único sustrato o sobre un grupo muy reducido de ellos.

El mecanismo catalítico de una enzima comienza cuando el sustrato se une a una región concreta de la enzima, llamada centro activo. El **centro activo** comprende un sitio de unión formado por los aminoácidos que están directamente implicados en el mecanismo de la reacción estando en contacto directo el sustrato y un sitio catalítico. Como se sabe, para que una reacción ocurra, tiene que existir un choque efectivo entre las moléculas. Estos choques dependen de la naturaleza de los reactivos, de su

concentración y de la orientación de las moléculas cuando ocurre el encuentro. La actuación del enzima permite que los reactantes (sustratos) se unan a su centro activo con una orientación óptima para que la reacción se produzca. La unión al centro activo modifica las propiedades químicas del sustrato, debilitando los enlaces existentes y facilitando la formación de otros nuevos.

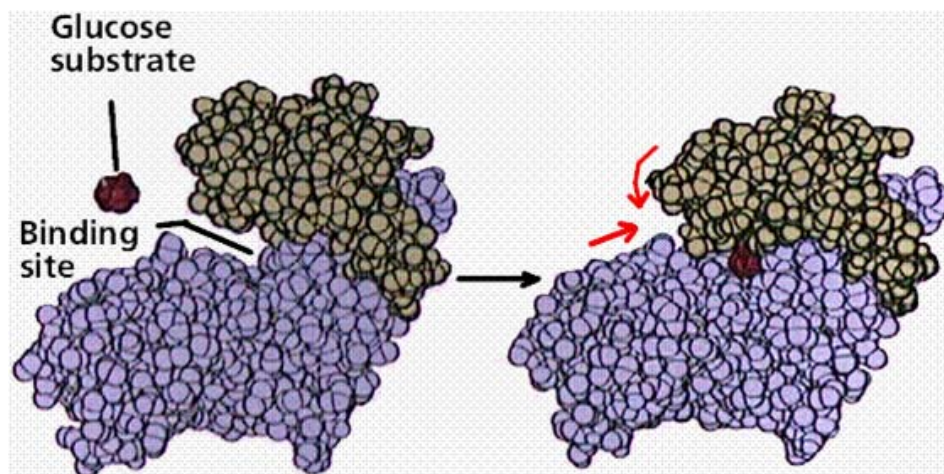


Figura 9-2. Ejemplo de la transformación del sustrato en el sitio activo de una enzima hacia la formación de producto. La enzima no permanece estática, sino que se mueve al unirse al primer sustrato. En el caso de la hexocinasa, la unión a glucosa hace que la enzima se cierre como una tijera, lo que crea un nuevo sitio para unión del ATP.

Los productos obtenidos a partir de ciertos tipos de reactivos dependen de las condiciones bajo las que se da la reacción química. No obstante, tras un estudio cuidadoso se comprueba que, aunque los productos pueden variar según cambien las condiciones, determinadas cantidades permanecen constantes en cualquier reacción química incluyendo, el número de átomos presentes, la carga eléctrica y la masa total.

La cantidad de producto que se suele obtener de una reacción química de forma práctica, es menor que la cantidad teórica. El rendimiento de una reacción depende de varios factores, como la pureza del reactivo y las reacciones secundarias que puedan tener lugar. Cuando uno de los reactivos esté en exceso, el rendimiento deberá calcularse respecto al reactivo limitante. En este contexto es importante recordar el concepto de equilibrio químico de *Le Chatelier*.

Tarea 9-5

Considere la reacción $AB + C \rightleftharpoons AC + B$

Usando el principio de *Le Chatelier*, ¿Qué debe agregar o quitar para desplazar la reacción hacia la derecha? ¿Qué debe hacer para maximizar la cantidad de producto AC?

Compare sus respuestas con otros compañeros y discútalos con su maestro de clases.

Tarea 9-6

Entre en Internet a <http://reacciones.colegiosandiego.com/ejercicios.html>, y realice los ejercicios, prácticas y juegos marcados ahí. Compare sus resultados con los compañeros. Corrija sus errores con la literatura especializada

9.1.6. Nomenclatura de las enzimas

Hay varias formas de asignarle un nombre a un enzima: nombre particular, nombre sistemático, código de la comisión enzimática (*enzyme comisión* o EC). Antiguamente, las enzimas recibían nombres particulares, asignados por su descubridor. También se les daba un nombre según su origen, por ejemplo, la papaina es una enzima que se extrae de la papaya. Al ir aumentando el número de enzimas conocidos, se hizo necesaria una nomenclatura sistemática que informara sobre la acción específica de cada enzima y los sustratos sobre los que actuaba.

El nombre sistemático de un enzima consta actualmente de 3 partes: el sustrato preferente, el tipo de reacción realizada terminación "asa". La mayoría de las enzimas se clasifican dependiendo la actividad que realicen sobre el sustrato.

Tipo de reacción	Ejemplos
Hidrolasa	Lipasa, Proteasa, Amilasa
Liasa	Aldolasa, Phenylamonia liasa
Oxido-reductasa	Alcohol deshidrogenasa
Transferasa	Hexocinasa
Isomerasa	G-6-P-isomerasa
Ligasa	Citrato sintasa

Muchas enzimas catalizan reacciones reversibles. No hay una manera única para fijar cual de los dos sentidos se utiliza para nombrar al enzima. La glucosa fosfato isomerasa cataliza la isomerización de la glucosa-6-fosfato en fructosa-6-fosfato. Así, la glucosa-6-fosfato isomerasa también podría llamarse fructosa-6-fosfato isomerasa. Cuando la acción típica del enzima es la hidrólisis del sustrato, el segundo componente del nombre se omite y por ejemplo, la lactosa hidrolasa se llama simplemente lactasa. Además de los nombres sistemáticos, aún persisten otros consagrados por el uso. Así, la glucosa-ATP fosforiltransferasa se llama habitualmente glucocinasa.

El nombre de cada enzima puede ser identificado por un código numérico, encabezado por las letras EC (*enzyme commission*), seguidas de cuatro números separados por puntos. El primer número indica a cual de las seis clases pertenece el enzima, el segundo se refiere a distintas subclases dentro de cada grupo, el tercero y el cuarto se refieren a los grupos químicos específicos que intervienen en la reacción. Así, la ATP:glucosa fosfotransferasa (glucocinasa) se define como EC 2.7.1.2. El número 2 indica que es una transferasa, el 7 que es una fosfotransferasa, el 1 indica que el aceptor es un grupo OH, y el último 2 indica que es un OH de la D-glucosa el que acepta el grupo fosfato.

En función de su acción catalítica específica, las enzimas se clasifican en 6 grandes grupos o clases:

Clase 1: OXIDORREDUCTASAS

Clase 2: TRANSFERASAS

Clase 3: HIDROLASAS

Clase 4: LIASAS

Clase 5: ISOMERASAS

Clase 6: LIGASAS

Como se mencionó anteriormente, una de las características de las enzimas es la cantidad de enzima, por ello es importante cuantificarla. La cantidad de enzima necesaria para transformar un mol de sustrato en un minuto se llama Unidad de enzima. La actividad enzimática también se puede medir en catales, que es la cantidad de enzima necesaria para transformar un mol de sustrato en un segundo. (Stryer,. 2007).

9.1.7. Regulación de la actividad enzimática

Una molécula de enzima no tiene por qué actuar siempre a la misma velocidad. Su actividad puede estar modulada por: pH, temperatura, presencia de cofactores, concentraciones del sustrato y productos finales, presencia de inhibidores, modulación alostérica, modificación covalente y

activación o inactivación por proteólisis. A continuación se explican algunos ejemplos de los tipos de regulación enzimática.

9.1.7.1. Efecto del pH sobre la actividad enzimática

Los enzimas poseen grupos químicos ionizables (carboxilos $-\text{COOH}$; aminos $-\text{NH}_2$; tioles $-\text{SH}$; imidazol, fenol, etc.). Estos representan algunos de los grupos funcionales de las cadenas laterales de sus aminoácidos. Según el pH del medio, estos grupos pueden tener carga eléctrica positiva, negativa o neutra. Como la conformación de las proteínas depende, en parte, de sus cargas eléctricas, habrá un pH en el cual la conformación será la más adecuada para la actividad catalítica. Este es el llamado pH óptimo. Es decir, la mayoría de las enzimas son sensibles a la concentración de protones en la solución, por lo que desviaciones por encima o por debajo del pH óptimo pueden afectar su actividad. La influencia del pH se da por un lado por la estructura terciaria de la proteína, y por otro lado por el mecanismo de reacción que a veces requiere protones como catalizadores. La mayoría de las enzimas celulares funcionan en un rango de pH de 6 a 8. Sin embargo, existen enzimas, como la pepsina estomacal, que tienen un pH óptimo de 2. Un pH extremo puede detener la actividad enzimática al provocar la desnaturalización de la estructura proteica rompiendo enlaces iónicos y puentes de hidrógeno.

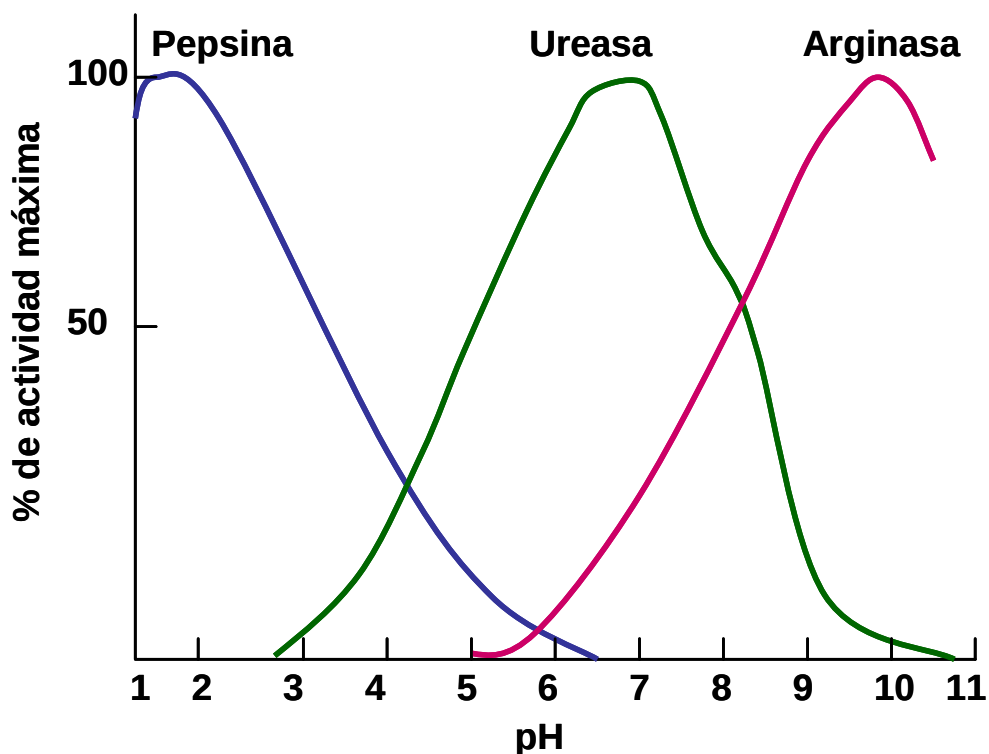


Figura 9-3. Gráfica del efecto del pH sobre la estructura de las proteínas. En los tres ejemplos se observa que el máximo de actividad enzimática (pH óptimo) se encuentra a diferentes valores de pH que representa que en el

sitio activo participan diferentes aminoácidos con grupos ionizables distintos.

9.1.7.2.Efecto de la temperatura sobre la actividad enzimática

En general, los aumentos de temperatura aceleran las reacciones químicas. La regla clásica de bioquímica dice que por cada 10 °C de incremento, la velocidad de reacción se duplica. La mayoría de las reacciones catalizadas por enzimas siguen esta ley general. Sin embargo, al ser proteínas, a partir de cierta temperatura, se empiezan a desnaturalizar por el calor. La temperatura a la cual la actividad catalítica es máxima, se llama temperatura óptima. Por ejemplo, las enzimas humanas se encuentran generalmente entre los 35 y los 40 °C, siendo 37 °C la temperatura media de las células humanas. La mayoría de las enzimas comienzan a desnaturalizarse rápidamente a partir de los 40 °C. Esto se debe a alteraciones de la estructura de la proteína debidas a la interrupción de uniones iónicas que resultan en la estabilización de la estructura tridimensional de la enzima. En cambio, las enzimas de las arqueas termófilas, que se encuentran en manantiales calientes, son estables hasta a 100 °C. Sin embargo, la idea de una temperatura "óptima" para una reacción enzimática puede ser engañosa, ya que la velocidad observada es el producto de la velocidad de la reacción química y la concentración de la enzima activa sin desnaturalizarse. La **Figura 9-4** muestra que por encima de esta temperatura, el aumento de velocidad de la reacción debido a la temperatura, es contrarrestado por la pérdida de actividad catalítica debida a la desnaturalización térmica y la actividad enzimática decrece rápidamente hasta anularse.

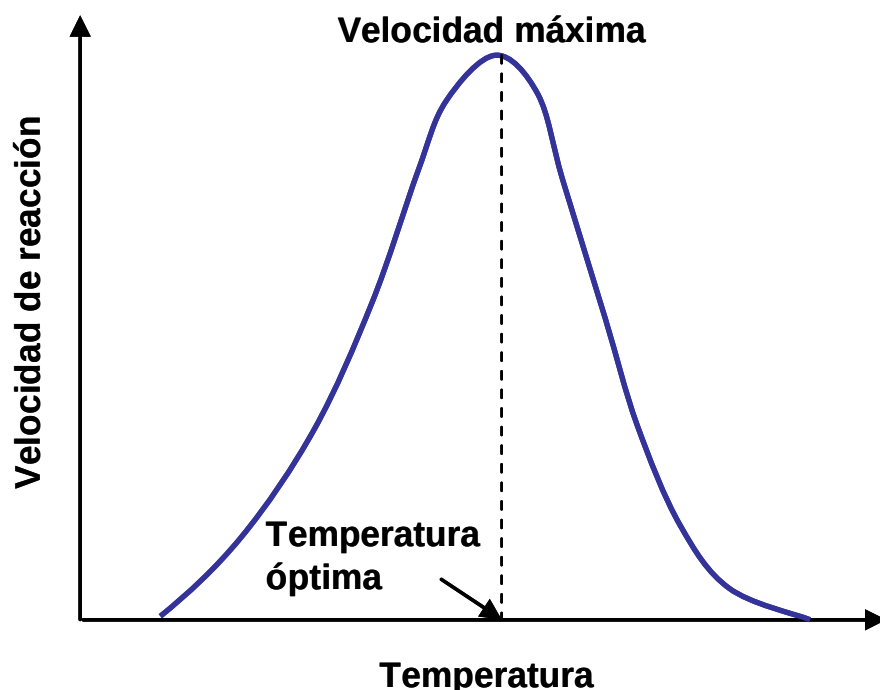


Figura 9-4. Gráfica que explica del efecto de la temperatura sobre la actividad enzimática. La velocidad aumenta a medida de que se incrementa la temperatura hasta alcanzar la temperatura óptima. Si la temperatura aumenta por encima de este punto, la enzima pierde su actividad debido al daño en la estructura tridimensional.

Tarea 9-7

Los cultivos vegetales no se dan igual de bien en todos los climas. Hay unos que crecen bien a ciertas temperaturas pero no a otras. El maíz no se crece óptimamente a las mismas temperaturas a las que se cultiva generalmente la papa. Haga una lista con más ejemplos. Postule algunas hipótesis de por que esto es así. Relacione los efectos fisiológicos conocidos a nivel de la productividad de las plantas, con los efectos que ha aprendido a nivel molecular de las células vegetales.

9.1.7.3. Efecto de los cofactores en la actividad enzimática

A veces, una enzima requiere para su función la presencia de moléculas adicionales para llevar a cabo su función. Si dicha molécula es inorgánica, como Fe^{2+} , Mg^{2+} , Mn^{2+} , Zn^{2+} , se denomina **cofactor**. En cambio, si se trata de una molécula orgánica se denomina **coenzima**. Casi un tercio de las enzimas conocidas requieren cofactores. En la Figura 9-5 podemos observar una molécula de mioglobina (proteína que transporta oxígeno al tejido muscular) y su coenzima (el grupo hemo, representado en color verde). Cuando los cofactores o las coenzimas se encuentran unidos covalentemente a la enzima se llaman grupos prostéticos. La enzima sólo es activa si esta unida al grupo prostético, lo que recibe el nombre de

holoenzima. La parte proteica sin grupo prostético es inactiva y recibe el nombre de **apoenzima**, de forma que: apoenzima + grupo prostético= holoenzima (ver Figura 9-5)

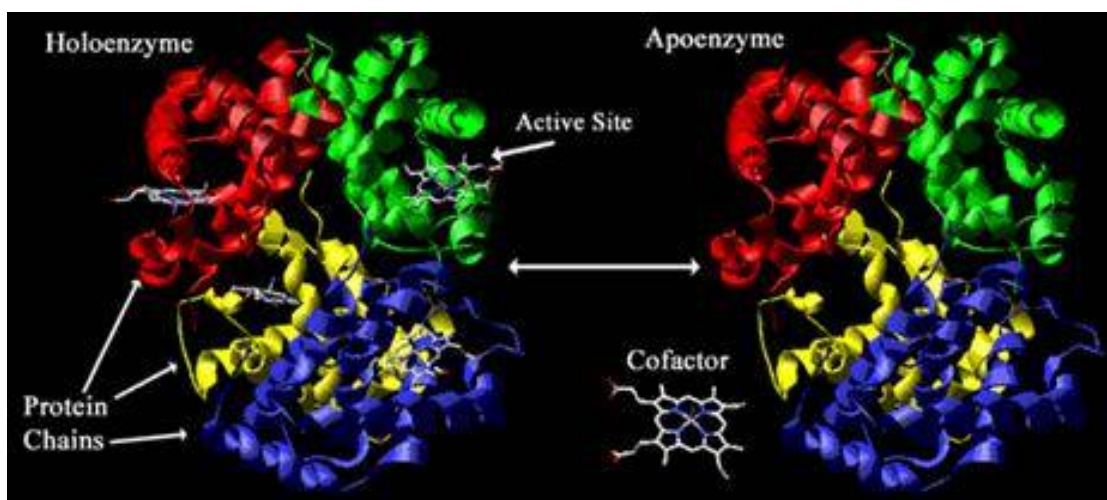


Figura 9-5. Imagen de la estructura cuaternaria de una enzima mostrando los cambios estructurales al unirse un cofactor y el sustrato en el sitio activo de pasar de holoenzima a apoenzima.

9.1.7.4. Efecto de la concentración de sustrato sobre la actividad enzimática

La velocidad de una reacción enzimática depende de la concentración de sustrato. Si aumenta la concentración de sustrato aumenta la velocidad de la reacción o actividad enzimática. Sin embargo, un grado extremo de saturación también puede limitar la velocidad máxima y hasta inhibirla. Una enzima está saturada cuando los sitios activos de todas las moléculas están ocupados la mayoría del tiempo. A partir del punto de saturación la reacción no puede acelerarse mediante adición de sustrato, sea cual sea la cantidad añadida. En un gráfico la velocidad de reacción habría alcanzado una meseta.

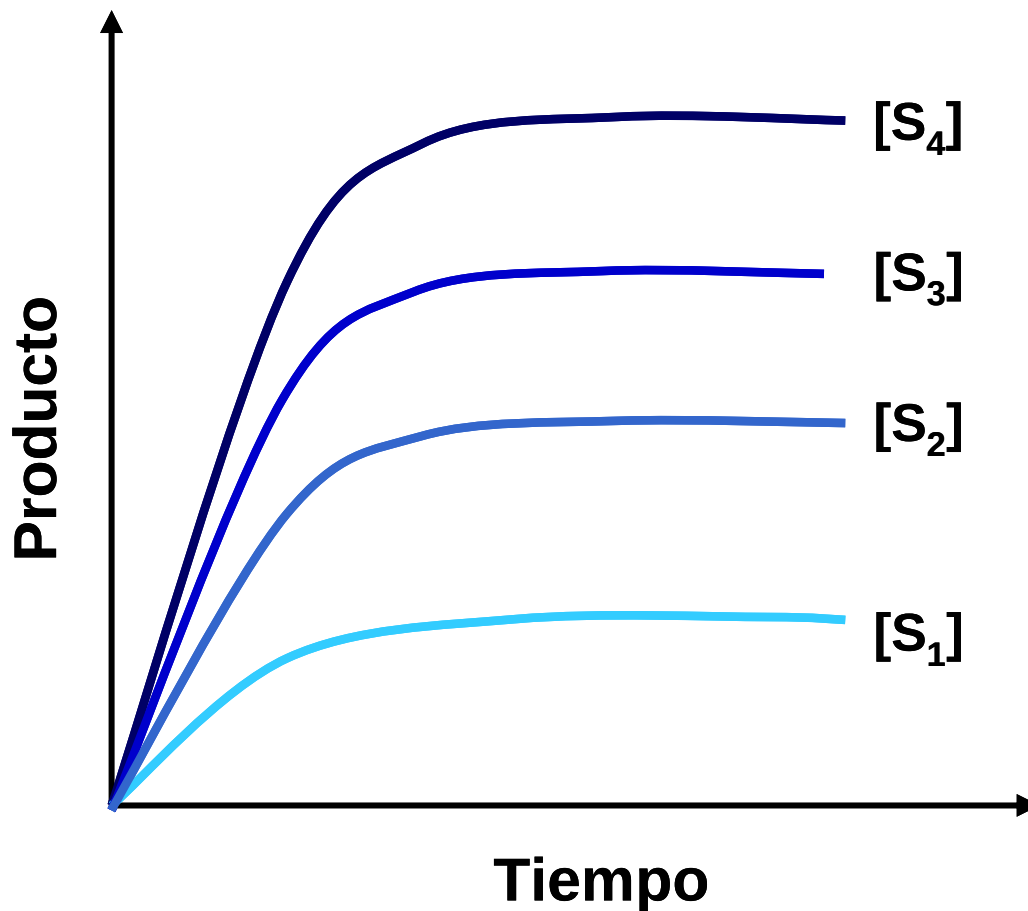


Figura 9-6. Gráfica de curso de reacción de una enzima a cuatro diferentes concentraciones de sustrato. Tanto la velocidad inicial, como el valor absoluto del producto final son distintos.

9.1.7.5. Efecto de la fuerza osmótica sobre la actividad enzimática

La mayoría de las enzimas no pueden tolerar concentraciones de sal extremadamente altas. Los iones de sodio interfieren con los enlaces débiles iónicos de las proteínas. Las enzimas típicas son activas en concentraciones salinas de 1 a 500 mM. Existen excepciones como las enzimas de organismos halófilos. Sin embargo, en la mayoría de las enzimas al encontrarse a una concentración alta de sal se debilitan las fuerzas básicas que estabilizan a la estructura y por lo tanto su función se pierde. Por lo regular los iones de sodio (Na^+) son más dañinos para las enzimas citosólicas que los iones de potasio (K^+).

Tarea 9-8

Investigue porque los cultivos como el maíz no pueden ser regados con agua salina de mar, mientras que las palmeras de coco aguantan muy bien ese tratamiento. Haga un comparativo entre los efectos de los iones de cloruro, sodio, potasio, magnesio, fósforo y sulfato. Relacione los efectos fisiológicos a nivel de la planta, con algunos efectos a nivel molecular de las células vegetales.

9.1.8. Modelo cinético de Michaelis-Menten

Los estudios sistemáticos del efecto de la concentración inicial del sustrato sobre la actividad enzimática comenzaron a realizarse a finales del siglo XIX. Ya en 1882 se introdujo el concepto del complejo enzima-sustrato como intermediario del proceso de catálisis enzimática. En 1913, Leonor Michaelis y Maud Menten (ver Figura 9-7) desarrollaron esta teoría y propusieron una ecuación de velocidad, que explica el comportamiento cinético de las enzimas.



Figura 9-7. Leonor Michaelis (izquierda) y Maud Menten (derecha).

Para explicar la relación observada entre la velocidad inicial (V_0) y la concentración inicial de sustrato ($[S]_0$) Michaelis y Menten propusieron que las reacciones catalizadas enzimáticamente ocurren en dos etapas: En la primera etapa se forma el complejo enzima-sustrato y en la segunda, el complejo enzima-sustrato da lugar a la formación del producto, liberando el enzima libre.

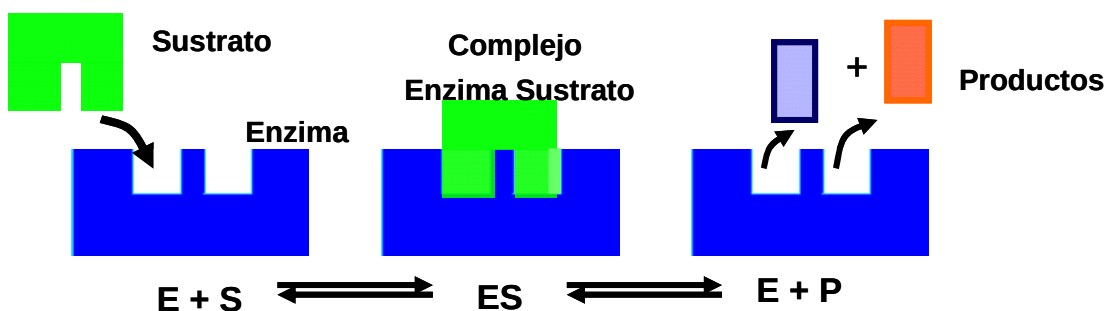


Figura 9-8. Representación de la unión del sustrato con la enzima y la formación de producto.

En las reacciones bioquímicas hay pasos intermedios que se llaman ES y EP, que se refieren al complejo de enzima (E) unido al sustrato (S) y al producto (P), respectivamente. Las flechas en las dos direcciones indican que la reacción es reversible en todos sus pasos. Es importante tener en cuenta que la enzima no se degrada, sino que se recicla durante la catálisis. Es decir, la suma de E + ES + EP se mantiene siempre constante. Lo que sí cambia después del tiempo, es la cantidad de E y de P, como en cualquier reacción en la que un sustrato se transforma en un producto. Sin embargo, la proporción de E y de P que finalmente se obtendrá depende del equilibrio químico y de las constantes de energía y no de la enzima. La enzima sólo determina la velocidad de la reacción y acelera el establecimiento del equilibrio químico.

La afinidad que existe entre la enzima y el sustrato, se caracteriza por su constante de Michaelis, K_m . El valor de K_m se refiere al valor de la concentración de S cuando la cantidad de enzima unida al sustrato (ES) o producto (EP) es igual a la cantidad de enzima libre (E). Es decir cuando $E=ES+EP$. La K_m es la concentración de sustrato con la que la enzima alcanza la mitad de su V_{max} , sus unidades se expresan en moles por litro ($\text{mol}\cdot\text{l}^{-1}$). Por su parte, la V_{max} es la velocidad máxima que puede alcanzar una enzima en su punto de saturación. Es decir cuando toda la enzima está unida al sustrato ($ES \gg E$) y casi no hay enzima libre ($E=0$). Las unidades de V_{max} se miden en micromoles por minuto ($\mu\text{moles}\cdot\text{min}^{-1}$).

Es importante saberse de memoria la ecuación de Michelis y Menten, ya que esta explica la grafica de saturación mostrada en la Figura 9-6.

$$v = \frac{[S] * V_{max}}{K_m + [S]}$$

La ecuación de Michelis-Menten explica la aparición de productos, el recambio del complejo enzima sustrato, la velocidad (v) y la afinidad de una enzima (K_m). La siguiente figura muestra la forma en que se deriva esta ecuación (ver Figura 9-9).

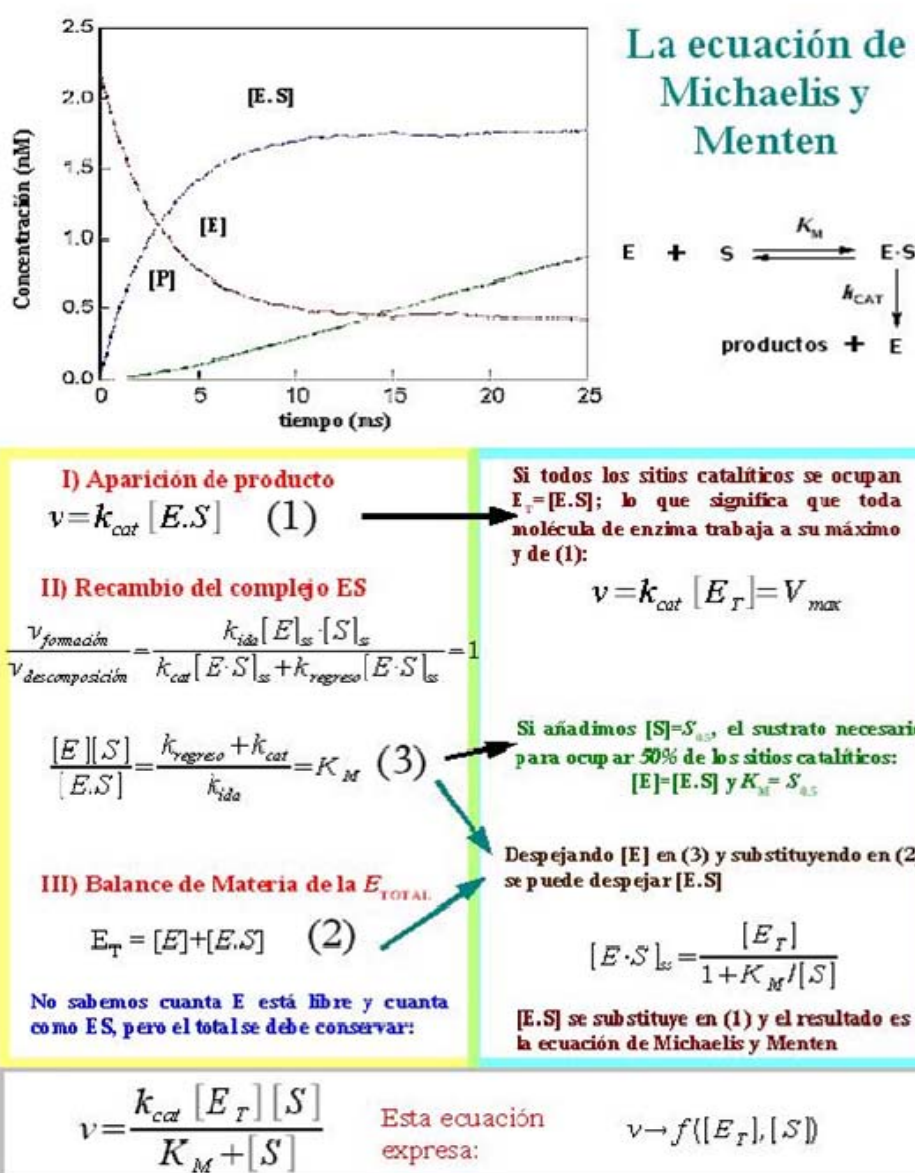


Figura 9-9. Forma de derivar la ecuación de Michaelis y Menten. Se puede distinguir entre enzima libre [E] y enzima unido al sustrato [ES], de forma que la concentración total de enzima, $[E_T]$, (que es constante a lo largo de la reacción) es:

$[E_T] = [E] + [ES]$. Como $[E] = [E_T] - [ES]$, resulta que: $v = k_1 [S] [E_T] - k_1 [S] [ES]$

Este modelo cinético La representación gráfica de la ecuación de Michaelis-Menten que se comporta como una hipérbola. La V_{max} corresponde al valor máximo al que tiende la curva experimental, y la K_m corresponde a la concentración de sustrato a la cual la velocidad de la reacción es la mitad de la V_{max} .

La derivación de Michaelis-Menten está descrita por Briggs y Haldane. Se obtiene de la siguiente manera: Se supone que la reacción enzimática es irreversible, y que el producto no se liga con la enzima después de la reacción.



Siguiendo la aproximación del estado estacionario, que señala que la concentración del complejo enzima-sustrato (ES) es pequeña y se mantiene casi constante a lo largo de la reacción enzimática:

$$\frac{\partial[ES]}{\partial t} = k_1[E][S] - k_{-1}[ES] - k_2[ES] = 0$$

$$[ES] = \frac{k_1[E][S]}{k_{-1} + k_2}$$

Se define:

$$K_m = \frac{k_{-1} + k_2}{k_1}$$

Entonces:

$$[ES] = \frac{[E][S]}{K_m} \quad 1$$

La velocidad de reacción es:

$$\frac{\partial[P]}{\partial t} = k_2[ES] \quad 2$$

La concentración total de la enzima:

$$[E_0] = [E] + [ES]$$

Por lo tanto:

$$[E] = [E_0] - [ES] \quad 3$$

Sustituyendo 3 en 1 da:

$$[ES] = \frac{([E_0] - [ES])[S]}{K_m}$$

Reordenando:

$$K_m [ES] + [S][ES] = [E_0][S]$$

$$[ES] = \frac{[E_0][S]}{K_m + [S]} \quad 4$$

Sustituyendo 4 en 2:

$$\frac{\partial[P]}{\partial t} = k_2[E_0] \frac{[S]}{K_m + [S]} = \frac{V_{\max}[S]}{K_m + [S]}$$

Representado gráficamente se obtiene una hipérbola que representa el cambio en la velocidad a medida que aumenta la concentración de sustrato como se ve en la siguiente Figura.

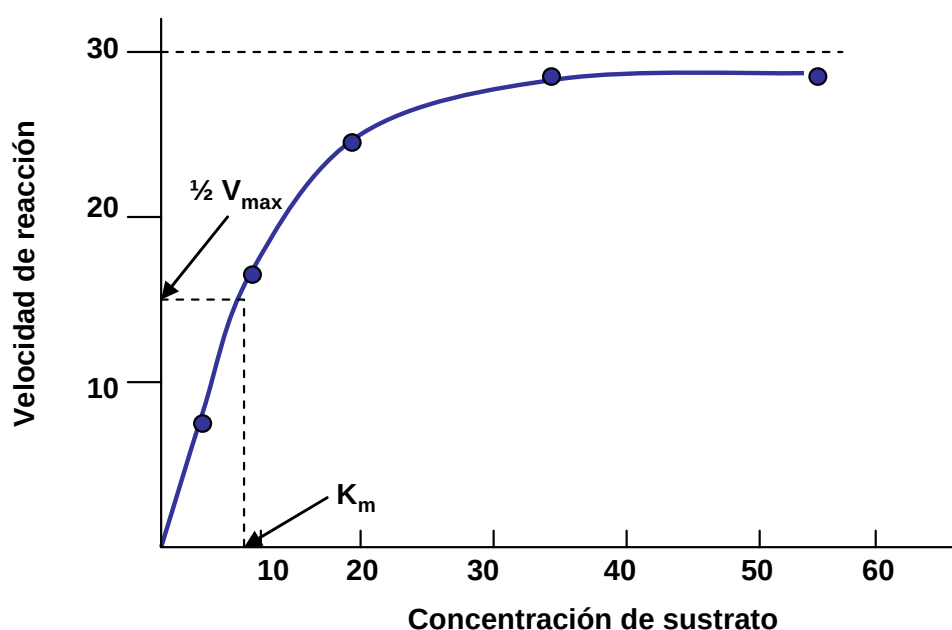


Figura 9-10. Representación gráfica de la curva de saturación de una enzima donde se puede observar como evoluciona la relación entre la concentración de sustrato y la velocidad de la reacción.

Esta ecuación puede ser analizada experimentalmente con un diagrama de Lineweaver-Burke o un diagrama de Eadie-Hofstee.

Para determinar gráficamente por medio Lineweaver-Burke de los valores de K_m y $V_{\max}$ es más sencillo utilizar la representación doble recíproca ($1/v_0$ frente a $1/[S_0]$), ya que es una línea recta. Esta representación doble recíproca recibe el nombre de representación de Lineweaver-Burk. Es una recta en la cual:

- La pendiente es $K_m/V_{\max}$.
- La abscisa en el origen ($1/V_0 = 0$) es $-1/K_m$.
- La ordenada en el origen ($1/[S_0] = 0$) es $1/V_{\max}$.

De esta forma, a partir de los datos experimentales se puede calcular gráficamente, los valores de K_m y $V_{\max}$ de un enzima para diversos sustratos.

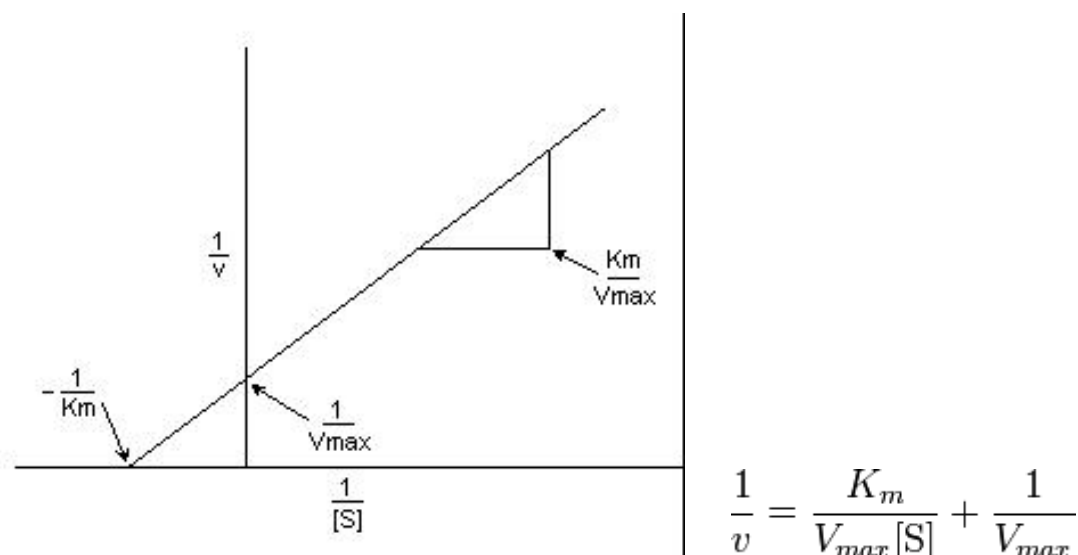


Figura 9-11. Gráfica de Lineweaver-Burke o de doble recíproco muestra el significado de los puntos de corte entre la recta y los ejes de coordenadas, así como el gradiente de velocidad.

9.1.8.1. Actividad enzimática

Se define la unidad de actividad enzimática (U), como la cantidad de enzima que cataliza la conversión de 1 μmol de sustrato en un minuto. La actividad específica es el número de unidades de enzima por miligramo de proteína (U/mg prot), por mililitro de disolución (U/ml) o por gramo de peso fresco (U/gFW).

9.1.8.2. Efectos alostéricos

Hay enzimas que no obedecen la ecuación de Michaelis-Menten. Se dice que su cinética no es Michaeliana. Esto ocurre con las enzimas alostéricas, cuya gráfica de la velocidad frente a la concentración de sustrato [S] no es una hipérbola, sino una sigmoidea. En la cinética sigmoidea, pequeñas variaciones en la [S] en una zona crítica (cercana a la K_m) se traduce en grandes variaciones en la velocidad de reacción.

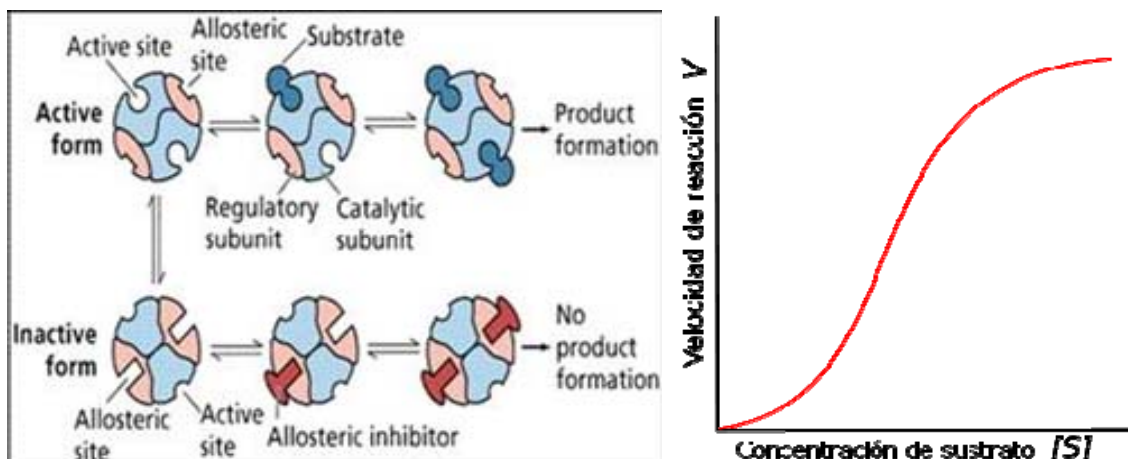


Figura 9-12. Ejemplificación de cómo se realiza la inhibición alostérica (izquierda). Curva de saturación de una enzima alostérica, donde se puede apreciar una cinética sigmoidea (derecha).

9.1.8.3. Tipos de inhibición de la actividad enzimática

Las enzimas catalizan, virtualmente, todos los procesos celulares, por lo que no es sorprendente, que los inhibidores enzimáticos se encuentren entre los agentes farmacéuticos y agroquímicos más importantes. Los herbicidas más potentes como los glifosatos, triazinas o fosfonotricinas, son agentes químicos que inhiben la actividad enzimática de ciertas proteínas clave para la sobrevivencia de la célula vegetal.

Tarea 9-9

Investigue cuáles son los agentes activos de algunos de los herbicidas comerciales como Basta, Roundup, Sanson, etc. Averigüe cuáles son las rutas metabólicas que inhiben estos compuestos. Dibuje las fórmulas químicas de estos herbicidas, y cómo estas moléculas se pueden unir al sitio activo de una enzima para inhibirla. Postule una hipótesis de por qué algunos herbicidas son selectivos para hierbas de hoja ancha (plantas dicotiledóneas) y otros son más efectivos para pastos (plantas monocotiledóneas). Use los conocimientos que aprendió sobre inhibición enzimática para explicarlo.

Un aspecto muy importante sobre la actividad de las enzimas, es su capacidad de activación o inhibición selectiva por otras moléculas o proteínas. Esta ocurre por los metabolitos que inhiben la actividad enzimática, ya sea contaminantes inespecíficos o inhibidores específicos. También se lleva a cabo por medio de proteínas que modifican las enzimas, por ejemplo por fosforilación de grupos hidroxilo o por modificaciones en los enlaces disulfuro entre diferentes residuos de cisteína (regulación redox). Esta regulación permite a las enzimas regular el flujo metabólico en una determinada vía.

Los inhibidores se pueden clasificar en varios grupos, entre ellos los del tipo reversible y los que causan una inhibición irreversible. Un tipo de inhibición reversible es la denominada competitiva. Un inhibidor competitivo compite con el sustrato por el sitio activo del enzima, pero la reacción normalmente no tiene lugar cuando se ha fijado el inhibidor. Los inhibidores competitivos son, frecuentemente, compuestos que se parecen al sustrato. Debido a que el inhibidor se une de manera reversible, se puede cambiar el sentido de la competición en favor del sustrato simplemente añadiendo más sustrato.

Se distinguen 3 tipos de inhibiciones:

- Competitiva: El inhibidor compite por el sitio activo de la enzima.
- No competitiva: El inhibidor se une con la enzima libre en otro sitio diferente al catalítico, afectando su velocidad de reacción.
- Acompetitiva: El inhibidor se une al complejo enzima-sustrato, mas no a la enzima libre.

A continuación se muestran los diferentes tipos de inhibición y como se pueden diferenciar de manera gráfica.

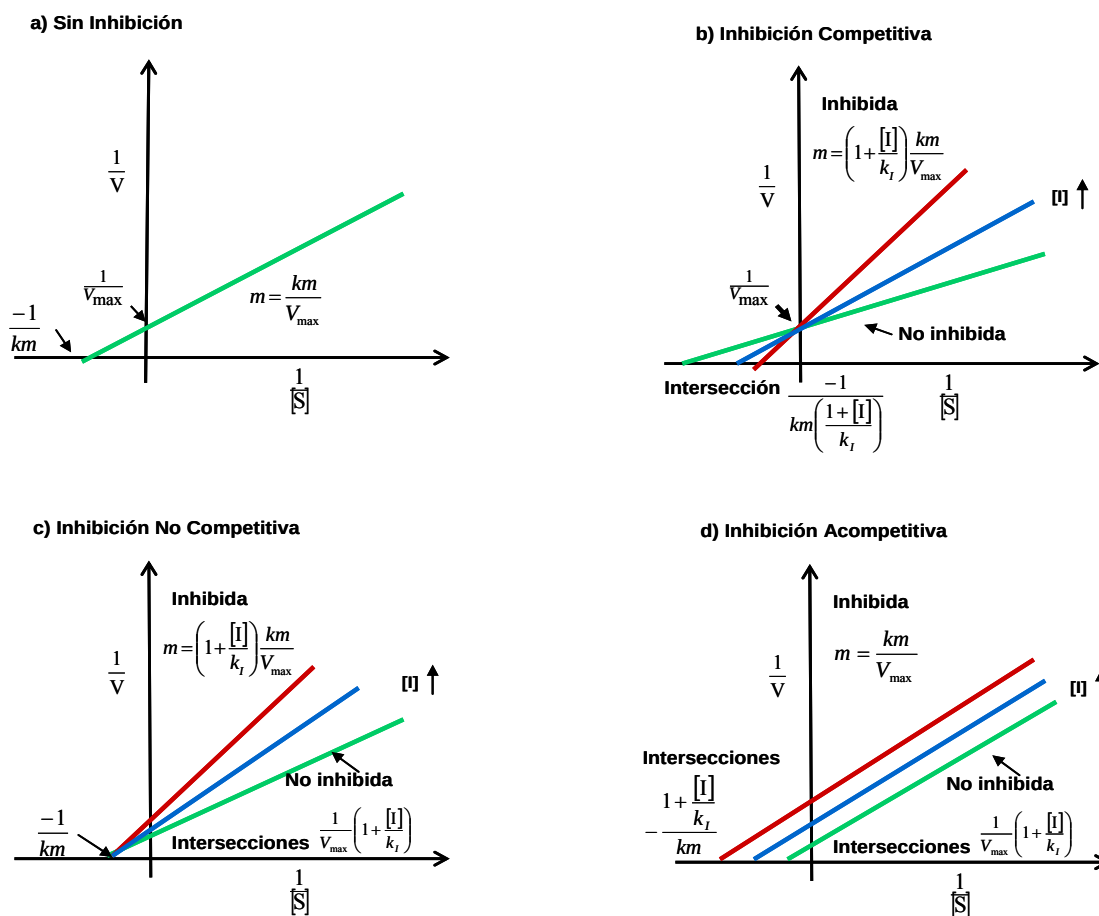


Figura 9-13. Gráfica comparativas de los inversos del sustrato y la velocidad para los 3 tipos de inhibición; a). Enzima no inhibida; b) inhibición competitiva; c) inhibición no competitiva y d) inhibición acompetitiva.

En la Figura 9-14 se muestra en forma tridimensional una cinasa que es inhibida por ATP.

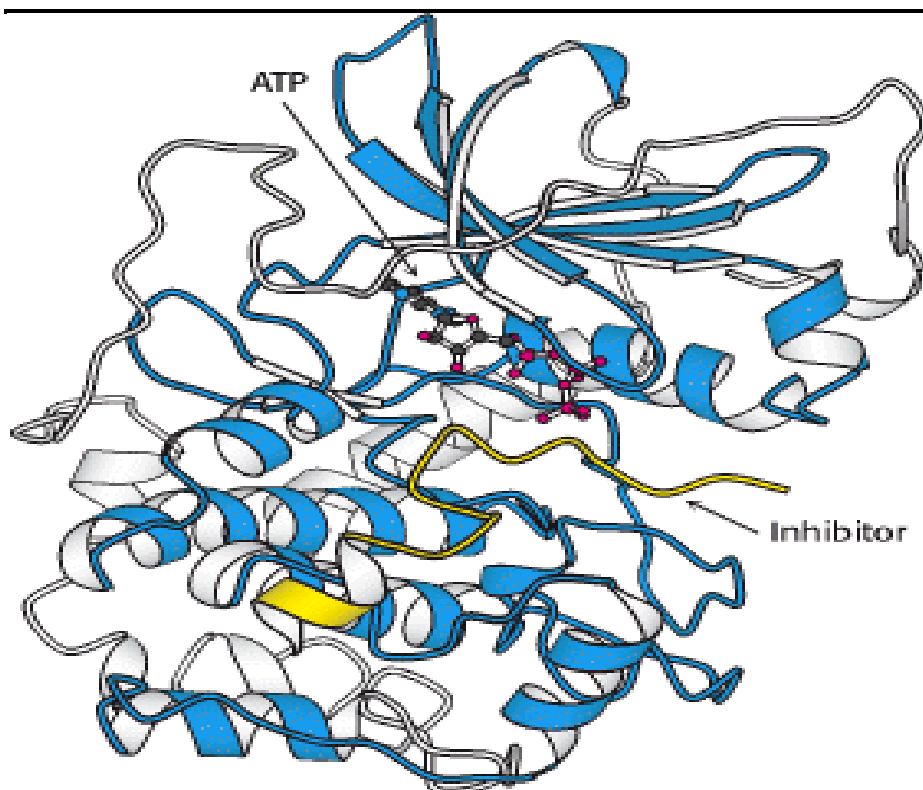


Figura 9-14. Ejemplificación de un complejo catalítico en tres dimensiones de la proteína cinasa con un inhibidor, el ATP está adyacente al sitio activo.(Strayer, 2007)

Tarea 9-10

Observe las gráficas, e indique cuáles son las diferencias significativas

- En cada gráfica ¿Qué indica cada una de las rectas? ¿Qué ocurre con la $V_{\max}$ y con la K_m ?
- Compare cada tipo de inhibición, ¿Qué ocurre con la $V_{\max}$ y con la K_m ?, ¿En cuáles se modifica y en cuáles no?

Tarea 9-11

Investigue, ¿Qué es la inhibición alostérica? Observe y explique la figura de arriba.

9.2. Fundamentos de espectrofotometría

9.2.1. La radiación electromagnética

La espectrofotometría se refiere a la medición de la cantidad de luz absorbida por una muestra en función de la longitud de onda. El sufijo de *metría* indica que es una metodología cuantitativa y no sólo cualitativa. Además, es rápida, precisa, versátil, fácil de usar y eficiente en costo. Es uno de los métodos más usados en la bioquímica analítica. Para explicar los usos prácticos de la espectrofotometría, primero debemos repasar algunos fundamentos teóricos sobre la luz y los cromóforos.

La luz tiene una naturaleza corpuscular y ondulatoria. La cantidad mínima de luz (paquete cuántico) se llama **fotón**. Los fotones son partículas y ondas a la vez. El principio de incertidumbre de Heisenberg dice que no podemos definir la ubicación exacta de una partícula al mismo tiempo que determinamos otras características. Los fotones no tienen masa, pero sí energía, la cual está definida, no por su velocidad que es constante en un determinado medio, sino por su frecuencia ν (seg^{-1}) o longitud de onda (λ).

$$E = h \times \nu$$

Ecuación 11

$$c = \lambda \times \nu$$

Ecuación 12 En donde:

E = energía de la molécula

h = constante de Planck

ν = frecuencia de la onda

λ = longitud de onda

c = constante de la velocidad de la luz en medio vacío ($\sim 3 \times 10^8$ m/s)

I = intensidad de la luz. Se refiere al número de fotones, o al equivalente de watts de potencia

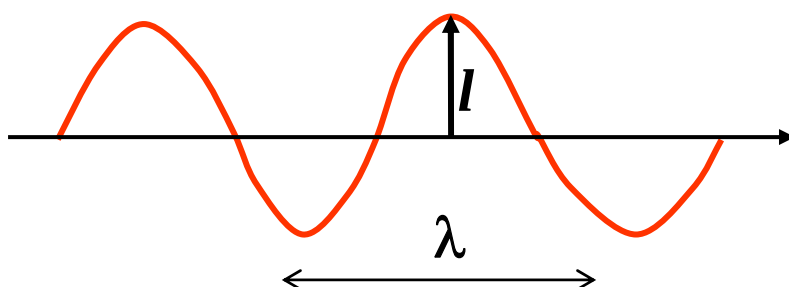


Figura 9-15. Representación de una onda senoidal con su amplitud o intensidad I y su longitud λ (lambda).

Se denomina espectro electromagnético a la distribución energética del conjunto de las ondas electromagnéticas. Referido a un objeto se denomina espectro electromagnético o simplemente espectro a la radiación electromagnética que emite (espectro de emisión) o absorbe (espectro de absorción) una sustancia. Dicha radiación sirve para identificar la sustancia de manera análoga a una huella dactilar. Los espectros se pueden observar mediante espectroscopios que, además de permitir observar el espectro, permiten realizar medidas sobre este, como la longitud de onda, la frecuencia y la intensidad de la radiación.



Figura 9-16. Espectro de absorción visible.

Tarea 9-12.

Para evaluar el contenido de nitrógeno de las plantas, a veces se usa un dispositivo que mide el contenido de clorofila de las hojas. Esto se hace con un espectrofotómetro portátil que se usa mucho en las evaluaciones de campo, por ejemplo, para el mejoramiento de maíz. Averigüe a que longitudes de onda trabajan estos equipos y haga una relación de los fundamentos de espectrofotometría que necesita saber para entender el funcionamiento de este aparato. Investigue también a nivel bioquímico que tiene que ver la concentración de clorofila con el contenido de nitrógeno y el tipo de fertilizante que se aplica en el suelo.

9.2.2. Cromóforos

Las moléculas que pueden absorber fotones se llaman cromóforos. Originalmente sólo se refería a la captura de fotones del rango visible (Vis), pero ahora también se incluyen los rangos mas amplios desde el ultravioleta (UV), hasta el rojo lejano (FR). La absorción de energía en el rango UV-VIS se debe a que estas moléculas tienen electrones que pueden ser excitados de un nivel basal a un nivel energético más alto. Debido a que estos niveles energéticos tienen valores discretos, únicamente los fotones de una determinada longitud de onda pueden excitar esos electrones. La energía de la luz incidente ($h\nu$) tiene que ser igual a la variación de energía entre dos estados electrónicos (fundamental E^0 y excitado E^*) de los orbitales moleculares del cromóforo.

Los cromóforos típicos encontrados en proteínas y ácidos nucleicos absorben luz únicamente por debajo de 300 nm, aunque las proteínas pueden contener otros grupos asociados que absorban a longitudes de onda mayores. Esta absorción espectroscópica en la región del ultravioleta de las proteínas y ácidos nucleicos contiene información importante de la composición de estas moléculas complejas y de su conformación o estructura tridimensional en solución.

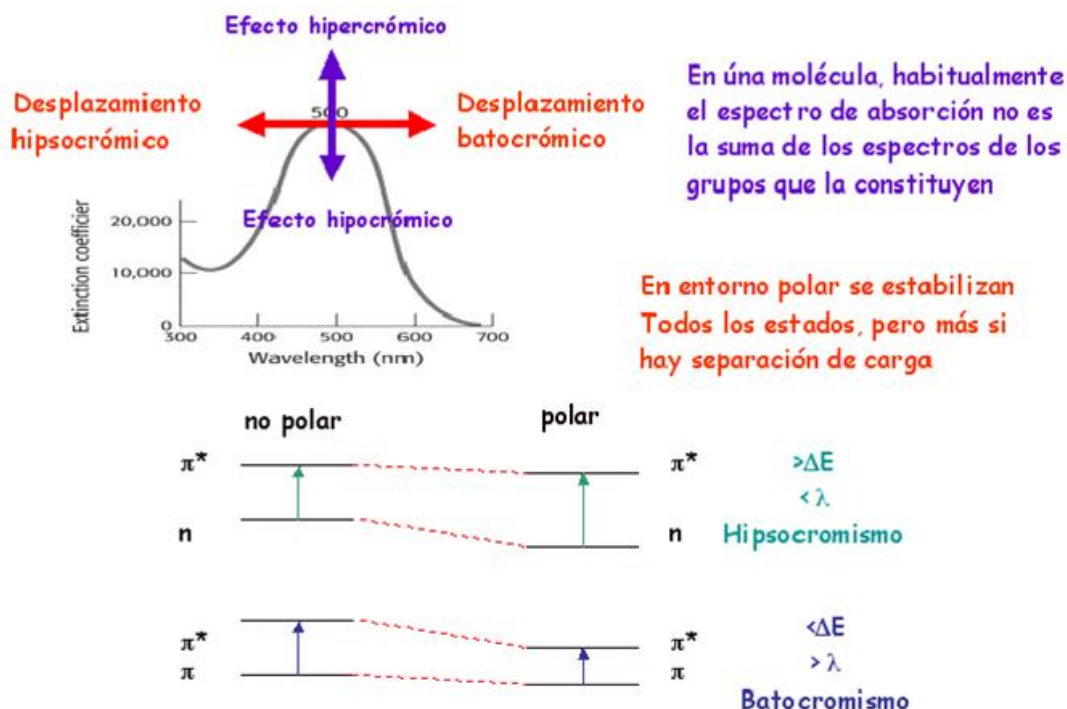


Figura 9-17. La Intensidad y posición de las bandas del espectro de absorción

9.2.3. Los auxocromos

Los auxocromos son grupos funcionales unidos a un cromóforo que modifican su absorbancia dependiendo del estado de ionización del auxocromo. Estos grupos ionizables son por ejemplo grupos carboxilo ($-\text{COOH}$), hidroxilo ($-\text{OH}$) o amino ($-\text{NH}_2$). Si estos grupos están en conjugación directa con el sistema π (π) del cromóforo, estos modifican la longitud de onda a la que la luz es absorbida, y como resultado, cambian el espectro de absorción del cromóforo. Una característica de los auxocromos es la presencia de al menos un par libre de electrones que extienden el sistema conjugado por resonancia (Figura 9-18).

Los ejemplos más comunes de moléculas con propiedades auxocrómicas son los indicadores de pH, como la fenolftaleína, que cambian su color dependiendo de la concentración de protones en el medio. El cambio de color de los indicadores de pH se debe a un cambio en los niveles energéticos de los orbitales, y esto se ve reflejado en el espectro de absorción en el rango de la luz visible.

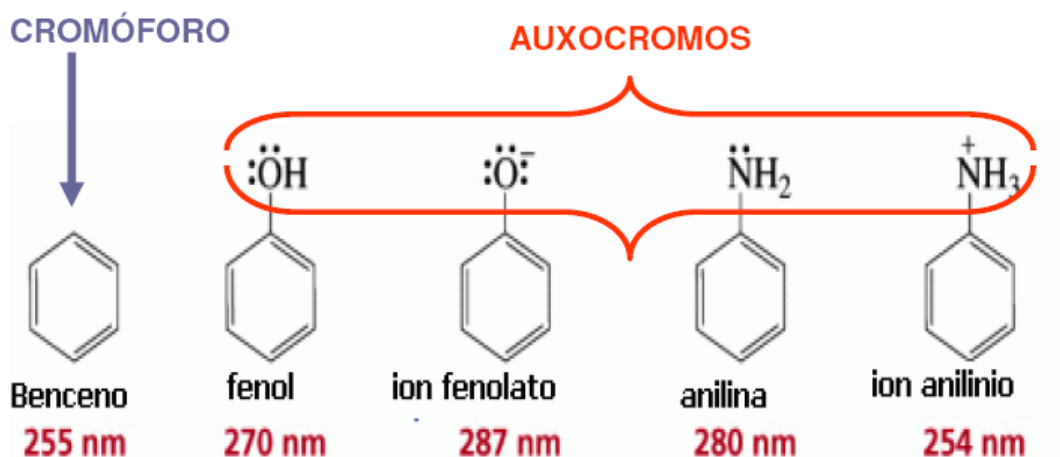


Figura 9-18. Ejemplos de Cromóforos y auxocromos. Modificado de Hesse *et al*, 1997

9.2.4. Espectro de compuestos biológicos

Muchos de los compuestos biológicos de interés, como los ácidos nucleicos y las proteínas, son cadenas de monómeros. Para comprender sus espectros de UV-VIS debemos examinar varias contribuciones al espectro:

- absorción espectral de los monómeros individuales
- contribución de la cadena de polímero
- estructuras secundarias y terciarias, incluyendo la formación de hélices
- Efectos de la cubeta y del solvente.

Una de las limitaciones de la espectrometría es que es necesario trabajar en contenedores de vidrio o de plástico. Esto restringe la espectrometría en el rango UV porque muchos de los plásticos usados para las microplacas y las micro cubetas bloquean la luz UV de 170-310 nm.

9.2.4.1. Espectro de los ácidos nucleicos

Tarea: Compare los espectros UV de adenina y de ADN y ARN.

9.2.4.2. Espectro de las proteínas

Los cromóforos presentes en una proteína pueden dividirse en tres clases: el enlace peptídico, la cadena lateral de los aminoácidos y el grupo prostético.

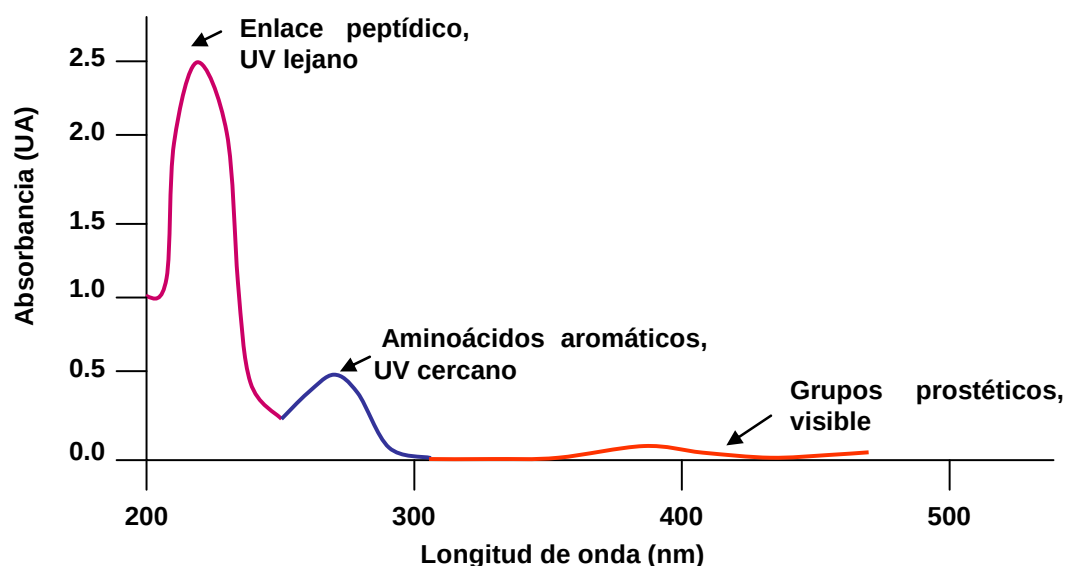


Figura 9-19. Espectro de una proteína

En cuanto a las cadenas laterales de las proteínas, las bandas de absorción más significativas se encuentran en el ultravioleta cercano, en el intervalo de longitud de onda entre 230 y 300 nm. Concretamente, los aminoácidos aromáticos absorben en el rango de 260-280 nm (figuras 10-20 y 10-21).

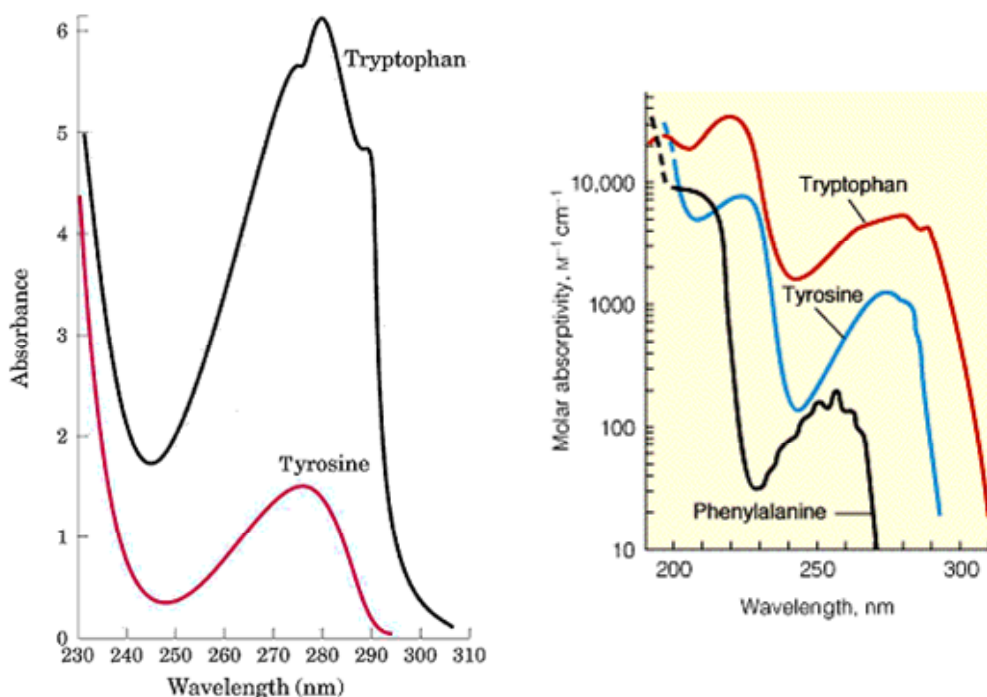


Figura 10-21. Espectros de absorción de aminoácidos aromáticos.

La fenilalanina es la que menos absorbe, aunque muestra un espectro complejo con múltiples bandas diferenciadas en torno a 250-260 nm. Presenta también bandas de mayor energía que solapan con el espectro del enlace peptídico en 215 y 180 nm. Sin embargo como no absorbe por

encima de 280 nm su contribución al espectro en la zona del ultravioleta cercano de proteínas suele ser pequeña.

La tirosina presenta un máximo a 275 nm con un hombro a 285 nm. La banda se desplaza a mayores longitudes de onda cuando se produce la desprotonación del OH del anillo aromático. Un cambio de pH produce grandes efectos en los espectros de tirosina ya que el lugar de protonación afecta directamente a la conjugación electrónica de la cadena lateral. El mayor cambio en el espectro de la tirosina por diferentes pH se da a 240nm, pero también a 280nm.

El triptófano también presenta un máximo a 280 nm. A diferencia de la tirosina, el espectro del triptófano se mantiene casi inalterado a diferentes pH.

La cisteína presenta una banda centrada a 250 nm, $n \rightarrow \sigma^*$ ($\epsilon_{\max} \sim 300 \text{ M}^{-1} \cdot \text{cm}^{-1}$), pero apenas se puede considerar, ya que es muy débil.

9.2.4.3. Grupos prostéticos

Muchas proteínas contienen además de los aminoácidos otros grupos que pueden encontrarse unidos a la proteína tanto de forma covalente como no covalente. En muchos casos puede que estos grupos sean cromóforos y contribuyan al espectro de absorción de la proteína en el UV-cercano o en el visible. Por ejemplo, flavinas, fosfato de piridoxal, metales presentan bandas de absorción intensas en estas regiones espectrales, que pueden resultar casi imprescindibles para muchos estudios físico-químicos de la acción de estas proteínas. Cambios espectrales en los distintos estados redox o según el entorno en que se encuentran.

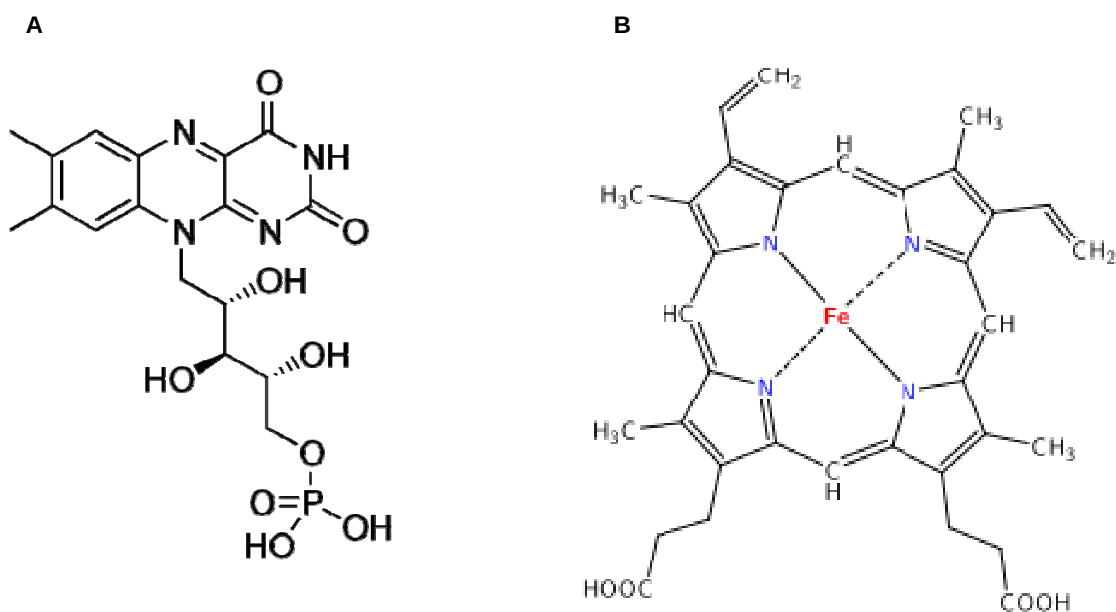


Figura 9-22. Estructuras de algunos grupos prostéticos. A) flavin mononucleótido, coenzima de varias oxidorreductasas. B) grupo hemo de la hemoglobina.

9.2.5. La ley de Beer-Lambert

Cuando la luz atraviesa un medio acuoso su intensidad decae. Es decir, no todos los fotones que entran a un medio salen, ya que algunos son absorbidos (**¡Error! No se encuentra el origen de la referencia. 23**). La ley de Beer-Lambert dice que la caída de intensidad es exponencial y depende de tres parámetros:

- 1) el factor molar de extinción del cromóforo, ϵ
- 2) la concentración del cromóforo en la solución, c
- 3) la longitud de la cubeta que la luz atraviesa, d

Esto se puede representar de la siguiente forma:

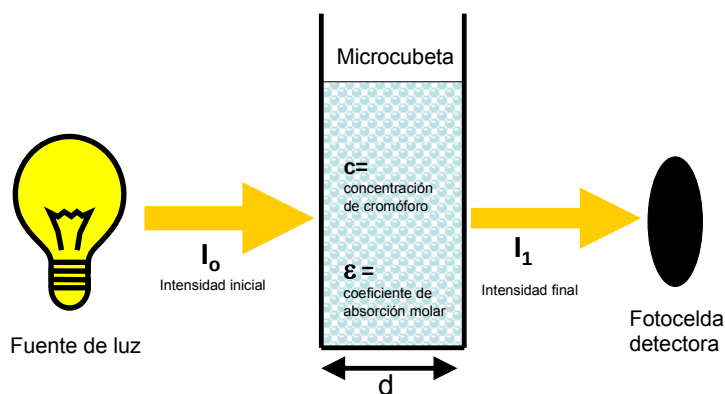


Figura 9-20 Ley de Beer- Lambert. Usando una microcubeta, se mide la intensidad que entra y que sale.

$\log \frac{I_0}{I_1} = A = \epsilon * c * d$ Ecuación 13	En donde: log = logaritmo decimal I_0 = intensidad de la luz incidente I_1 = intensidad de la luz una vez que ha atravesado el medio A = valor decimal de absorbancia ϵ = coeficiente de absorción molar de la sustancia c = concentración de sustancia absorbente en el medio d = distancia que la luz atraviesa por el medio

Las unidades de ϵ dependen del modo en que se exprese la distancia y la concentración de la sustancia absorbente. Por lo regular, la distancia se especifica en centímetros (cm), la concentración en micro moles por litro ($\mu\text{mol/l}$), mientras que el coeficiente molar se expresa en las unidades reciprocas ($\text{l}/\mu\text{mol}\cdot\text{cm}$). De esta forma, el valor de absorbancia (A) no tiene unidad.

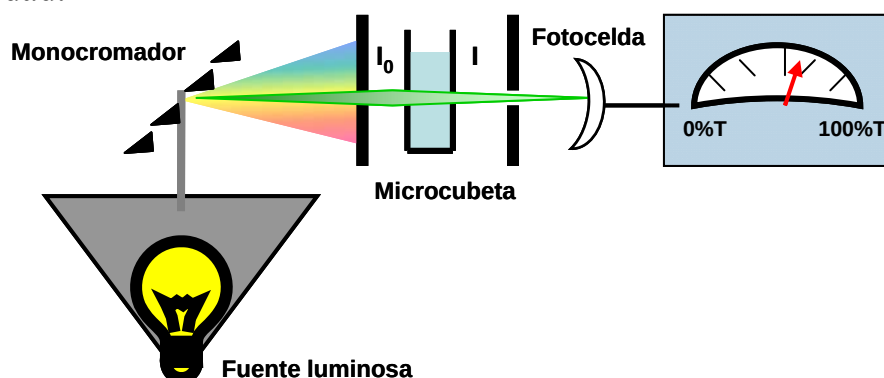


Figura 9-21. Medición de la luz que atraviesa un medio acuoso. Ley de Beer-Lambert

La relación entre la concentración y la absorción de luz a una determinada longitud de onda es la base de la espectrofotometría para identificar y cuantificar diversas sustancias. Si conocemos d y ϵ , la concentración de la sustancia puede ser deducida a partir de la cantidad de luz absorbida, y esto se logra con un espectrofotómetro. El parámetro de lectura que arroja un espectrofotómetro es el valor de transmitancia (%T) o el valor de absorbancia (A) a una determinada longitud de onda usando una cubeta de definido grosor (d) (

Figura 9-21).

El por ciento de transmitancia (%T) se refiere a la cantidad de radiación que pasa a través de la muestra y alcanza el detector. Una solución límpida, no absorbente, mostrará una lectura de 100% de transmitancia, mientras que un cuerpo opaco absorberá toda la radiación que le llega resultando en una transmitancia de 0%. Las unidades de absorbancia son al revés. Una solución translúcida da una absorbancia de cero, mientras que entre más opaco sea, mayor será el valor de A . Las unidades de absorbancia van de 0 al infinito, pero los aparatos por lo regular solamente dan una respuesta lineal en el intervalo de 0 a 2 (Skoog *et al.*, 1998). La transmitancia (T), intensidad (I) y absorbancia (A) se relacionan con la siguiente fórmula:

$$T = \frac{I_1}{I_0} = 10^{-A}$$

Ecuación 14

9.2.6. Métodos enzimáticos.

9.2.6.1. ¿Qué es un ensayo enzimático?

Un ensayo enzimático es un procedimiento experimental, mediante el cual se puede medir la velocidad de una reacción enzimática, o bien, medir la concentración de un sustrato. Como las enzimas no se consumen en la reacción que catalizan, los ensayos enzimáticos suelen medir los cambios, bien en la concentración de sustrato (que va decreciendo), bien en la concentración de producto (que va aumentando). Existen diversos métodos para realizar estas medidas. La espectrofotometría permite detectar cambios en la absorbancia de luz por parte del sustrato o del producto (según la concentración de estos) y la radiometría implica incorporación o liberación de radiactividad para medir la cantidad de producto obtenido por tiempo. Los ensayos espectrofotométricos son los más utilizados, ya que permiten medir la velocidad de la reacción de forma continua. Por el contrario, los ensayos radiométricos requieren retirar las muestras para medirlas, por lo que son ensayos discontinuos. Sin embargo, estos ensayos son extremadamente sensibles y permiten detectar niveles muy bajos de actividad enzimática. También se puede utilizar la espectrometría de masas para detectar la incorporación o liberación de isótopos estables cuando el sustrato es convertido en producto.

Los ensayos enzimáticos más sensibles utilizan láser dirigidos a través de un microscopio para observar los cambios producidos en enzimas individuales cuando catalizan una reacción. Estas medidas pueden utilizar cambios producidos en la fluorescencia de cofactores que intervienen en el mecanismo de catálisis o bien unir moléculas fluorescentes en lugares específicos de la enzima, que permitan detectar movimientos ocurridos durante la catálisis. Estos estudios están dando una nueva visión de la cinética y la dinámica de las moléculas individuales, en oposición a los estudios de cinética enzimática tradicionales, en los que se observa y se mide el comportamiento de una población de millones de moléculas de enzima.

En la Figura 9-22 se puede observar la típica evolución de una curva obtenida en un ensayo enzimático. Inicialmente, la enzima transforma el sustrato en producto siguiendo un comportamiento lineal. A medida que avanza la reacción, se va agotando la cantidad de sustrato y va disminuyendo la cantidad de producto que se genera por unidad de tiempo (disminuye la velocidad de la reacción), lo que se manifiesta en forma de curva asintótica en la gráfica. Dependiendo de las condiciones del ensayo y del tipo de enzima, el período inicial puede durar desde milisegundos hasta horas. Los ensayos enzimáticos suelen estar estandarizados para que el período inicial dure en torno a un minuto, para llevar a cabo las medidas más fácilmente. Sin embargo, los modernos equipos de mezcla rápida de líquidos permiten llevar a cabo medidas cinéticas de períodos iniciales cuya duración puede llegar a ser inferior a un segundo. Este tipo de ensayos

rápidos son esenciales para medidas de la cinética del estado estacionario, discutida más abajo.

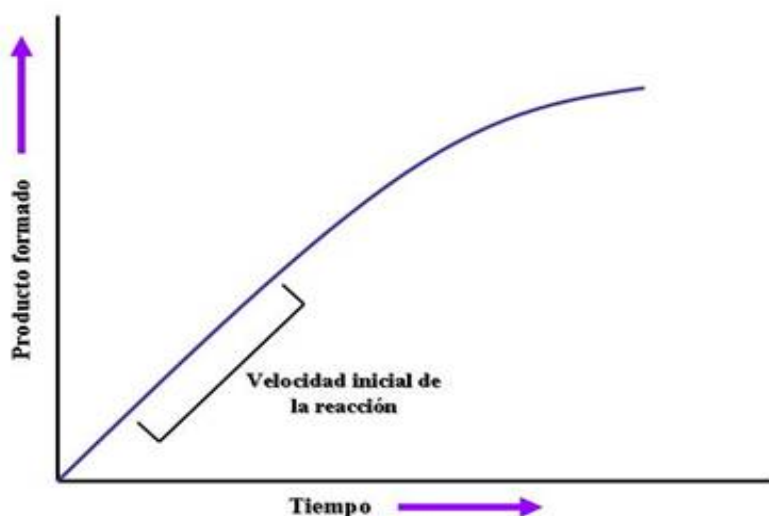


Figura 9-22. Curva de saturación de una reacción enzimática.

En la figura anterior la pendiente representa, en el periodo inicial, la velocidad de la reacción. La ecuación de Michaelis-Menten describe como va variando la pendiente con la concentración de sustrato o de enzima. En estos experimentos los parámetros cinéticos se determinan a partir de expresiones de las concentraciones de las diferentes especies como funciones del tiempo. La concentración del sustrato o del producto se mide en momentos posteriores a la rápida fase inicial y durante un tiempo lo suficientemente largo que permita a la reacción acercarse al equilibrio. Este tipo de ensayos se intenta evitar por el error matemático que se puede introducir y es cada vez menos practicado, pero fue muy ampliamente utilizado en los primeros periodos del estudio de la cinética enzimática.

La mayoría de los estudios de cinética enzimática se centran en el periodo inicial, es decir, en la zona lineal de la reacción enzimática (Figura 9-22). Sin embargo, también es posible medir toda la curva de la reacción y ajustar estos datos a una ecuación no lineal. La forma de medir las reacciones enzimáticas es denominada análisis de la curva de progreso. Esta aproximación es muy útil como alternativa a las cinéticas rápidas, cuando el periodo inicial es demasiado rápido para ser medido con precisión.

Según la fase de la reacción estudiada, todos los ensayos enzimáticos miden el sustrato consumido o el producto generado en la reacción durante un tiempo. Por ejemplo, se miden los micromoles (μmol) de producto generados por minuto. A la velocidad reacción también se le llama

actividad enzimática. Existe un número grande de diferentes métodos para medir la concentración de sustratos y productos y muchas enzimas pueden ser ensayadas de diferentes maneras.

Según el seguimiento de la reacción, los ensayos enzimáticos pueden dividirse en dos grupos dependiendo de la manera en que se sigue la medición. Por una parte están los ensayos **continuos**, en los que el método ofrece una lectura continua de la actividad monitorizando a tiempo real el sustrato consumido o el producto generado. Por otra parte existen ensayos **discontinuos**, en los que la reacción se detiene en un punto y se mide la concentración de los sustratos o los productos.

9.2.6.2. Ensayos continuos.

Los ensayos del tipo continuo son los más convenientes, al obtener la velocidad de la reacción directamente en tiempo real (*online*). Para que se puedan realizar, la detección tiene que hacerse de manera no intrusiva.

9.2.6.3. Ensayos discontinuos.

En un ensayo discontinuo se toman muestras de una reacción enzimática en intervalos y se mide en ellas la cantidad de producto formado o sustrato consumido.

9.2.6.4. Métodos de detección.

Tanto para los ensayos continuos como discontinuos hay muchos métodos diferentes de detección.

Espectrofotométricos: En los ensayos espectrofotométricos puede seguirse el curso de una reacción observando como cambia la luz absorbida por la solución en la que se está dando la reacción. Para poder utilizar un ensayo de este tipo debe haber entre los sustratos o los productos alguno que absorba luz a una longitud de onda determinada, y que sea la única molécula de la mezcla de reacción que lo hace a la misma; de esta forma, se puede observar el aumento o la disminución de la absorbancia a dicha longitud de onda.

Cuando la luz absorbida se encuentra en el espectro visible es posible observar un cambio en el color de la muestra, y entonces hablamos de método colorimétrico. También es usual el uso de la luz ultravioleta (luz UV), especialmente porque sirve para monitorizar reacciones en las que participan NADH y NADPH, coenzimas que participan en infinidad de reacciones bioquímicas y que absorben luz UV en forma reducida (NADH y NADPH), pero no en forma oxidada (NAD^+ y NADP^+). Así, la actividad

de una oxidorreductasa que use NADH como coenzima puede ser monitorizada siguiendo el descenso de la absorbancia a 340 nm a medida que se consume el NADH

Calorimétricos: La calorimetría es la medida del calor liberado o absorbido por una reacción química. Estos ensayos son muy generales, ya que muchísimas reacciones llevan asociado algún cambio de calor y utilizando un microcalorímetro no es necesario utilizar grandes cantidades de enzima o sustrato. Estos ensayos pueden ser usados para medir reacciones imposibles de medir con otros métodos.

Quimioluminiscencia: La quimioluminiscencia es la emisión de luz por una reacción química. Algunas reacciones enzimáticas producen luz y esta puede ser medida para detectar la formación de producto. Estos tipos de ensayo pueden ser extremadamente sensibles, ya que la luz producida puede ser registrada en una película fotográfica por días e incluso semanas, pero al mismo tiempo son de difícil cuantificación, porque no toda la luz emitida por la reacción será detectada.

La detección de la peroxidasa de rábano por quimioluminiscencia enzimática (ECL, *enzymatic chemiluminescence*) es un método común para detectar anticuerpos en el *western blot*. Otro ejemplo de quimioluminiscencia es la luciferasa, enzima presente en las luciérnagas que produce luz de forma natural a partir de su sustrato, luciferina.

Radiométricos: En los ensayos radiométricos se mide la incorporación de radiactividad en los productos o su liberación desde sustratos. Los radioisótopos más usados en estos ensayos son ^{14}C , ^{32}P , ^{33}P , ^{35}S y ^{125}I . Como los isótopos radiactivos permiten el marcaje de un solo átomo de un sustrato, estos ensayos son extremadamente sensibles y específicos. Frecuentemente son utilizados en bioquímica y a menudo son la única manera de medir una reacción específica en extractos crudos (mezclas complejas de las enzimas liberadas al lisar células). En estos procedimientos la radiactividad es normalmente medida en contadores de centelleo.

Cromatográficos: En los ensayos cromatográficos se mide la formación de producto separando los componentes de la mezcla de reacción por cromatografía. Normalmente se lleva a cabo mediante HPLC. Aunque esta aproximación puede requerir grandes cantidades de material de partida, su sensibilidad puede incrementarse mediante marcaje radiactivo de los sustratos o los productos.

Fluorimétricos: En el fenómeno de la fluorescencia una molécula emite luz de una longitud de onda determinada tras absorber luz a otra longitud

de onda. Los ensayos fluorimétricos usan la diferencia en la fluorescencia entre producto y sustrato para medir la reacción enzimática. Estos ensayos son generalmente mucho más sensibles que los espectrométricos, ya que son más específicos al tener que utilizar dos longitudes de onda en lugar de una, pero pueden sufrir más interferencias por impurezas presentes en el medio o por la inestabilidad que muchos compuestos fluorescente presentan al ser expuestos a la luz.

Un ejemplo de estos ensayos es, una vez más, el uso de las coenzimas NADH y NADPH. En este caso, las formas reducidas son fluorescentes y las oxidadas no. Las reacciones de oxidación pueden, por tanto, ser seguidas por el descenso de la intensidad de fluorescencia, y las de reducción por el aumento. También existen sustratos sintéticos que liberan un fluoróforo en reacciones catalizadas por enzima, como por ejemplo el 4-metilumbeliferil- β -D-glucurónido para ensayar la β -galactosidasa.

9.2.6.5. Métodos enzimáticos acoplados a NADPH

En los ensayos espectrofotométricos puede monitorearse una reacción bioquímica observando cómo cambia la densidad óptica de la solución en la que se está dando. Para utilizar un ensayo de este tipo debe haber un sustrato que absorba la luz de manera diferente al producto que genera. Por medio de espectrofotómetro se puede observar la disminución de la concentración del sustrato (o el aumento del producto) en tiempo real. Cuando la luz absorbida se encuentra dentro del rango del espectro visible se observa un cambio en el color de la muestra, y entonces hablamos de un método colorimétrico. En muchos casos se usa luz ultravioleta (luz UV). Para esos casos se necesitan cubetas especiales, que sí sean transparentes a una longitud de onda menor de 400nm. La luz UV sirve para monitorizar reacciones en las que participan NADH y NADPH. Estas coenzimas participan en infinidad de reacciones bioquímicas, principalmente de oxidorreducción. NADH y NADPH en su forma reducida absorben luz UV a 340nm, mientras que en su forma oxidada (NAD⁺ y NADP⁺) son casi transparentes a esta longitud de onda (Figura 9-23). Así, la actividad de una oxidoreductasa que use NADP como coenzima puede ser monitoreada, siguiendo el incremento de la absorbancia a 340 nm a medida que se produce el NADPH.

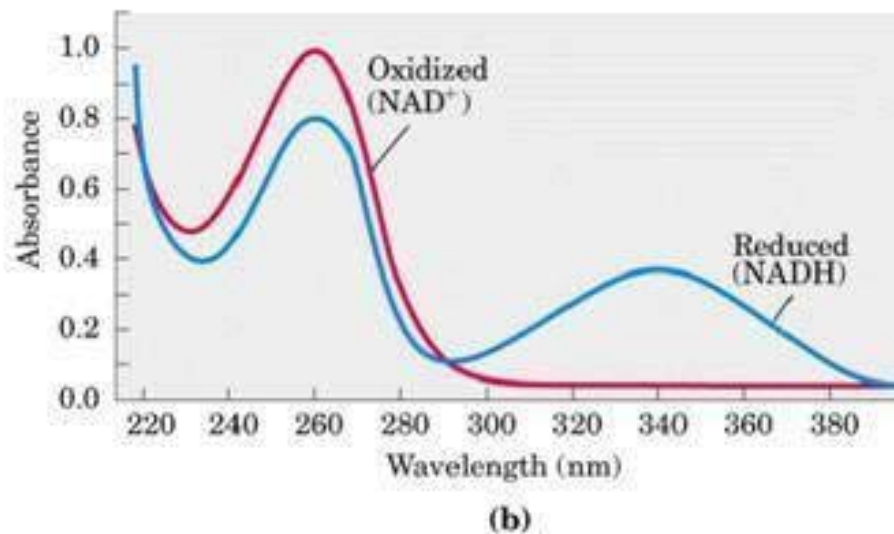


Figura 9-23. Especto del Absorción del NAD^+ y del NADH

Incluso cuando la reacción enzimática estudiada no provoca ningún cambio de absorbancia o esta no es específica, pueden usarse ensayos espectrofotométricos para la enzima realizando un ensayo acoplado. En este, el producto de la reacción a estudiar se usa como sustrato de una segunda reacción, que ha de ser de fácil seguimiento y cuyos otros sustratos, coenzimas y enzimas deben encontrarse a concentraciones saturantes para que la velocidad de la segunda reacción no sea limitante. Por ejemplo, en la Figura 9-24 puede verse el esquema de un ensayo acoplado para la hexocinasa, que puede medirse acoplando la producción de glucosa-6-fosfato a la producción de NADPH usando la glucosa-6-fosfato deshidrogenasa

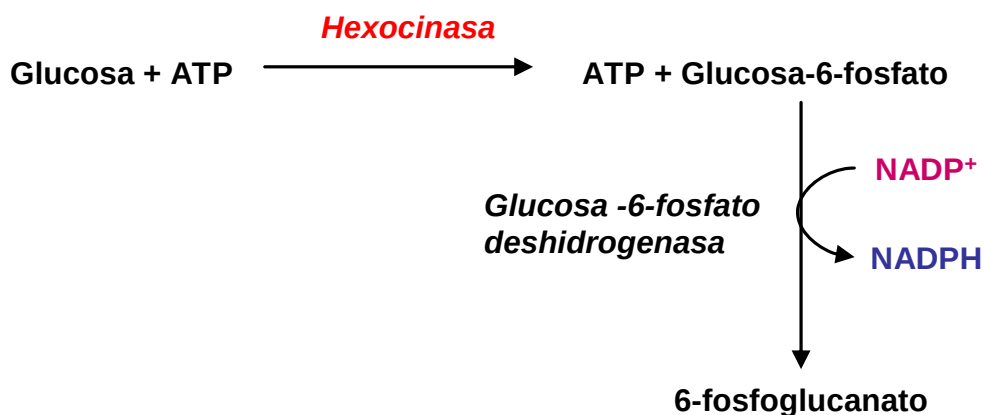


Figura 9-24. Esquema del ensayo acoplado para la hexocinasa

9.2.6.6. Ejemplo: análisis de azúcares

En la siguiente figura se puede observar el esquema de un ensayo acoplado para la medición de tres azúcares de forma secuencial: glucosa, fructosa y sacarosa. Por cada molécula de hexosas se produce un mol de NADPH, mientras que por cada mol de sacarosa se produce 2 moles de NADPH.

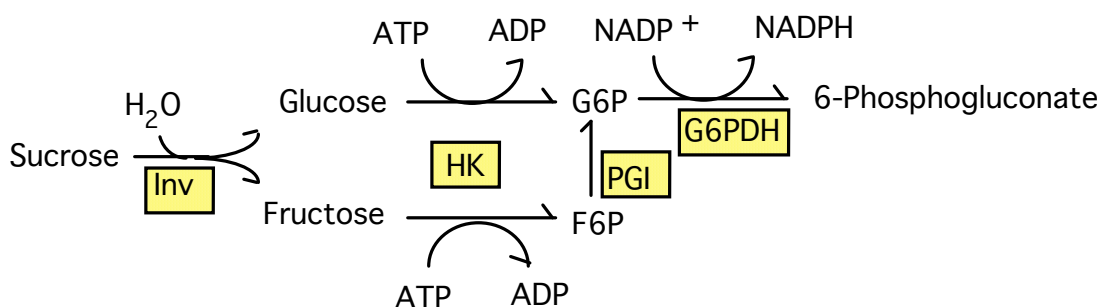


Figura 9-25. Ejemplificación de la producción de NADPH a partir de glucosa, fructosa y sacarosa.

Tarea 9-13

El maíz produce sacarosa a través de la fotosíntesis en las hojas. Bajo condiciones limitantes de agua (sequía), las estomas de las hojas se cierran, evitando así el escape de vapor de agua. Esto también limita la entrada de dióxido de carbono, y por consiguiente, inhibe la fotosíntesis. Un fisiólogo de maíz ha propuesto un proyecto en donde se medirán los niveles de glucosa y sacarosa en las hojas y tallos de diferentes genotipos de maíz contrastantes, ya sea más, o menos tolerantes a la sequía. ¿Cree usted que la medición de azúcares es pertinente? ¿Se justifica el uso de esta herramienta bioquímica para un proyecto de mejoramiento genético de maíz? Asuma una postura como evaluador de este proyecto y defiéndala con argumentos.

9.2.6.7. Métodos cinéticos

Uno de los métodos recientes para cuantificar metabolitos en un tejido es conocido como *cycling assay*. Esta técnica consiste en la medición o cuantificación de metabolitos a muy bajas concentraciones basado en la actividad de dos enzimas que en su conjunto catalizan una reacción cíclica. Una enzima convierte el metabolito A en el metabolito B y una segunda enzima convierte el metabolito B otra vez en el A. Es decir, la reacción global consiste en un reciclaje de una misma molécula. El truco de los *cycling assays* es acoplarlos a la generación de una molécula reportera, es decir, cogenerar algún metabolito que pueda ser fácilmente cuantificado, como por ejemplo, NADH o MTT. Como las enzimas y cofactores están en exceso, la velocidad del ciclo depende de la suma de concentración A y B. Las concentraciones pueden ser determinadas específicamente ya que la

velocidad de la reacción será linealmente dependiente de la concentración de estos sustratos. Según la ecuación de Michaelis-Menten la relación entre la concentración de los sustratos cíclicos y la velocidad de reacción, es lineal en el rango de $S < K_m$. Podemos decir que los *cycling assay* permiten detecciones a muy bajas concentraciones por la alta afinidad de las enzimas por sus sustratos. La Figura 9-26 ejemplifica un ensayo de *cycling* basados en la determinación de metabolitos producidos por glicerol-3-P oxidasa/glicerol-3-P deshidrogenasa, Gly3POX y Gly3PDH, respectivamente.

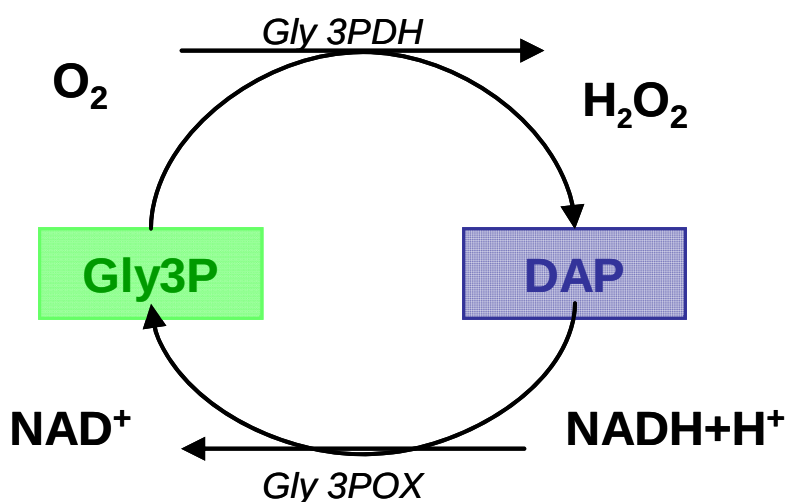


Figura 9-26. Reacción de *cycling assay* basada en el ciclo de glicerol-3-fosfato.

La (Gly3POX) cataliza la conversión de oxígeno (O_2) del gliceraldehído 3 fosfato (Gly3P) en dihidroxiacetona-fosfato (DAP) mientras que la (Gly3PDH) simultáneamente convierte DAP a Gly3P y oxida el NADH a NAD^+ , el cuál es detectado espectrofotométricamente a 340 nm. Este tipo de reacciones acopladas que no dependen directamente del sustrato utilizado para monitorear la reacción se llaman pseudo orden cero. En el caso de la Figura 9-26 la velocidad de la reacción cíclica no depende, ni de la concentración de O_2 , ni de NADH. La velocidad solo dependerá de la suma de DAP y Gly3P. La velocidad de disminución de NADH determinará la concentración de este metabolito (Gibon et al., 2002). Idealmente, los *cycling assay* se llevan a cabo a concentraciones de sustrato que son menores que el K_m de las dos enzimas participantes (Gly3PDH y Gly3POx), que por lo regular son del orden de micromoles por litro, por lo que con este tipo de ensayos son muy sensibles, ya que se pueden detectar cantidades del rango de 10 picomoles de Gly3P.

9.3. *Análisis de macromoléculas*

La electroforesis y la cromatografía comprenden un conjunto diverso de métodos que permite a los científicos separar componentes en mezclas complejas, lo que en muchas ocasiones resulta imposible por otros medios. Se pueden separar macromoléculas en función de sus cargas, tamaños y masas moleculares. También a través de la polaridad de sus enlaces, sus potenciales redox, entre otros.

9.3.1. **Electroforesis**

La electroforesis es una técnica analítica de separación de fundamento cinético, basada en el movimiento o migración de las macromoléculas como resultado de la acción de un campo eléctrico. Por lo regular la migración se hace con macromoléculas disueltas en un determinado medio (solución tampón de electroforesis), a través de una matriz o soporte reticulado. El comportamiento de la molécula viene dado por su movilidad electroforética, y esta por la carga, tamaño y forma de la misma. Cuanto mayor es la relación carga/tamaño más rápido migra un ión en el seno del campo eléctrico.

Esta técnica tiene como ventaja su capacidad de separar macromoléculas de interés en la industria biotecnológica y química. Ha sido un método muy útil para la separación de proteínas (enzimas, hormonas) y ácidos nucleicos (DNA y RNA) con gran resolución.

La electroforesis fue desarrollada por primera vez por el químico sueco **Arne Tiselius**, merecedor del Premio Nobel en 1948 por sus trabajos realizados en esta área.

La electroforesis en gel es un grupo de técnicas empleadas por los científicos para separar moléculas basándose en propiedades como el tamaño, la forma o el punto isoelectrico. La electroforesis en gel, ya sea en base a agarosa o acrilamida, se utiliza generalmente con propósitos cualitativos, pero también analíticos. Puede ser una técnica preparativa para purificar moléculas parcialmente antes de aplicar espectrometría de masas, PCR, clonación o secuenciación de DNA.

9.3.1.1. **Ácidos nucleicos**

En los ácidos nucleicos la dirección de migración es del electrodo negativo al positivo, y esto es debido a la carga negativa presente en el esqueleto azúcar-fosfato. En los fragmentos de DNA dobles (con estructura de doble hélice) **la velocidad de migración es inversamente proporcional a su tamaño**. En fragmentos de RNA y fragmentos simples de de DNA (una sola cadena), dichas moléculas tienden a plegarse de forma compleja y a migrar de forma más complicada, según la estructura terciaria formada tras el plegamiento. Sin embargo, compuestos que puedan romper los enlaces

de puente de hidrógeno, como el hidróxido de sodio o la formamida, son empleados para 'desplegar' las moléculas lineales y permitir que la velocidad de migración dependa únicamente del tamaño y no de la estructura formada tras el plegamiento.

9.3.1.2. Proteínas

Por otra parte, las proteínas no tienen una estructura predecible como los ácidos nucleicos, y por tanto sus velocidades de migración no son similares entre ellas. Incluso puede que no migren ni al aplicar una fuerza electromotriz (al encontrarse en su punto isoeléctrico). Laemmli, 1970, modificó la electroforesis para proteínas mediante la adición de un detergente como el dodecilsulfato sódico/dodecilsulfato sódico (SDS/SDP) y un agente reductor como el 2-mercaptoetanol. Los detergentes otorgan una carga neta negativa a la proteína que les permite migrar a través del gel de poliacrilamida en relación directa a su masa, ya que la cantidad de cargas negativas que se unen a la proteína depende del tamaño de ésta, existiendo una relación carga/masa similar. Por otro lado, el agente reductor rompe los enlaces disulfuros, separando a la proteína en sus sub-unidades. Además, la desnaturalización hace que pierdan su estructura terciaria y cuaternaria por tanto su velocidad de migración es proporcional al tamaño y no a su estructura terciaria ni a su interacción con otras macromoléculas. Así, los más grandes se desplazan más lentamente.

Cuando se ha completado la electroforesis, las moléculas más pequeñas han llegado al ánodo. Entonces se pueden 'revelar' mediante la adición de un colorante específico para hacerlas visibles. Se emplean compuestos como el bromuro de etidio, para los ácidos nucleicos, o tinción de plata, azul de coomassie o tinción fluorescente, para las proteínas. Asimismo se emplean otros métodos para visualizar la separación de la mezcla en el gel. Si el reactivo es fluorescente bajo la luz UV, se puede simplemente hacer una fotografía de la placa bajo dicha luz. También, si las moléculas contienen átomos radiactivos se puede efectuar una autorradiografía.

Las bandas en diferentes separaciones paralelas, que están a la misma distancia del principio, significan que contienen moléculas que han atravesado el gel a la misma velocidad. Existen marcadores especiales que contienen una mezcla de moléculas de tamaño conocido. Si se hace una electroforesis de un marcador con una mezcla desconocida, las bandas observadas en el marcador pueden ser comparadas con las obtenidas en la mezcla desconocida para determinar su tamaño o punto isoeléctrico. La distancia a la que se encuentra la banda del principio es (aproximadamente) inversamente proporcional al logaritmo del tamaño de la molécula.

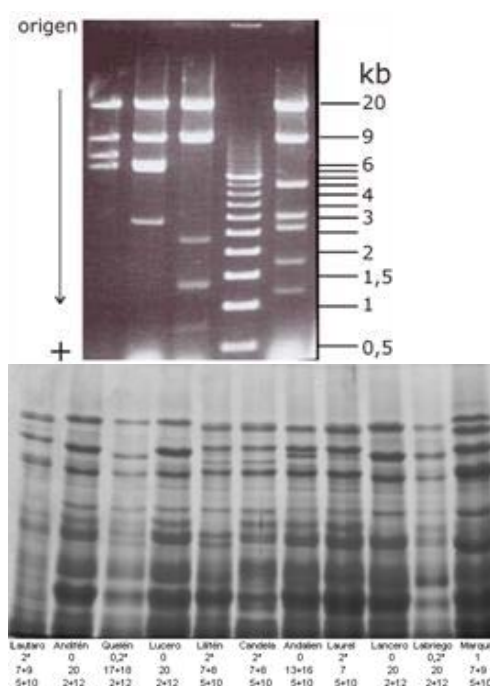


Figura 9-27. Ejemplos de separación de proteínas (SDS-PAGE) izquierda y de ácidos nucleicos derecha.

9.3.2. Cromatografía

La cromatografía se distingue de la electroforesis por la forma en que migran las partículas. En las técnicas cromatográficas, la migración de las moléculas está dada por un flujo de masas, y no por una fuerza eléctrica, como en la electroforesis. La cromatografía no sólo permite la separación de los componentes de una mezcla, sino también su identificación y cuantificación. Se pueden distinguir compuestos que están muy estrechamente relacionados. El análisis cualitativo está basado en la medida de parámetros cromatográficos (tiempos y volúmenes de retención) mientras que el análisis cuantitativo está basado en la medida de alturas o áreas de picos cromatográficos que se relacionan con la concentración. La columna cromatográfica y la forma con la que se diseña, constituye el corazón de la separación. El detector, situado al final de la columna es el que garantiza la respuesta de los componentes que se separan.

En todas las separaciones cromatográficas la muestra se desplaza con una fase móvil, que puede ser un gas o un líquido. La fase móvil puede pasarse a través de una fase estacionaria, con la que es inmisible y que se fija a una columna o a una superficie sólida. Las dos fases se eligen de forma, que los componentes de la muestra se distribuyen de modo distinto entre la fase móvil y la fase estacionaria. Aquellos componentes que son fuertemente retenidos, por la fase estacionaria se mueven lentamente con el flujo de la fase móvil; por el contrario los componentes que se unen débilmente a la fase estacionaria, se mueven con rapidez. Como

consecuencia de la distinta movilidad, los componentes de la muestra se separan en bandas o zonas discretas que pueden analizarse cualitativa y/o cuantitativamente. Estos resultados se recogen en forma de gráficos llamados cromatogramas.

Los métodos cromatográficos se pueden clasificar según la forma en que la fase móvil y la fase estacionaria se pongan en contacto. Así, en la cromatografía de columna, un tubo estrecho contiene la fase estacionaria a través de la cual se hace pasar la fase móvil por presión. En la cromatografía en plano, la fase estacionaria se fija sobre una placa plana o a los intersticios de un papel; en este caso la fase móvil se desplaza a través de la fase estacionaria por capilaridad o por gravedad.

Las tres clases de cromatografía desde un ángulo más general son cromatografía de líquidos, cromatografía de gases y cromatografía de fluidos supercríticos. Como su nombre lo indica, la fase móvil en las tres técnicas son líquido, gas y fluido supercrítico respectivamente.

Sólo la cromatografía de líquidos es la que puede llevarse a cabo en columna o sobre superficies planas, por otra parte tanto la de gas como la de fluidos supercríticos están restringidas a los procedimientos en columna de tal manera que las paredes de la columna contienen la fase móvil.

La técnica más usada es la cromatografía líquida de alta resolución, HPLC, por su sensibilidad, fácil adaptación a las determinaciones cuantitativas exactas, su idoneidad para la separación de especies no volátiles o termolábiles y su aplicación a sustancias de primordial interés en la industria, como son los aminoácidos, proteínas, ácidos nucleicos, hidrocarburos, carbohidratos, entre otros.

La cromatografía líquida de proteínas la podemos diferenciar en cuatro grupos:

- de reparto: separa los solutos basándose en la solubilidad
- de adsorción: se basa en la afinidad de adsorción
- de exclusión: separa solutos según el peso molecular
- de intercambio iónico: separa solutos según la carga iónica

Los métodos de la cromatografía de adsorción y reparto tienden a ser complementarios, aunque en algunos casos se superponen. La cromatografía de reparto está formada por la cromatografía Líquido-Líquido y la cromatografía unida químicamente. Estas técnicas se diferencian en la forma en que se retiene la fase estacionaria sobre las partículas del soporte relleno. En el segundo caso, como su nombre lo

indica, la fase estacionaria se une químicamente a la superficie del soporte. Esta técnica actualmente es más usada debido a que la cromatografía líquido-líquido necesita de un recubrimiento periódico de las partículas del soporte debido a la pérdida de la fase estacionaria por disolución en la fase móvil.

En general los soportes para casi todos los rellenos de fases unidas químicamente se preparan con sílice rígida o composiciones donde la sílice es el elemento básico. Estos sólidos están formados por partículas mecánicamente resistentes, porosas y uniformes.

Los rellenos de columna en la cromatografía de fase unida químicamente se clasifican como de fase inversa cuando el recubrimiento enlazado tiene carácter no polar, y de fase normal cuando el recubrimiento contiene grupos funcionales polares. Esta técnica puede utilizarse en la separación de mezclas edulcorantes artificiales, antioxidantes, aflatoxinas, bebidas refrescantes entre otras.

En la siguiente sección se explican algunos ejemplos de los métodos cromatográficos más usados hoy en día para el mejoramiento genético, por ejemplo, la determinación de la calidad nutricional del maíz.

9.4. Metodologías y equipos

9.4.1. Cromatografía de líquidos (HPLC)

La clave para conseguir la precisión deseada en la bioquímica analítica es el método de extracción y el método de cuantificación. El protocolo y la apropiada configuración de los instrumentos son muy importantes, así como el entrenamiento del personal que los utilizará. El instrumento mejor diseñado, el más resistente en el mundo, no va a dar resultados adecuados a largo plazo si los usuarios finales no entienden cómo usarlo ni mantenerlo. El soporte técnico y el personal calificado son factores críticos. El protocolo mejor escrito no funcionará si la persona que realiza la extracción no tiene experiencia en el laboratorio y al menos un mínimo de conocimientos químicos. Es muy importante que los estudiantes entiendan todos los fundamentos de la bioquímica y de la cromatografía para sacarle mayor provecho a estas herramientas analíticas dentro de sus proyectos experimentales.

Introducción a cromatografía



La Cromatografía líquida de alta resolución o *High Performance Liquid Chromatography* (HPLC) es el tipo de cromatografía utilizada con más frecuencia en la bioquímica analítica. Las columnas de HPLC separan los componentes de una mezcla basándose en diferentes tipos de interacciones químicas entre las sustancias analizadas y la fase de la columna cromatográfica. La separación física está basada en la distribución de los componentes químicos de una mezcla entre dos fases, una fase fija o estacionaria y otra móvil.

La cromatografía líquida “clásica” se lleva a cabo en columnas generalmente de vidrio o plástico, las cuales se rellenan de manera manual con una resina tipo gel. Luego de colocar la muestra en la parte superior, se hace fluir la fase móvil a través de la columna por efecto de la gravedad. Esto todavía se hace mucho para separar y purificar proteínas.

Con el objeto de aumentar la eficiencia en las separaciones, se fue disminuyendo el tamaño de las partículas de fase fija hasta el tamaño de los micrones, lo cual generó la necesidad de utilizar altas presiones para lograr que fluya la fase móvil. Así nació la técnica de cromatografía líquida de alta resolución (HPLC), que requiere de instrumental especial para trabajar con las altas presiones requeridas. En HPLC la fase móvil es un líquido que fluye a altas presiones a través de una columna de metal que contiene una resina empacada desde la fábrica. Dependiendo del tipo de fase móvil o fija, y del tipo de fenómeno físico que provoca la separación, existen diversas variantes de cromatografía líquida:

Cromatografía de exclusión

La cromatografía de exclusión, también llamada de filtración en gel o cromatografía de permeación en gel, se basa en la diferencia de penetración de las moléculas de la mezcla a través de los poros de la fase estacionaria, de tal manera que la separación obtenida depende del tamaño de la molécula. El tiempo de elución es proporcional al peso molecular de los

componentes de la mezcla, por lo que no es muy usada con los compuestos de alto peso molecular. Este tipo de separación por tamaño difiere de las demás técnicas de cromatografía en que no existen interacciones físicas o químicas entre el analito y la fase estacionaria. Es una técnica reproducible, escalable y rápida.

La fase fija está formada por partículas poliméricas o de sílice que contienen una red uniforme de poros por los que pueden penetrar las moléculas de pequeño tamaño. Las moléculas de tamaño grande se excluyen totalmente y son eluidas en primer lugar, mientras que las de pequeño tamaño tienen acceso a todo el volumen poroso y son las últimas que se eluyen; de esto se deduce que el volumen disponible para las moléculas pequeñas es mayor que para las grandes. Por lo tanto las moléculas se eluyen por su **tamaño** decreciente. En resumen los factores que determinan la separación de las moléculas son el tamaño del poro, el tamaño de la partícula y el flujo de elución.

Esta separación por tamaño depende del radio de Stokes de las partículas a separar. La ley de Stokes se refiere a la fuerza de fricción experimentada por objetos esféricos moviéndose en el seno de un fluido viscoso en un régimen laminar (de bajos números de Reynolds). En general la ley de Stokes es válida en el movimiento de partículas esféricas pequeñas moviéndose a velocidades bajas.

La ley de Stokes puede escribirse como: $F_r = 6\pi R\eta v$, donde R es el radio de la esfera, v su velocidad y η la viscosidad del fluido.

Los diferentes tipos de partículas usadas para la fase estacionaria deben ser estables, mecánica y químicamente, tener bajo contenido en grupos iónicos, uniformidad de poro y tamaño. Los compuestos pueden ser derivados de dextranos (Sephadex), de agarosa (Sephacosa), de acrilamidas (Biogel P), o bien, esferas de vidrio. Hay diferentes tamaños de partícula para un gel: a menor tamaño mayor resolución y menor gasto en la columna.

Esta técnica se emplea en la separación de proteínas de alimentos, determinación de glucosa y fructosa en zumos de fruta, etc.

La cromatografía de exclusión es un método de baja resolución, de forma que se suele utilizar en los pasos finales del proceso de purificación. También es muy útil para la determinación de las estructuras terciaria y cuaternaria de las proteínas purificadas.

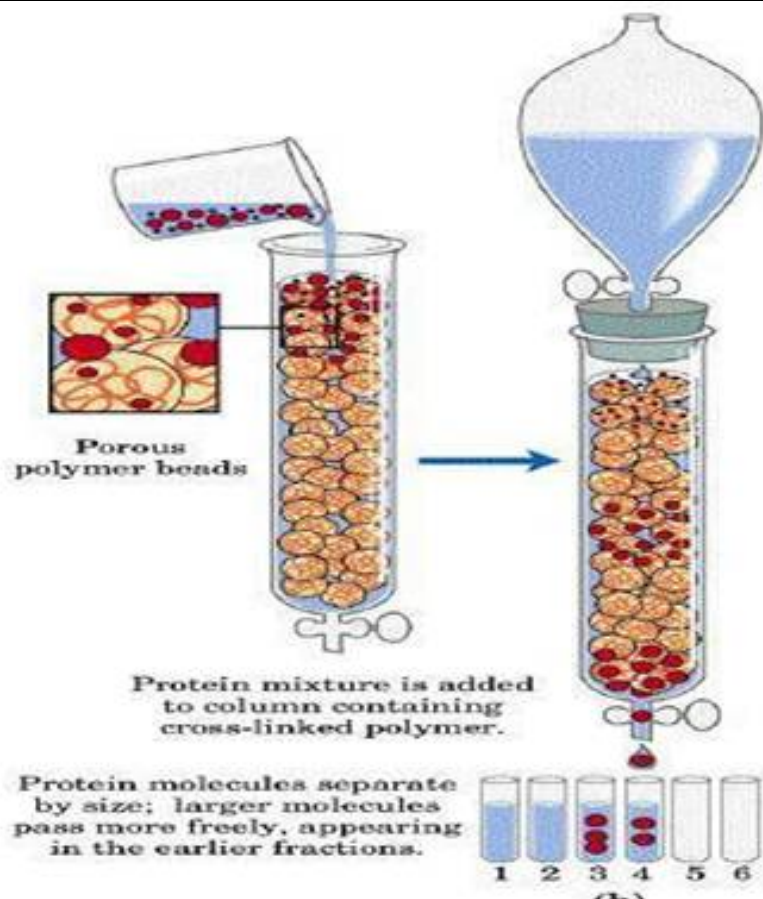


Figura 9-28 Cromatografía de exclusión molecular.

Cromatografía de adsorción

La cromatografía de adsorción o Líquido-Sólido es la forma clásica de la cromatografía de líquidos. Ha sufrido algunas transformaciones que la han convertido en un método importante de la HPLC. Las únicas fases que se utilizan en este tipo de cromatografía son la sílice y la alúmina, siendo la primera la preferida. Es adecuada para compuestos no polares probablemente con masas moleculares inferiores a 5000.

En general la cromatografía Líquido-Sólido es más adecuada para muestras que son solubles en disolventes no polares y por ello tienen solubilidad limitada en disoluciones acuosas que son las que se utilizan en cromatografía de reparto en fase inversa. En este tipo de cromatografía también se pueden separar compuestos con diferentes grupos funcionales. Una característica particular de este método es su capacidad para diferenciar compuestos isómeros en mezclas.

Cromatografía de intercambio iónico

La cromatografía de intercambio iónico está basada en la atracción entre iones del soluto y puntos cargados que existen en la fase estacionaria. En el caso de intercambiadores aniónicos, grupos cargados positivamente en la

fase estacionaria atraen aniones del soluto. Los intercambiadores catiónicos contienen puntos cargados negativamente que atraen cationes del soluto.

Los intercambiadores iónicos se clasifican en ácidos o básicos, fuertes o débiles. Las resinas ácidas fuertes siguen ionizadas incluso en disoluciones muy ácidas, en cambio las resinas ácidas débiles se protonan a un pH próximo a 4 y pierden su capacidad de intercambio catiónico. Los grupos muy básicos de amonio cuaternario siguen siendo catiónicos a cualquier valor de pH. Los básicos débiles de amonio terciario se desprotonan en disoluciones moderadamente básicas y pierden entonces su capacidad.

Las resinas de intercambio iónico tienen aplicación en estudios donde intervienen moléculas pequeñas ($PM=500$) que pueden penetrar en los poros pequeños de la resina. Los intercambiadores iónicos de poliestireno son tan grandes que las macromoléculas muy cargadas, como las proteínas, se pueden enlazar irreversiblemente a ellos. Los de celulosa y dextranos sirven para intercambio iónico de macromoléculas. Los geles de intercambio iónico se usan en el caso de moléculas grandes (proteínas y ácidos nucleicos). Cuando las separaciones exigen condiciones químicas fuertes se emplean intercambiadores iónicos inorgánicos.

En la cromatografía de intercambio iónico, la retención se basa en la atracción electrostática entre los iones en solución y las cargas inmovilizadas a la fase estacionaria. Los iones de la misma carga son excluidos mientras que los de carga opuesta son retenidos por la columna. Un incremento en la concentración del contraión (respecto a los grupos funcionales de la resina) reduce el tiempo de retención. Un incremento en el pH reduce el tiempo de retención en las cromatografías de intercambio catiónico mientras que una disminución del pH reduce el tiempo de retención en las cromatografías de intercambio aniónico.

Entre los intercambiadores iónicos podemos encontrar: resinas de poliestireno, intercambiadores iónicos de celulosa y dextranos (geles) y sílice porosa o vidrio de tamaño de poro controlado. En general los intercambiadores iónicos favorecen la unión de iones de elevada carga y radio pequeño. Este tipo de cromatografía es ampliamente utilizado para la purificación de agua, concentración de componentes traza, separación de proteínas, carbohidratos y oligosacáridos.

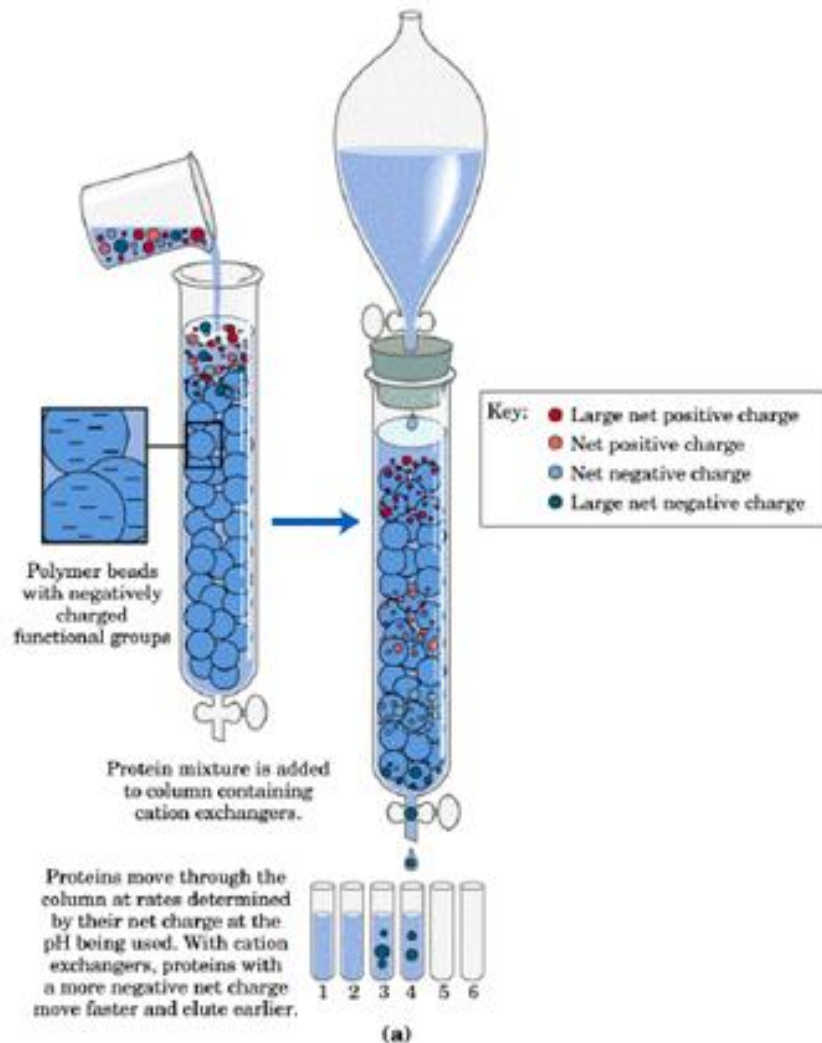


Figura 9-29. Cromatografía de intercambio iónico

Cromatografía basada en bioafinidad

Este tipo de cromatografía se basa en la capacidad de las sustancias biológicamente activas de formar complejos estables, específicos y reversibles. La formación de estos complejos implica la participación de fuerzas moleculares como las interacciones de *Van der Waals*, electrostáticas, dipolo-dipolo, hidrofóbicas así como puentes de hidrógeno entre las partículas de la muestra y la fase estacionaria.

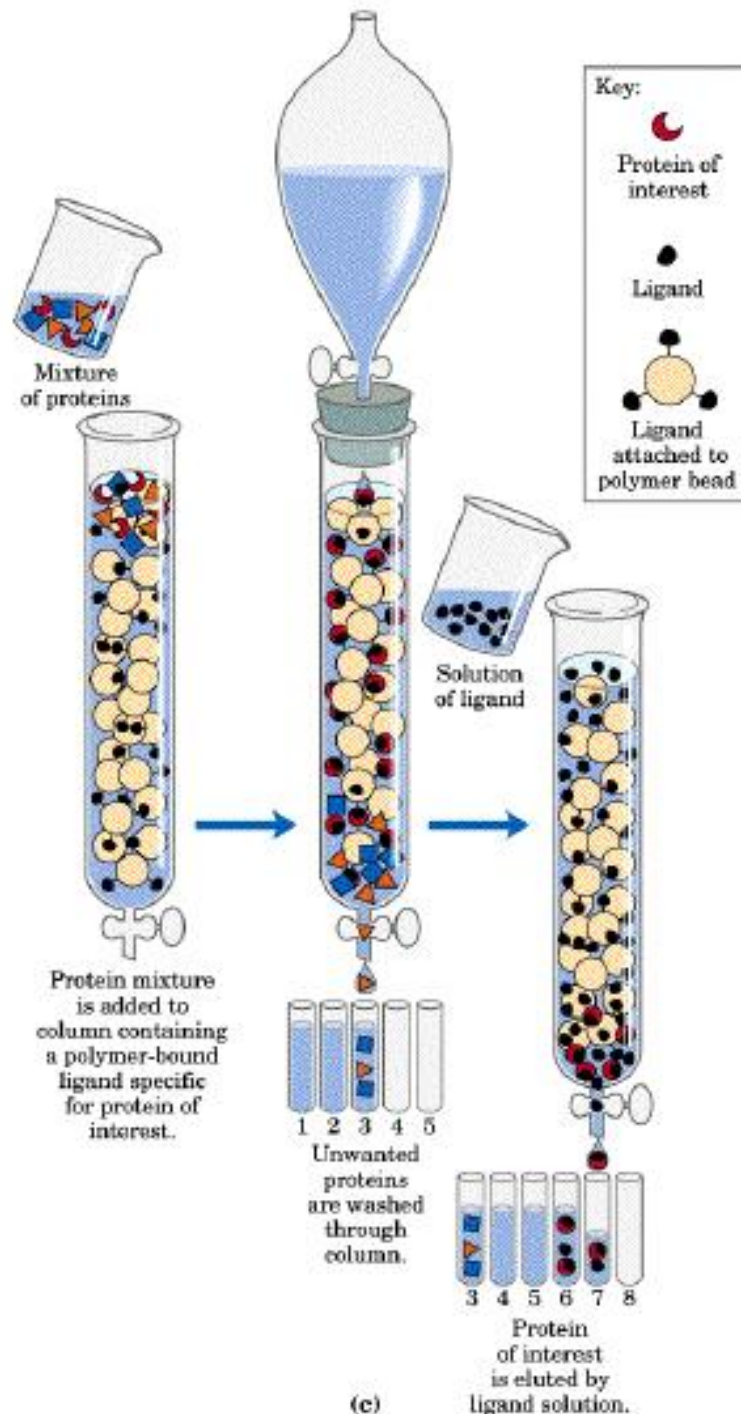


Figura 9-30. Cromatografía de biofinidad, tomado de www.unizar.es/departamentos/bioquimica_biologia/docencia/FMBvirtual/PurProt/Croma.htm

Tarea 9-14:

Realice diagramas explicando cada una de las cromatografías, con un ejemplo ya sea que tomes de la literatura o que inventes. Explica, si quisiera medir, azúcares, carotenos, nucleótidos o antocianinas ¿Cuál de las cromatografías escogería para cada familia de compuestos y por qué?

En resumen la gran variabilidad de los métodos cromatográficos se debe a las diferentes condiciones que pueden utilizarse para separar los componentes de una sustancia.

Todas estas técnicas tienen en común la existencia de una fase estacionaria a través de la cual fluye una fase móvil, así como la presencia de un mecanismo de inyección de la muestra y un mecanismo de detección de los diferentes componentes separados.

Cromatografía de fase normal

La cromatografía de fase normal (NP-HPLC) fue el primer tipo de sistema HPLC utilizado en el campo de la química, y se caracteriza por separar los compuestos en base a su polaridad. Esta técnica utiliza una fase estacionaria polar y una fase móvil apolar, y se usa cuando el compuesto de interés es bastante polar. El compuesto polar se asocia y es retenido por la fase estacionaria. La fuerza de adsorción aumenta a medida que aumenta la polaridad del compuesto y la interacción entre el compuesto polar y la fase estacionaria polar (en comparación a la fase móvil) aumenta el tiempo de retención.

La fuerza de interacción no sólo depende de los grupos funcionales del compuesto de interés, sino también de los factores estéricos, de forma que los isómeros estructurales a menudo se pueden diferenciar el uno del otro. El uso de disolventes más polares en la fase móvil disminuye el tiempo de retención de los compuestos mientras que los disolventes más hidrofóbicos tienden a aumentar el tiempo de retención.

La NP-HPLC cayó en desuso a los años setenta con el desarrollo del HPLC de fase reversa debido a la falta de reproducibilidad de los tiempos de retención puesto que los disolventes próticos cambiaban el estado de hidratación de la sílice o alúmina de la cromatografía.

Cromatografía de fase reversa

La cromatografía en fase reversa (RPC) permite separar moléculas en base a su polaridad. El principio de la cromatografía en fase reversa es semejante al de la cromatografía en capa fina. Sin embargo, aquí la fase estacionaria es de partículas de sílice químicamente modificadas con hidrocarburos saturados, insaturados o aromáticos de diferentes tipos. Esto convierte a la fase estacionaria en una matriz apolar. Por lo tanto, para este tipo de cromatografías se emplean mezclas de solventes polares, tales como agua, acetonitrilo, acetato de etilo, acetona y alcoholes alifáticos.

Por tal razón, este tipo de cromatografía ha sido también llamada como cromatografía de interacción hidrofóbica (HIC). En general, este último término se ha empleado para referirse a las aplicaciones en las que se emplean sustituyentes y matrices compatibles con fluidos biológicos. Por tanto, el término HIC es común cuando se habla de purificación de proteínas y carbohidratos, en tanto que RPC es más empleado para referirse

a la separación de moléculas en sistemas que son inadecuados para la separación de macromoléculas biológicas.

La HPLC de fase reversa (RP-HPLC) consiste en una fase inmóvil apolar y una fase móvil de polaridad moderada. Una de las fases estacionarias más comunes de este tipo de cromatografía es la de sílice tratada con RMe_2SiCl , donde la R es una cadena alquil tal como $\text{C}_{18}\text{H}_{37}$ ó C_8H_{17} . El tiempo de retención es mayor para las moléculas de naturaleza apolar, mientras que las moléculas de carácter polar eluyen más rápidamente.

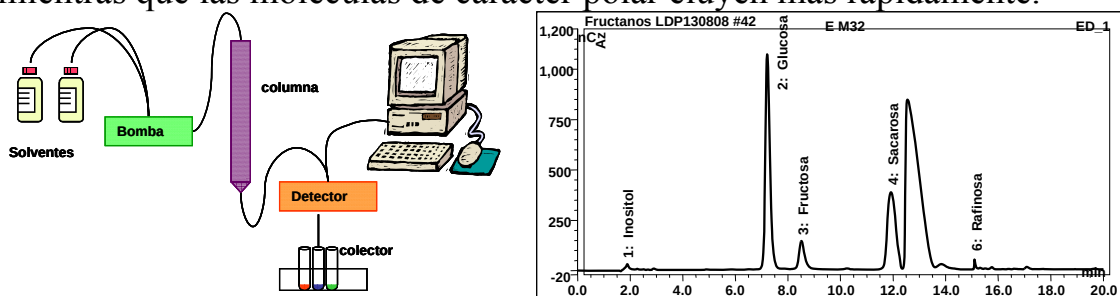


Figura 9-31. Diagrama del funcionamiento de la cromatografía de fase reversa. Y un cromatograma típico.

El tiempo de retención aumenta con la adición de disolvente polar a la fase móvil y disminuye con la introducción de disolventes más hidrofóbicos. La cromatografía de fase reversa es tan utilizada que a menudo se lo denomina HPLC sin ninguna especificación adicional. La cromatografía de fase reversa se basa en el principio de las interacciones hidrofóbicas que resultan de las fuerzas de repulsión entre un disolvente relativamente polar, un compuesto relativamente apolar, y una fase estacionaria apolar. La fuerza conductora en la unión del compuesto a la fase estacionaria es la disminución del área del segmento apolar del analito expuesto al disolvente. Este efecto hidrofóbico está dominado por la disminución de la energía libre de la entropía asociada con la minimización de la interfase compuesto-disolvente polar. El efecto hidrofóbico disminuye con la adición de disolvente apolar a la fase móvil. Esto modifica el coeficiente de partición de forma que el compuesto se mueve por la columna y eluye.

Las características del compuesto de interés juegan un papel muy importante en la retención. En general, un compuesto con una cadena alquil larga se asocia con un tiempo de retención mayor porque aumenta la hidrofobicidad de la molécula. Aun así, las moléculas muy grandes pueden ver reducida la interacción entre la superficie del compuesto y la fase estacionaria. El tiempo de retención aumenta con el área de superficie hidrofóbica que suele ser inversamente proporcional al tamaño del compuesto. Los compuestos ramificados suelen eluir más rápidamente que sus isómeros lineales puesto que la superficie total se ve reducida.

Aparte de la hidrofobicidad de la fase móvil, otras modificaciones de la fase móvil pueden afectar la retención del compuesto; por ejemplo, la adición de sales inorgánicas provoca un aumento lineal en la tensión superficial, y como que la entropía de la interfase compuesto-disolvente está controlada precisamente por la tensión superficial, la adición de sales tiende a aumentar el tiempo de retención.

Otra variable importante es el pH, puesto que puede cambiar la hidrofobicidad del compuesto. Por este motivo, la mayoría de métodos utilizan un tampón como el fosfato de sodio por controlar el valor del pH. Estos tampones controlan el pH, pero también neutralizan la carga o cualquiera resto de sílice de la fase estacionaria que haya quedado expuesta y actúan como contraiones que neutralizan la carga del compuesto. El efecto de los tampones sobre la cromatografía puede variar, pero en general mejoran la separación cromatográfica.

Las columnas de fase reversa se echan a perder con menor facilidad que las columnas de sílice normales. Aun así, muchas columnas de fase reversa están formadas por sílice modificada con cadenas alquil y no se deben utilizar nunca con bases en medio acuoso puesto que estas podrían dañar el esqueleto de sílice subyacente. Las columnas se pueden utilizar en medio acuoso ácidos, pero no deberían estar expuestas demasiado tiempo al ácido porque puede corroer las partes metálicas del aparato de HPLC.

9.4.2. Cromatografía de gases (GC)

La cromatografía de gases (GC por sus siglas en inglés) es una técnica cromatográfica en la que la muestra se volatiliza y se inyecta en la cabeza de una columna capilar. La elución se produce por el flujo de una fase móvil de gas inerte. A diferencia de los otros tipos de cromatografía, la fase móvil no interacciona con las moléculas del analito; su única función es la de transportar el analito a través de la columna.

Existen dos tipos de cromatografía de gases (GC): la cromatografía gas-sólido (GSC) y la cromatografía gas-líquido (GLC). Esta última se utiliza más ampliamente, y que se puede llamar simplemente cromatografía de gases (GC). En la GSC la fase estacionaria es sólida y la retención de los analitos en ella se produce mediante el proceso de adsorción. Precisamente este proceso de adsorción, que no es lineal, es el que ha provocado que este tipo de cromatografía tenga aplicación limitada, ya que la retención del analito sobre la superficie es semipermanente y se obtienen picos de elución con colas. Su única aplicación es la separación de especies gaseosas de bajo peso molecular. La GLC utiliza como fase estacionaria moléculas de líquido inmovilizadas sobre la superficie de un sólido inerte.

La GC se lleva a cabo en un cromatógrafo de gases. Este consta de diversos componentes como el gas portador, el sistema de inyección de muestra, la columna (generalmente dentro de un horno), y el detector.

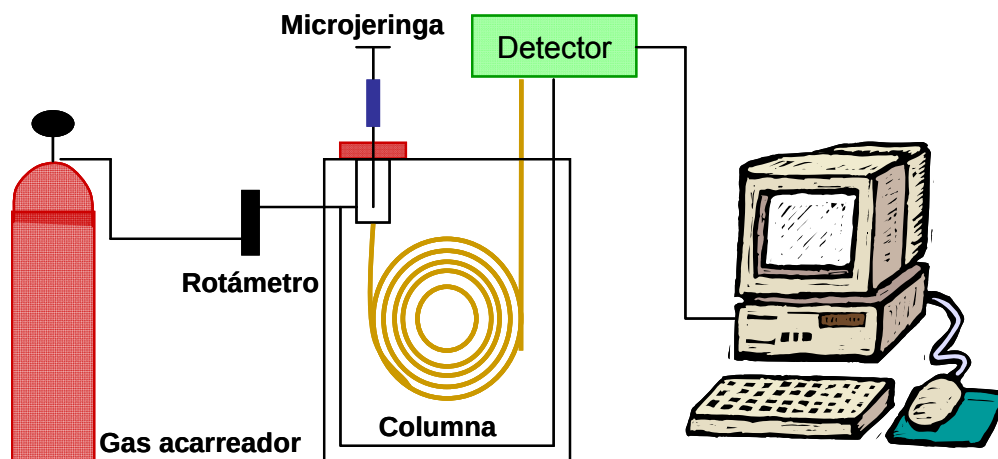


Figura 9-32. Diagrama de un cromatógrafo de gases.

Rampas de temperatura

La temperatura es una variable importante, ya que de ella va a depender el grado de separación de los diferentes analitos. Para ello, debe ajustarse con una precisión de décimas de grado. Dicha temperatura depende del punto de ebullición del analito o analitos, y por lo general se ajusta a un valor igual o ligeramente superior a él. Para estos valores, el tiempo de elución va a oscilar entre 2 y 30-40 minutos. Si tenemos varios componentes con diferentes puntos de ebullición, se ajusta la llamada rampa de temperatura, con lo cual esta va aumentando ya sea de forma continua o por etapas. En muchas ocasiones, el ajustar correctamente la rampa puede significar separar bien o no los diferentes analitos. Es recomendable utilizar temperaturas bajas para la elución ya que, aunque a mayor temperatura la elución es más rápida, se corre el riesgo de descomponer el analito.

Columnas capilares

En el GC se emplean dos tipos de columnas: las empaquetadas o de relleno, y las tubulares abiertas o capilares. Estas últimas son más comunes en la actualidad (2005) debido a su mayor rapidez y eficiencia. La longitud de estas columnas es variable, de 2 a 50 metros, y están construidas en acero inoxidable, vidrio, sílice fundida o teflón. Debido a su longitud y a la necesidad de ser introducidas en un horno, las columnas suelen enrollarse en una forma helicoidal con diámetros de 10 a 30 cm, dependiendo del tamaño del horno.

Las columnas capilares son de dos tipos básicos: las de pared recubierta (WCOT) y las de soporte recubierto (SCOT). Las WCOT son simplemente tubos capilares donde la pared interna se ha recubierto con una finísima

capa de fase estacionaria. Las columnas SCOT tienen en su parte interna una fina capa de material adsorbente como el empleado en las columnas de relleno (tierra de diatomeas) donde se ha adherido la fase estacionaria. Las ventajas de las SCOT frente a las WCOT es la mayor capacidad de carga de esta última, ya que en su fabricación se emplean mayores cantidades de fase estacionaria, al ser la superficie mayor de intercambio. Las columnas WCOT se fabrican a partir de sílice fundida, conocidas como columnas tubulares abiertas de sílice fundida o FSOT. Estas columnas se fabrican a partir de sílice especialmente pura, sin apenas contenido de óxidos metálicos. Debido a la fragilidad inherente a este material, en el mismo proceso de obtención del tubo se recubre con una capa de poliimida, de esta forma la columna puede enrollarse con un diámetro de unos pocos centímetros. Estas columnas, con propiedades como baja reactividad, resistencia física y flexibilidad, han sustituido a las WCOT clásicas.

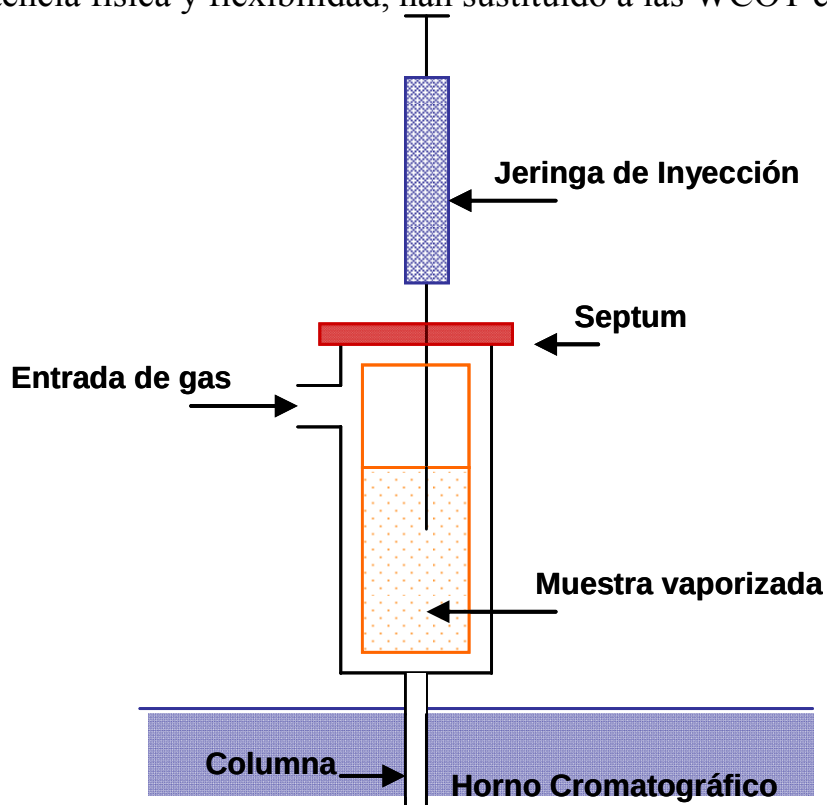


Figura 9-33. Columna capilar.

Las columnas FSOT tienen diámetros internos variables, entre 250 y 320 μm (para columnas normales) y 150-200 μm para columnas de alta resolución. Estas últimas requieren menor cantidad de analito y un detector más sensible, al eluir menor cantidad de gas. Existen columnas macrocapilares con diámetros de hasta 530 μm , que admiten cantidades de analito comparables a las de relleno pero con mejores cánones.

En estas columnas existe un problema debido a la adsorción del analito sobre la superficie de la sílice fundida, adsorción debida a la presencia de grupos silanol (Si-OH), los cuales interaccionan fuertemente con moléculas polares orgánicas. Este inconveniente se suele solventar inactivando la

superficie por sililación con dimetilclorosilano (DMCS). La adsorción debida a los óxidos metálicos se ve paliada en gran parte por la elevada pureza de la sílice empleada. Por orden de eficacia, en primer lugar están las WCOT, luego las SCOT y por último las columnas de relleno.

Detector de ionización de llama, FID

El detector de ionización de llama, es un detector utilizado en cromatografía de gases. Es uno de los detectores más usados y versátiles. Básicamente, es un quemador de hidrógeno/oxígeno, donde se mezcla el efluente de la columna (gas portador y analito) con hidrógeno. Inmediatamente, este gas mezclado se enciende mediante una chispa eléctrica, produciéndose una llama de alta temperatura. La mayoría de compuestos orgánicos al someterse a altas temperaturas, pirolizan y se producen iones y electrones, que son conductores eléctricos. Este hecho se aprovecha estableciendo una diferencia de potencial de unos centenares de voltios entre la parte inferior del quemador y un electrodo colector situado por encima de la llama. La corriente generada es baja (del orden de los 10-12 $\text{\AA}$), por lo tanto debe ser amplificada mediante un amplificador de alta impedancia.

Detector por ionización de flama

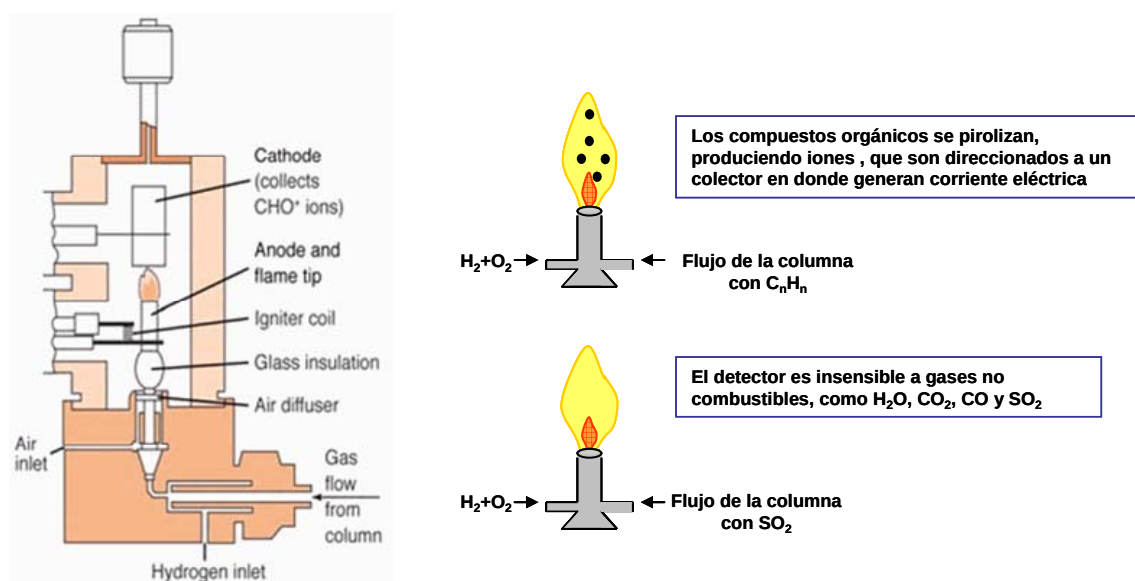


Figura 9-34. Detector por ionización de flama y formación de iones cuando un compuesto orgánico es quemado en una flama de H_2 y O_2

El proceso de ionización que se da en la flama es complejo, pero se puede aproximar el número de iones producidos al número de átomos de carbono transformados en la flama. Esto produce que sea un detector sensible a la masa (al número de átomos de carbono que salen de la columna), más que a

la concentración, por lo tanto no le afectan demasiado los cambios en el flujo de salida.

Existen algunos grupos funcionales que no dan respuesta en este detector, como el carbonilo, alcohol, halógeno o amina, y tampoco responden gases no inflamables como el CO₂, SO₂, agua y óxidos de nitrógeno. Este hecho, más que limitar el ámbito de aplicación de este detector, permite el análisis de muestras contaminadas con alguno de los compuestos mencionados.

Ventajas:

- Alta sensibilidad, del orden de 10⁻¹³ g/s.
- Amplio intervalo lineal de respuesta, 10⁷ unidades.
- Bajo ruido de fondo (elevada relación señal/ruido).
- Bajo mantenimiento, fácil de fabricar.

Desventajas:

- Destruye la muestra (la piroliza).

9.4.2.1. Ejemplo: Análisis de alcoholes y ácidos por GC-FID

El control de calidad por análisis de laboratorio usualmente se efectúa en las reservas alimenticias, en materiales procesados y en los productos terminados. La cromatografía líquida HPLC se usa comúnmente para analizar materiales durante el proceso de fermentación, para monitorear la ruptura de las moléculas de fécula en glucosa, luego la conversión a etanol. El etanol en presencia de oxígeno puede ser convertido en ácido acético. Esto es la forma en que se produce el vinagre a partir de jugo de caña de azúcar. El análisis de alcohol por GC-FID no es el mismo que se hace con el alcoholímetro. EL GC-FID permite distinguir diferentes alcoholes y el ácido acético por sus tiempos de retención.

9.4.3. Ejemplos de metodologías usadas para el mejoramiento de maíz

9.4.3.1. Análisis de triptófano

El maíz es uno de los cultivos más importantes de América, por el área cultivada, el valor de la cosecha y su amplia utilización en la alimentación. Se ha establecido que alrededor del 70% del suministro mundial de proteínas es de origen vegetal y aproximadamente el 30% es de origen animal. Los cereales, son los alimentos de principal consumo en la gran mayoría de los países en desarrollo y aportan a la dieta promedio el 80% de las calorías totales y el 70% de las proteínas requeridas. El desarrollo de los híbridos QPM, con un alto contenido de aminoácidos esenciales (lisina y

triptófano), presenta nuevas posibilidades y esperanzas para reducir las deficiencias nutricionales de las familias de bajos ingresos en los países subdesarrollados, en los cuales el maíz representa una parte muy importante en la alimentación. En la actualidad se conocen genes recesivos, por ejemplo, el gen *opaque2*, cuya incorporación a variedades corrientes permiten mejorar sustancialmente la calidad de las proteínas en dichas variedades.

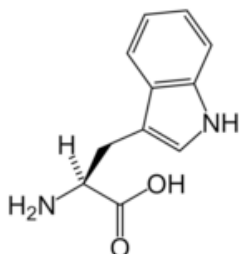


Figura 9-35. Molécula de triptófano.

La cuantificación de triptófano es difícil debido a la inestabilidad del triptófano en condiciones ácidas. Los métodos convencionales de detección de los demás aminoácidos principalmente se basan en la hidrólisis ácida de las proteínas. Esto se hace a 110 °C -125 °C de 6 a 24 horas con ácido clorhídrico 6 a 7N. Hay evidencia de que la destrucción de triptófano durante la hidrólisis ácida con ácido clorhídrico, se debe a que este es oxidado por el enlace disulfuro de la cisteína, ya que se ha observado que hay un incremento de esta en hidrolizados de proteínas que contienen cistina y triptófano (Olocott y Fraenkel-Conrat, 1947). En ausencia de oxígeno, el triptófano es relativamente estable en ácidos calientes ya sea solo o en presencia de otros aminoácidos (Lugo, 1933). Sin embargo, cuando están presentes cistina u otros bisulfatos, el triptófano es destruido. La destrucción de triptófano también se ha observado cuando es calentado en solución ácida con ácido pirúvico. La reacción de triptófano con ácido pirúvico, así como la reacción con carbohidratos, resulta en soluciones negras o cafés, y precipitados (huminas) dependiendo de la cantidad de triptófano destruido. En cambio la reacción con cistina promueve la destrucción de triptófano sin la formación de huminas y las soluciones permanecen claras (Olocott y Fraenkel-Conrat, 1947).

Los métodos desarrollados hasta ahora no están bien adaptados para una determinación rutinaria en muestras de alimentos vegetales. Para detectar y cuantificar triptófano en mezclas complejas en general, es necesario seguir algunos pasos, dependiendo del método a emplear. La mayoría de las técnicas requieren de una previa extracción pues el triptófano debe de estar en forma libre (no formando parte de una cadena polipeptídica) para poder ser detectado, (algunos casos de métodos colorimétricos). Después de la extracción puede seguir un proceso de separación (principalmente cromatográfica), y por último, un proceso de detección y cuantificación.

Tarea 9-15

Haga una búsqueda bibliográfica sobre las publicaciones sobre el gen *opaque 2* y los maíces QPM que han sido liberados por el CIMMYT.

Tarea 9-16

Para un fitomejorador que quiere generar variedades de maíz con mayor calidad de proteínas, ¿Qué tan importante es el método analítico que va a utilizar para medir esas características nutricionales? Haga un listado de lo que esperaría de un servicio analítico en términos de: rapidez, cercanía, confiabilidad, número de muestras que pueden procesar, precio, garantía, etc. ¿Cuáles serían sus prioridades como fitomejorador? ¿Qué porcentaje de su presupuesto utilizaría para trabajo de campo, qué tanto para marcadores moleculares, y qué tanto para análisis bioquímicos?

Métodos de extracción**a) Hidrólisis ácida**

Para extraer triptófano por hidrólisis ácida existen varias técnicas, algunas recurren a la protección del anillo indólico contra la oxidación que sufre por la hidrólisis ácida. Algunos compuestos protectores son *p*-tolueno, ácido sulfónico (Liu & Chang, 1971), ácido tioglicólico (Gruen & Nicholls, 1972; Matsubara & Sasaki, 1969), β -Mercaptoetanol (Pascaud *et al.*, 1987), ácido 3,3 indoilpropiónico y otros reactivos (Penke *et al.*, 1974). Sin embargo, la utilización de estos compuestos no da buenos porcentajes de recuperación de triptófano, además, estas técnicas sólo han sido utilizadas para determinar triptófano en proteínas puras, ya que algunos compuestos del grano suelen disminuir su poder protector como son los carbohidratos resultantes de la hidrólisis del almidón (Piombo y Lozano, 1980). Alternativamente se han usado otros compuestos para hidrolizar en lugar del HCl, que mejoran la recuperación, como el ácido metanosulfónico conteniendo 3-(2- aminoetil) indol, y el ácido mecaptometanosulfónico (Creamer & Matheson, 1976).

b) Hidrólisis básica

Se ha optado también por la hidrólisis básica utilizando NaOH en concentraciones de 4 a 6 M y protegiendo con algunos compuestos para su degradación como stanita (Lugg, 1938) con acetato e histidina (Spies, 1967) o con almidón (Dreze, 1960). Con la extracción básica se llegan a obtener mejores porcentajes de recuperación aunque involucran procedimientos más elaborados: Si se va a utilizar NaOH es necesaria una predigestión de almidón pues este suele gelificar en soluciones básicas. También son necesarias condiciones de vacío durante la hidrólisis en tubos de polipropileno (Oelshlegel, 1970) para evitar la oxidación del triptófano

(Spies, 1967; Oelshlegel, 1970) y calentando a temperaturas de aproximadamente 135 °C (Hugli & Moore, 1972).

El uso de de Ba (OH)₂ puede ayudar a evitar la formación de gel por el almidón del grano, pero también puede formar compuestos que co-precipitan al triptófano como BaSO₄ o BaCO₃ (Robel, 1967).

c) Amortiguadores de proteínas totales

La extracción de proteínas también puede ser llevada a cabo con los llamados amortiguadores (buffers) de proteínas totales, que pueden contener principalmente agentes caotrópicos como la urea, además de detergentes y agentes reductores como dithiothreitol (DTT) o β-Mercaptoetanol entre otros compuestos. Sin embargo, algunos métodos de cuantificación no pueden trabajar adecuadamente si el triptófano todavía está unido a una cadena polipeptídica.

d) Hidrólisis enzimática

Y por último están los métodos de extracción con enzimas como la pronasa (Spies, 1967), pepsina, tripsina (Opienska-Blauth *et al.*, 1963) y papaína (Hernández & Bates, 1969; Oste *et al.*, 1976). Las enzimas se encargan de romper los enlaces peptídicos que hay entre aminoácidos siendo algunas de ellas muy específicas. Estas enzimas requieren de una molécula de agua para realizar su función, por eso a las proteasas también se les puede catalogar como hidrolasas. Los tiempos de hidrólisis son muy similares a los usados con la digestión básica. Una desventaja de estos métodos es que la hidrólisis puede no ser completa. La hidrólisis de todos los enlaces peptídicos es importante por ejemplo, para la cuantificación por HPLC, ya que la presencia de oligopéptidos causa grandes cambios en los tiempos de retención, pero no se requiere una digestión completa para métodos espectrofotométricos. Si se usa hidrólisis enzimática y HPLC en combinación, existe el riesgo de no cuantificar todo el triptófano que en realidad hay en la muestra.

Métodos	Características	Ventajas	Desventajas
Hidrólisis ácida	El trp debe ser protegido de su degradación por HCl.	Se puede analizar trp junto con los demás aminoácidos	Bajas recuperaciones. Usados solo para proteínas puras
Hidrólisis básica	NaOH y BaOH ₂ a concentraciones 4 a 6 N	Buenas recuperaciones. Menor degradación de trp.	Gelificación del almidón si se usa NaOH. Digestión al vacío y a altas temperaturas. El Ba (OH) ₂ puede formar BaSO ₄ o

			BaCO ₃ que co-precipita con trp.
Buffers de proteínas totales	Contienen principalmente: Agentes caotrópicos Detergentes Agentes reductores	Eficiencia en la extracción de proteínas.	Algunos compuestos interfieren con algunos métodos de detección.
Hidrólisis enzimática	Proteasas: Pronasa, pepsina, tripsina y papaína	La papaína es barata e inespecífica	Digestión a veces incompleta y por lo tanto no compatible con algunos métodos que requieren que el trp se encuentre en forma libre.

Tabla 9-1. Características de métodos de extracción/digestión

9.4.3.2. Análisis de carotenos

Las frutas y vegetales de la dieta humana contienen entre 20 y 30 carotenoides que se absorben en el plasma. Los carotenoides y los tocoferoles presentan especial interés para los investigadores médicos debido a sus características antimutagénicas y anticarcinogénicas, como también a su asociación con mayor longevidad así como disminución del riesgo de cáncer y enfermedades cardiovasculares observadas en estudios epidemiológicos. Aunque el suplemento con β -carotenos no ha sido asociado con efectos beneficiosos sobre la reducción del riesgo de cáncer o enfermedades cardiovasculares, en estudios recientes se ha señalado que tanto la luteína como la zeaxantina dietarias o plasmáticas elevadas se asociaron con reducción de la progresión de lesiones ateroscleróticas y con menor riesgo de cataratas.

La deficiencia en vitamina A (VAD) causa síntomas que van desde ceguera, xeroftalmia, insuficiencias en el sistema inmune, hasta mortalidad materna e infantil. Este problema es frecuente en zonas como el sudeste de Asia y países de África, donde el arroz sin cáscara, que carece de vitamina A, es la fuente principal de la alimentación. La administración oral de la vitamina A en estos países es problemática debido a la falta de infraestructura y capital, por lo que son requeridas soluciones alternativas (Ye, 2000).

La vitamina A es producida naturalmente por carotenos con actividad provitamina A y el β -caroteno es la más importante provitamina A para los mamíferos.

Se ha reportado una línea genéticamente modificada de arroz, llamada Arroz Dorado, diseñada para sintetizar y acumular provitamina A (β -caroteno) en el endospermo. A partir de esto que se desarrolla el concepto de ingeniería genética basada en el aumento nutricional del arroz para lograr una reducción en la deficiencia de vitamina A (Hoa, 2003).

El desarrollo se basó en la presencia del precursor geranil-geranildifosfato, el cual es capaz de ser sintetizado en el endospermo del arroz salvaje.

Proveyendo mediante ingeniería genética todos los genes necesarios para completar la ruta biosintética de la provitamina A, se obtienen granos de arroz amarillos que contienen altas cantidades de β -caroteno (más de 1.6 $\mu\text{g/g}$) (Ye, 2000).

Se ha mostrado que el uso de una desaturasa de bacteria (*crtI*) disminuye el esfuerzo de transformación ya que esta es capaz de producir ambas desaturaciones (Kaspar, M. 2005). Además, no es necesaria la transformación con el gen Licopeno-ciclasa para la obtención de β -caroteno, esto puede deberse a una expresión constitutiva en el endospermo o una inducción por licopeno de esta enzima (Ye, 2000).

Sin embargo, la línea mas coloreada producida tenía problemas tales como múltiples eventos de integración de los genes foráneos, el marcador de selección era un gen antibiótico y la línea transformada era poco consumida en los países con VAD, todos estos problemas hacían que el producto obtenido fuera indeseable para la desregulación como cultivo transgénico (Hao, 2003).

A partir de esos antecedentes, se desarrolló un proyecto para obtener una nueva línea de Arroz Dorado que satisficiera todos los requerimientos regulatorios para su liberación como producto (Hoa, 2003). La siguiente hipótesis fue; “La transformación de líneas de arroz, mediante técnicas moleculares que cumplan con los requerimientos regulatorios, con los genes *psy* y *crtI*, es suficiente para generar líneas que acumulen provitamina A en el endospermo”. Se transformaron mediante el sistema mediado por *Agrobacterium tumefaciens*, por el patrón de integración simple que produce, dos cultivares de arroz, *Indica* y *Japonica*.

La construcción combina el gen fitoeno-desaturasa (*crtI*), fusionado con un péptido señal de transporte a plástidos bajo el promotor constitutivo 35S

del virus del mosaico de la coliflor y el gen psy, bajo el promotor específico de endospermo Gt-1. El gen pmi (Foso-Man-Isomerasa) fue usado como gen selector, el cual permite la selección de las líneas transformadas, pero no tiene actividad antibiótica.

Las semillas obtenidas en la T1 fueron seleccionadas fenotípicamente por el color visible antes de someterlas a otro tipo de análisis. Las mejores líneas desde el punto de vista fenotípico fueron examinadas para confirmar la ausencia de integración más allá de los bordes derecho e izquierdo del vector. Este ensayo se llevó a cabo utilizando distintas combinaciones de primers y el posterior análisis de amplificación en distintas reacciones de PCR (Reacción en Cadena de la Polimerasa).

La técnica HPLC permite la cuantificación de muchos carotenos simultáneamente. La cantidad de carotenos fue variable entre las distintas líneas. Se obtuvieron valores mas altos con la línea Japónica que con los cultivares Indica y la cantidad parece ser independiente del número de integraciones.

9.4.3.3. Análisis de flavonoides.

La ingeniería genética de flavonoides se ha utilizado para aumentar o suprimir la concentración de flavonoides en las plantas, o bien para sintetizarlos a través de cultivos de bacterias. La ingeniería metabólica de los flavonoides empezó en 1987 (Meyer P, Heidmann I, Forkmann G, Saedler H. 1987.) y ha sido un área de investigación muy fructífera en la década de 1990 (Dixon R, Steele C. 1999.)

La ingeniería metabólica consiste en la introducción o supresión de genes específicos en una planta objetivo. Su realización es cara y trabajosa. Por eso se realiza previamente una caracterización minuciosa de los genes, la ruta biosintética, la especificidad de sustrato de las enzimas concernientes y la disponibilidad de sustratos definidos de la planta objetivo.

El conocimiento profundo de la planta objeto de estudio mejora mucho la probabilidad de éxito para el mejoramiento y la ingeniería genética. Por eso se realizan experimentos previos como la aplicación de inhibidores específicos de las enzimas, que puede proveer información adicional. Existen inhibidores específicos para enzimas dioxigenasas dependientes de 2-oxoglutarato, como son: flavanona 3-hidroxilasa (F3H), flavonol sintasa (FLS), antocianidina sintasa (ANS) y flavona sintasa I (FS I); y para enzimas monooxigenasas dependientes de citocromo P450: flavonoide 3' hidroxilasa (F3'H), flavonoide 3'5' hidroxilasa (F3'5'H), flavona sintasa II (FSII) e isoflavona sintasa (IFS). Suplementar a las plantas con intermediarios de flavonoides que no se encuentran naturalmente en dichas plantas es una manera de probar si la planta tiene un sistema interno de enzimas necesario para convertir dichos

intermediarios en los flavonoides deseados. La aplicación in vivo de inhibidores puede reflejar los resultados de estrategias de antisentido o de supresión.

El desarrollo de estos experimentos previos permite hacer una predicción del resultado final de un experimento de ingeniería metabólica, ahorrando tiempo y dinero. Gracias a estos experimentos previos o aproximaciones podemos hacer también análisis químicos de alteraciones en el patrón de metabolitos, incluyendo la síntesis de nuevos compuestos o, en el caso de cambios de color, incluso inspección ocular.

La innovación en el color de las flores, en particular la generación de variedades azules y amarillas desconocidas, es una de las mayores fuerzas conductoras en el desarrollo del cultivo de plantas ornamentales mediante una combinación de métodos clásicos y moleculares. La ingeniería metabólica es una herramienta muy poderosa para generar nuevos colores de flores sin cambiar otras características preexistentes en una determinada variedad o cultivar.

El éxito conseguido en la síntesis de pelargonidina expresando unos genes adecuados de la enzima dihidroflavonol 4-reductasa (DFR) procedentes de otras plantas, que permite conseguir flores naranjas de *Petunia*, ha conllevado mucho trabajo, incluyendo la introducción de nuevos genes en determinadas especies de plantas, la complementación de mutantes definidos, y la supresión de la síntesis de flavonoides por antisentido o supresión de sentido (Forkmann, G., S. Martens. 2001).

La formación de delfinidina por la introducción del gen *F3'5'H* también ha sido alcanzada en cultivos ciánicos de *Dendranthema*, *Dianthus* y *Rosa*, pero las antocianinas basadas en cianidina y/o pelargonidina aún permanecen en estas variedades.

Un nuevo avance concerniente a la manipulación de la síntesis de chalcona ha llevado a la formación de flores amarillo pálido de *Petunia*. Expresando el gen chalcona queretido-reductasa (CHKR) de alfalfa en una línea aciánica de *Petunia* se permite la síntesis de 6'-deoxichalcona isoliquiritigenina, a través de la acción conjunta de la enzima introducida con la chalcona sintasa (CHS). Como la isoliquiritigenina no es un sustrato de la chalcona isomerasa (CHI) de *Petunia*, cierta cantidad de chalcona se acumula en las flores, provocando una coloración amarilla pálida. Una concentración mayor de chalcona puede llevar a flores verdaderamente amarillas, lo cual podría ser alcanzado por la glicosilación e hidroxilación de la isoliquiritigenina. La introducción de CHKR podría ser exitosa en todas las especies de plantas con alta actividad CHS.

La coloración invariable amarilla de los pétalos de *Forsythia* es causada por la acumulación de carotenoides, sin embargo, algunas antocianinas se forman en los sépalos. Se ha estudiado la ruta de estos flavonoides de los sépalos, revelando que la formación de antocianinas en los pétalos es deteriorada en los

pasos controlados por las enzimas DFR y ANS, sobre todo en este último paso. Introduciendo el gen DFR de *Antirrhinum* y el ANS de *Matthiola* se consigue una variedad transgénica de *Forsythia* con flores ligeramente marrones, debido a la síntesis de algunas antocianinas en un fondo amarillo producido por los carotenoides.

Introduciendo y expresando el gen *diff* en las flores de *Dianthus* que serán también transformadas por el gen F3 '5 'H de Petunia conseguimos unas flores casi negras, debido a la gran acumulación de derivados de delphinidina.

Transformado la variedad de *Dianthus* 'Eliat' con el antisentido del gen *F3H* que codifica para la enzima flavanona 3-hidroxilasa, conseguimos flores amarillo crema, debido a la acumulación de naringenina-chalcona 2'-glucósido.

Además de la supresión de la formación de antocianina, la fragancia media de las plantas transgénicas era significativamente mayor que en las flores control debido a los muy elevados niveles de metilbezoato. Así, suprimiendo la biosíntesis de flavonoides con el F3H antisentido se desvía el flujo metabólico a la biosíntesis de derivados del ácido benzoico.

¿Qué conclusiones se pueden sacar sobre los trabajos previos de ingeniería genética en los flavonoides? A pesar del éxito de los trabajos recientes, es muy difícil generar o suprimir compuestos flavonoides específicos de forma altamente controlada por ingeniería metabólica de la ruta de los flavonoides. El aislamiento y clonación de más genes implicados en la ruta biosintética de los flavonoides y de secuencias promotoras, permitirá conseguir grandes avances. No sólo será posible promover una determinada ingeniería metabólica o flavonoides con funciones terapéuticas en plantas de cultivo, sino que también será concebible desarrollar mejoras en las plantas, tales como un aumento de la resistencia generando o manipulando fitoalexinas, o mejorando la protección a los rayos UV y mejorar la eficiencia de nodulación. Además, un mejor conocimiento de los mecanismos implicados en la co-pigmentación y el pH vacuolar, así como la disponibilidad de respectivos genes y la mejora de sistemas de transformación para plantas ornamentales, hará que la modificación del color de las flores sea más factible.

9.4.3.4. Análisis de lípidos

Figura 9-36. Extracción por solventes y saponificación de lípidos.

9.4.3.5. Análisis de proteínas totales

Los granos de maíz común son fuente alimenticia para humanos y animales domésticos, y contienen en su mayor parte hidratos de carbono (74%), y en menor proporción, proteínas (9%), aceite comestible (3.4%) y fibra (1%) (Paliwal *et al.*, 2001). Análisis bromatológicos de los maíces comunes que llenan actualmente el mercado mundial de granos indican que los niveles de proteína cruda están en la banda de 7.5 a 8.4 % con bajo contenido de aminoácidos esenciales, especialmente de lisina y triptófano; el porcentaje de grasa está en el intervalo de 3.0 a 3.5%; excepción hecha en los maíces altamente especializados, sea para calidad proteica o para alto contenido de aceite, cuyos valores son significativamente más altos que los anteriores (Dale, 1997).

Comparado con otros cereales, el grano de maíz es una fuente importante de energía, pero menor como fuente de proteína, tanto en proporción como en calidad, dada las carencias en aminoácidos esenciales; el porcentaje de lisina en grano es menor a 0.29, y el triptófano no rebasa el 0.07 (Dale, 1997; Araba, 1998, citado en Feed and Grain, 1998).

Desde la mejora genética del maíz, se han desarrollado variedades especializadas tanto para calidad proteica (ricos en lisina y triptófano) como para alto contenido de aceite (de 6 a 9% de grasa cruda), las cuales están en proceso de adopción por los agricultores y usuarios. En México, los maíces de alta calidad proteica (conocidos como QPM, por su denominación en lengua inglesa: *quality protein maize*) se han venido promoviendo durante los últimos seis años con éxito limitado todavía; los maíces de alto contenido en aceite (conocidos como HOC: *high oil corn*) no ingresan al mercado nacional de semillas por su origen transgénico, así como las variantes en el manejo especializado para su producción.

La condición gemelar o poliembrionía en semillas de maíz es una característica natural que puede ser aprovechable como una vía alterna en el diseño de variedades de aplicación especial, buscando además de potencial de rendimiento, el valor nutritivo del grano, incrementando cantidad y calidad de aceites y proteína; esto, bajo la hipótesis de que dos o más embriones por semilla, permitirán incrementar la capacidad de almacenamiento de nutrientes de calidad.

9.4.3.6. Espectroscopía en el infrarrojo cercano (NIR)

La espectroscopía en el infrarrojo cercano (*Near Infrared Spectroscopy NIRS*) es un complemento de las técnicas analíticas tradicionales. El NIRS se fundamenta en un principio según el cual, al irradiar con un haz de luz monocromática los materiales orgánicos, estos (en función de la naturaleza de los enlaces y cargas electrostáticas existentes entre sus átomos y

moléculas), absorben una determinada cantidad de energía. La absorción de energía en el rango NIR se debe a las vibraciones de los diferentes tipos de enlaces C-O, C-H y C-N. El espectro NIR es un poco más complejo, ya que contiene bandas de absorción sobrelapadas que corresponden a vibraciones simples y múltiples de los diferentes enlaces.

La absorción puede expresarse en términos de la reflectancia (log reflectancia estándar/ reflectancia de la muestra) a partir de lo cual se logra un espectro característico de cada material que refleja su composición química. La cuantificación por medio de NIRS se fundamenta en la quimiometría, es decir, la aplicación de las matemáticas a la química analítica. Esta técnica combina la espectroscopía, la estadística, y la computación, generando modelos matemáticos que relacionan la composición química (presencia de grupos químicos activos) con cambios de energía en la región correspondiente al rango infrarrojo cercano (longitudes de onda entre 800 y 2500nm). Las ventajas de esta técnica son: proveer información acerca de un metabolito en segundos, ser un método no destructivo, requiere un mínimo o ningún tratamiento de la muestra, minimiza el daño ambiental y es una técnica multianalítica de alta precisión que permite predecir varios factores simultáneamente. Una vez calibrado el espectrofotómetro, el uso del NIRS redundará en bajos costos de análisis para los usuarios que los requieran, por lo que, a nivel internacional, es una metodología que tiene amplia acogida.

La aplicación de las ecuaciones de calibración para el NIRS es un procedimiento sencillo, pudiéndose combinar estas calibraciones dentro de un mismo programa, de manera que con un único barrido de cada muestra, se pueden realizar simultáneamente todos los análisis citados, facilitándose el análisis de materia seca, ceniza, lípidos, proteína, fibra, humedad, etc.

El método de selección de ecuaciones de calibración desarrollado en el Handbook of agriculture No. 643 (ARS, agricultural research service, USDA, Washington, DC) para el análisis de forrajes por NIR establece que para cada componente, el error estándar de calibración (SEC) y el coeficiente de determinación (RSQ) de cada ecuación de calibración, representan criterios importantes para tomar una decisión respecto de la selección de la ecuación.

En la calibración, un grupo de **espectros** y un grupo de datos químicos se usan para derivar el coeficiente de correlación, mientras que en la validación y la subsecuente predicción, un grupo de espectros y la derivada del coeficiente de correlación se utilizan para predecir un dato químico desconocido.

En la siguiente sección se describe en inglés el principio de espectroscopia de infrarrojo cercano:

Near infrared spectroscopy (NIRS) is based on molecular overtone and combination vibrations. Such transitions are forbidden by the selection rules of quantum mechanics. As a result, the molar absorptivity in the near IR region is typically quite small. One advantage is that NIR can typically penetrate much farther into a sample than mid infrared radiation. Near infrared spectroscopy is therefore not a particularly sensitive technique, but it can be very useful in probing bulk material with little or no sample preparation.

The molecular overtone and combination bands seen in the near IR are typically very broad, leading to complex spectra; it can be difficult to assign specific features to specific chemical components. Multivariate (multiple wavelength) calibration techniques (e.g., principal components analysis or partial least squares) are often employed to extract the desired chemical information. Careful development of a set of calibration samples and application of multivariate calibration techniques is essential for near infrared analytical methods.

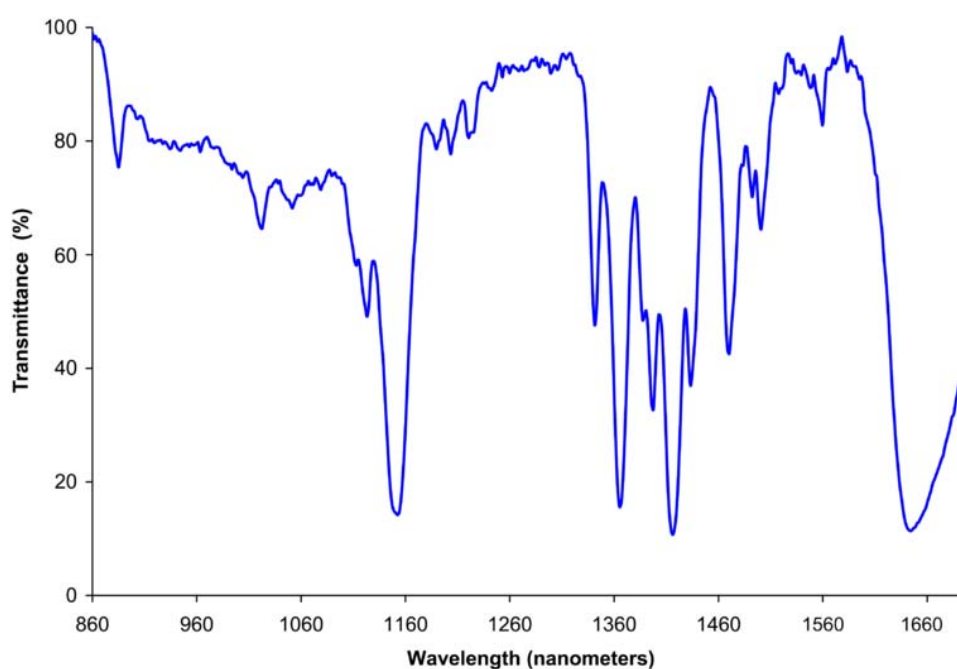


Figura 9-37. Ejemplo de un espectro en el rango NIR

Tarea 9-17

¿En qué principios está basada la metodología NIR? Haga una búsqueda en la biblioteca y averigüe los detalles.

10. Fisiología vegetal

10.1. Introducción

La fisiología vegetal es fundamental para el mejoramiento genético de un cultivo como el maíz. Se requieren conocimientos profundos sobre los distintos procesos que ocurren dentro de las células vegetales para plantear estrategias adecuadas de selección fenotípica y genética. Por ejemplo, para incrementar la tolerancia a sequía de un cultivo, primero hay que entender cuales son las partes de la planta que están involucradas. ¿Qué procesos metabólicos están afectados? A qué órgano hay que darle más atención: ¿a la raíz, a las hojas, al tallo o a la mazorca? ¿Qué es más importante bajo sequía? ¿El metabolismo en los órganos fotosintéticos o en los órganos reproductores? ¿Que es mas limitante? ¿El metabolismo biosintético (anabólico) o el metabolismo degradativo (catabólico)? ¿Qué encimas participan? ¿Qué hormonas podrían estar involucradas en el crecimiento vegetal? ¿Qué genes se pueden mejorar? ¿Cuáles son los alelos superiores? Sólo así se pueden identificar las limitantes y seleccionar de manera apropiada los genotipos que se comportan mejor bajo ese tipo de estrés.

Antes de convertirse en un fitomejorador, primero hay que dominar todos los temas de la fisiología vegetal. Estos incluyen tópicos tan amplios como: fotosíntesis, transporte vascular, xilema, floema, hormonas, etc. Sin estas bases, no se puede llegar muy lejos. Es decir, sin fundamentos sólidos no se puede construir algo perdurable. Es como diseñar un coche de carreras sin antes conocer los principios básicos de la mecánica. Si no sabe nada del fenómeno de combustión, ¿cómo decidir si es mejor inyectarle más o menos oxígeno a los cilindros? Si no sabemos de aerodinámica, ¿cómo podríamos diseñar el prototipo más veloz?

Los estudiantes que quieran convertirse en buenos fitomejoradores, no pueden conformarse con ser solo técnicos de campo con conocimientos limitados. También tienen que aprender ciertos principios científicos, para así evitar trabajar sólo a base de prueba y error. Como un chimpancé tratando de componer un poema sin saber usar bien la maquina de escribir. Si lo intenta suficientes veces durante millones de años, tarde o temprano logrará escribir algún soneto que rime. Lo escribirá pero no sabrá como lo logró, ni lo podrá repetir la próxima vez. El que quiera ser más creativo y eficiente que un mono, le conviene primero aprender el vocabulario y dominar algo de gramática y mecanografía, antes que escribir una novela.

En este capítulo se dará una recapitulación de los fundamentos y los principales temas de la fisiología vegetal, con un enfoque tanto de ciencia básica, como aplicada. Pero antes, es necesario explicar un **paradigma de**

la **fisiología vegetal** por medio de una analogía. ¿Qué es lo que estudia la fisiología vegetal? Digamos que estudia la **función** de las células **autotróficas** y **heterotróficas**, y la interacción que existe entre los distintos órganos vegetales que forman una planta. Pero ¿Qué son las interacciones?

Tal vez es más fácil explicar las interacciones en el ámbito ecológico. Por ejemplo, la **ecología** se enfoca a estudiar las interacciones que existen entre diversos organismos y especies de seres vivos. Por ejemplo, se estudian las interacciones ecológicas entre plantas, microorganismos e insectos. El tipo de interacción puede ser benéfica como maligna, por lo que se distingue la **simbiosis**, **mutualismo**, y la **patogénesis**. Por ejemplo, los insectos que interactúan con plantas pueden ser tan benéficos como los polinizadores que ayudan a la reproducción sexual de las plantas, o tan malignos como los herbívoros que se las comen y las matan. La forma que la planta responde a los patógenos se le llama mecanismo de defensa, que puede ser **susceptibilidad**, **resistencia** o **tolerancia**. La forma que la planta se acopla a las bacterias u hongos benéficos se le llama mecanismos de simbiosis. Todas esas interacciones forman parte de la ecología.

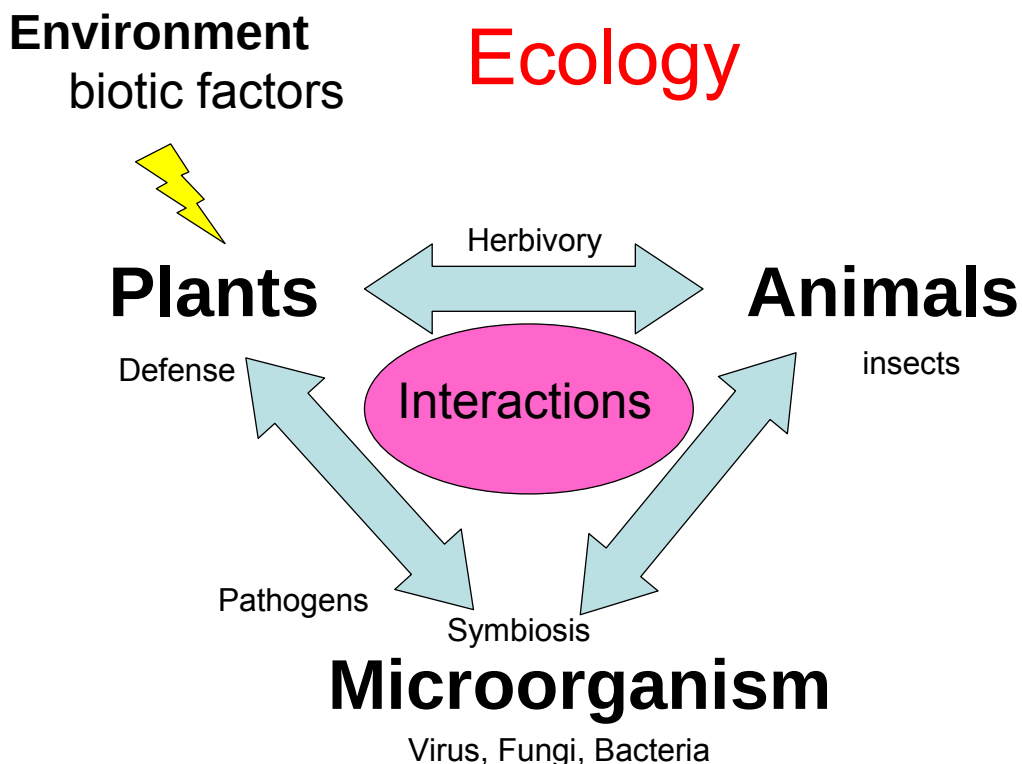


Figura 10-1. La ecología estudia las interacciones entre diferentes especies en un determinado ambiente. Estas interacciones están influenciadas por la luz, la temperatura, la humedad, el pH, la disponibilidad de minerales, etc. A los factores abióticos y los factores bióticos se les agrupa como factores ambientales. Estos factores son muy importantes para el rendimiento agronómico de un cultivo como el maíz.

Ahora bien, la fisiología vegetal también estudia las interacciones entre diferentes partes, pero en lugar de especies diferentes, es la misma planta con sus distintos órganos, como las hojas, el tallo, la raíz, los frutos, etc.

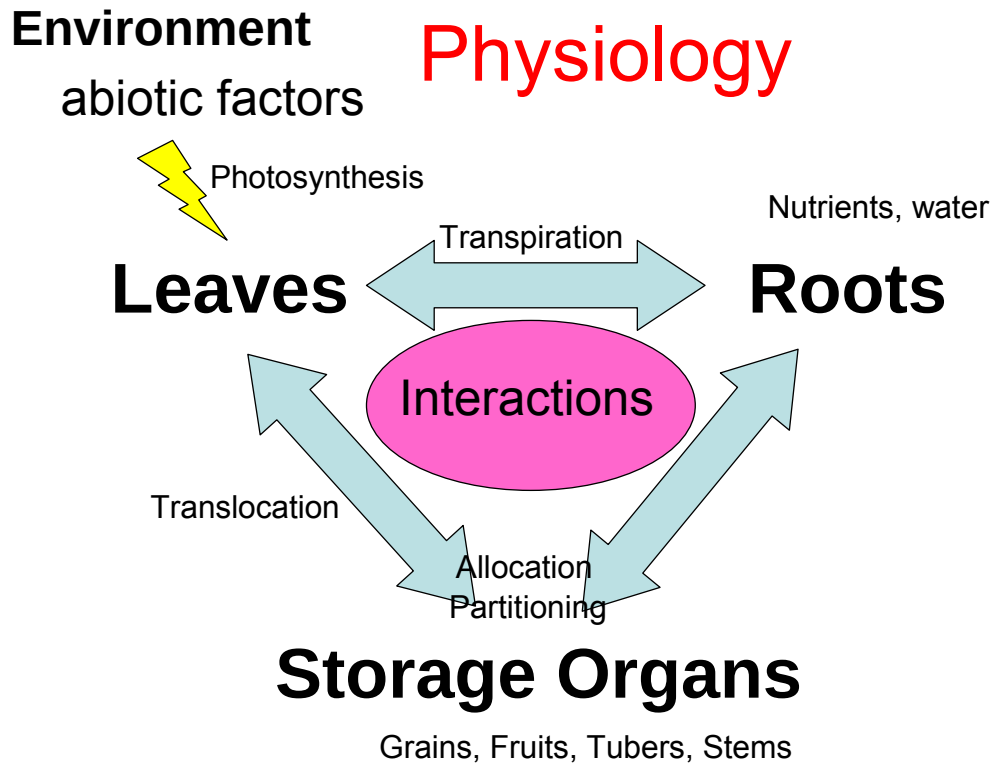


Figura 10-2. La fisiología estudia las interacciones entre diferentes órganos de la planta en un determinado ambiente. Estas interacciones están influenciadas por la luz, la temperatura, la humedad, el pH, la disponibilidad de minerales, etc. Como se puede apreciar al comparar con la figura anterior, la ecología y la fisiología son muy parecidas porque ambas estudian las interacciones internas o externas de la planta. En algunos casos, la ecología se enfoca más a los factores bióticos, mientras que la fisiología más a los factores abióticos. Ambos, tanto la ecología como la fisiología, son relevantes para incrementar los rendimientos del maíz.

Para la fisiología es de suma importancia entender las funciones que tiene cada una de las órganos de la planta (ver Figura 10-3).

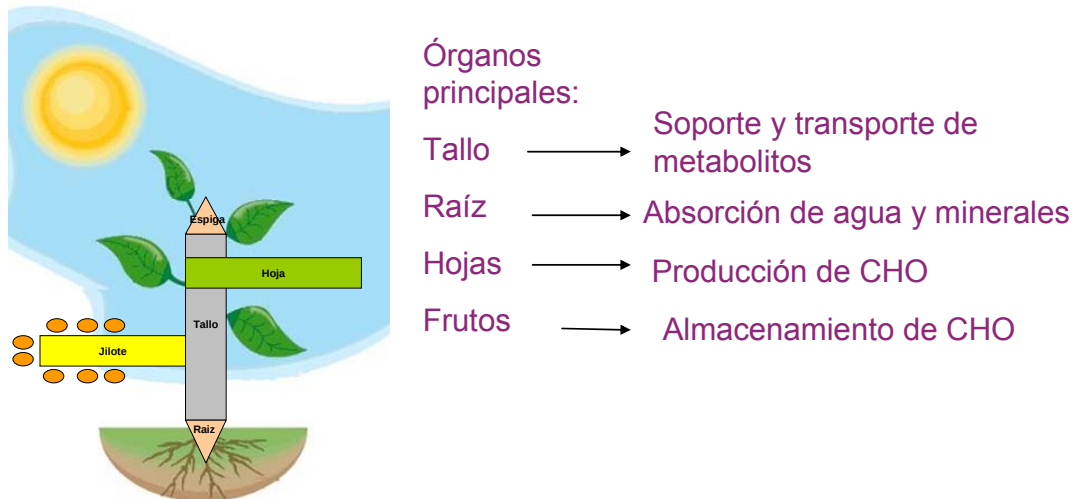


Figura 10-3. Los diferentes órganos de la planta.

También es muy importante entender el concepto fisiológico de fuente (*source*) y sumidero (*sink*). Los **tejidos fuente** son aquellos órganos fotosintéticamente activos que producen carbohidratos y los aportan para otros órganos y para sí mismos. Los **tejidos sumidero** son órganos que demandan los carbohidratos que se producen en otros órganos. El carbohidrato principal que se transporta en el floema es la sacarosa. Pero existen indicios de que otros azúcares y metabolitos también se presentan abundantemente en los tejidos vasculares.

Sink-Source Model

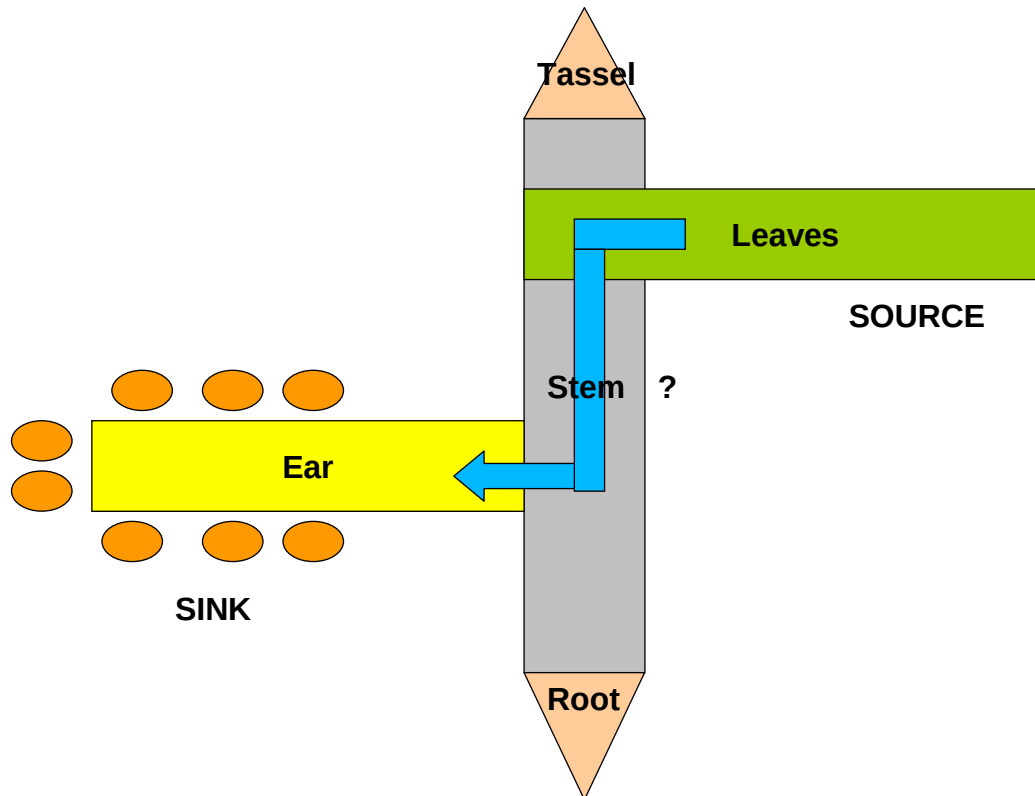


Figura 10-4. Modelo fisiológico *sink-source* de los diferentes órganos de una planta de maíz. La sacarosa, algunos aminoácidos y otros nutrientes se transportan de un tejido a otro por medio del floema. En el caso del maíz, las hojas verdes son el tejido fuente, mientras que las raíces, la mazorca y la espiga son tejidos sumidero. El tallo, es un tejido intermediario, ya que importa, acumula, degrada, sintetiza y exporta azúcares al mismo tiempo. Los granos que se forman en la mazorca son lo que importan y acumulan el almidón de reserva. Las semillas son los individuos de la siguiente generación. Reciben el material genético de ambos progenitores, pero no están conectados de manera intracelular por plasmodesmos, por lo que se dice que no hay conexiones simplásticas, sino solo se comunican y transfieren nutrientes e información por medio del apoplasto.

10.2. Asimilación de carbono

10.2.1. La fotosíntesis

10.2.1.1. La evolución de la foto-autotrofia

En el comienzo de la vida, la tierra estuvo dominada por diferentes tipos de bacterias. Algunas eran heterotróficas mientras que otras tenían un metabolismo autotrófico. Las bacterias fotoautotróficas podían utilizar la energía de la luz para producir su propio alimento. Las cianobacterias aparecieron hace 2.7 millones de años, como una combinación genómica de dos grupos de bacterias: las que tenían un fotosistema del tipo uno (bacterias verdes), con las que tenían un fotosistema tipo dos (bacterias púrpuras). Las bacterias que solo tenían un tipo de fotosistema sólo podían usar H_2S como donador de electrones para la fijación y reducción de dióxido de carbono. Las cianobacterias fueron las primeras que lograron acoplar los dos tipos de fotosistemas de manera seriada para un mismo proceso de fotofosforilación. Esta innovación biológica permitió por primera vez usar agua (H_2O) como donador de electrones para la fotosíntesis. Esto tuvo como consecuencia que en lugar de liberarse S_2 a la atmósfera como antes sucedía, las cianobacterias empezaron a liberar otro gas, el O_2 . En un millón de años aproximadamente, las cianobacterias incrementaron el nivel de oxígeno de la atmósfera hasta un 5%, casi la cuarta parte del nivel actual (21% O_2).

La razón por la que el incremento de oxígeno tardó tantos años es que en la tierra primitiva había mucho hierro en su forma reducida Fe^{2+} que reaccionaba con el oxígeno, oxidándose así a Fe^{3+} . Sólo hasta que se terminaron de oxidar miles de millones de toneladas de hierro fue cuando el oxígeno atmosférico incrementó más. Hoy en día, todavía podemos admirar el legado de las primeras cianobacterias fotosintéticas en forma de diversos colores. El pigmento que utilizan las cianobacterias, la clorofila, refleja la parte media del espectro, por lo que es de color verde. El oxígeno que se libera de la fotosíntesis a partir de agua, causa una difracción especial de la luz solar en la atmósfera. Es la razón por la que nuestro cielo es de color azul. Por si fuera poco, el color rojizo de algunas suelos se debe al óxido ferroso que se produjo por el oxígeno, también por culpa de las cianobacterias. Podemos decir entonces que algunos de los colores más característicos de nuestro planeta se deben precisamente a la fotosíntesis, que inventaron por primera vez las cianobacterias hace millones de años. Podemos decir también que uno de los inventos biológicos con más repercusión evolutiva, el acoplamiento de los fotosistemas I y II, se logró a través de la combinación de cientos de genes de organismos diferentes, lo que hoy en día se llama **ingeniería genética**. La diferencia es que en ese entonces la transferencia genética horizontal de dio por el azar (“la mano de Dios”), mientras que hoy en día la transferencia de genes en los laboratorios está supervisada por científicos mas o menos

creativos, que por lo regular prefieren tardarse menos de cien años, no millones como la evolución natural.

El éxito de las cianobacterias pudo deberse no sólo a la ventaja que supuso el aprovechamiento de la energía solar (la fotosíntesis), sino también a la abundancia del substrato H_2O en comparación de H_2S (que solo se presenta en zonas volcánicas o dentro de fumarolas marinas con emisión de azufre). Otra ventaja competitiva pudo ser la inhibición de otros organismos competidores por el oxígeno liberado (como agente tóxico generador de radicales oxidantes). La emergencia de los primeros eucariontes data de 1.8 millones de años, y se debió a la incorporación de otra bacteria heterotrófica a una célula eucariótica con un sistema de membranas internas (núcleo y retículo endoplasmático). Así fue como las células eucarióticas adquirieron sus primeros organelos por endosimbiosis. Las células eucarióticas con mitocondrias pudieron multiplicarse mucho más rápido, ya que podían aprovechar el oxígeno para oxidar alimentos (Ciclo de Krebs y fosforilación oxidativa), y de esta forma obtenían mucho más ATP que las otras células fermentadoras. Muchos millones de años después, algunas de estas células eucarióticas heterotróficas, incorporaron por endosimbiosis a las cianobacterias fotosintéticas formándose así células eucarióticas con cloroplastos, dando lugar a las algas verdes. A partir de entonces, la fotosíntesis de las algas verdes disminuyó los niveles de dióxido de carbono e incrementó el nivel de oxígeno hasta llegar al actual. Por último, los descendientes de las algas verdes, las plantas terrestres, aparecieron hace 500 millones de años creando así la tierra que hoy habitamos, con sus diversos paisajes y ecosistemas. Si antes la vida dependía más de la fotosíntesis marina, hoy en día la mayor parte de la biomasa y de nuestros alimentos la obtenemos de la fotosíntesis de plantas terrestres.

10.2.1.2. El esquema global

La vida sobre la Tierra depende de la luz solar. La energía lumínica es capturada por los organismos fotosintéticos quienes la usan para formar ATP, NADPH, carbohidratos y oxígeno libre a partir del dióxido de carbono y del agua, en una serie compleja de reacciones. En la fotosíntesis, la energía lumínica se convierte en energía química y el carbono se fija en compuestos orgánicos. La ecuación generalizada para este proceso es:



en la cual H_2A significa agua o alguna otra sustancia como H_2S cuyos electrones puedan ser desprendidos fácilmente. Esta fórmula deja claro que el oxígeno (A_2) que se libera a partir de la fotosíntesis viene de las dos moléculas de agua (H_2A) y no del dióxido de carbono. El oxígeno de la

molécula de dióxido de carbono (CO_2) se queda una parte en los carbohidratos (CH_2O) y otra parte se libera en forma de agua (H_2O).

La energía lumínica es capturada por el mundo vivo por medio de pigmentos especiales unidos a complejos de proteínas llamados fotosistemas. La fotosíntesis en las plantas ocurre dentro de organelos celulares, cloroplastos, que están rodeados por dos membranas. Dentro de las membranas del cloroplasto está contenida una solución de compuestos orgánicos e iones, conocida como estroma, y un sistema complejo de membranas internas fusionadas que forman sacos llamados tilacoides. Los tilacoides se apilan formando la grana. Se llaman así porque se ven como pequeños manchones verdes (gránulos) cuando uno mira los cloroplastos con un microscopio. Los pigmentos y los complejos fotosintéticos responsables de la captura de la luz están situados en las membranas tilacoides.

Hace unos 200 años se demostró que se requiere luz para la fotosíntesis de las plantas con generación de oxígeno. La evidencia de que este proceso puede ser influenciado por distintos factores llevó a distinguir una etapa 100% dependiente de la luz -las reacciones lumínicas- y una etapa más bien dependiente de la temperatura. Es decir, se llegó a reconocer que existía una etapa fotoquímica dependiente de luz y una etapa enzimática no dependiente de la luz. A esta segunda etapa se le llama reacciones "oscuras". Ahora se sabe que la etapa que requiere luz es la que se lleva a cabo en los tilacoides (liberación de oxígeno), mientras que la fase oscura se lleva a cabo en el estroma (fijación de dióxido de carbono).

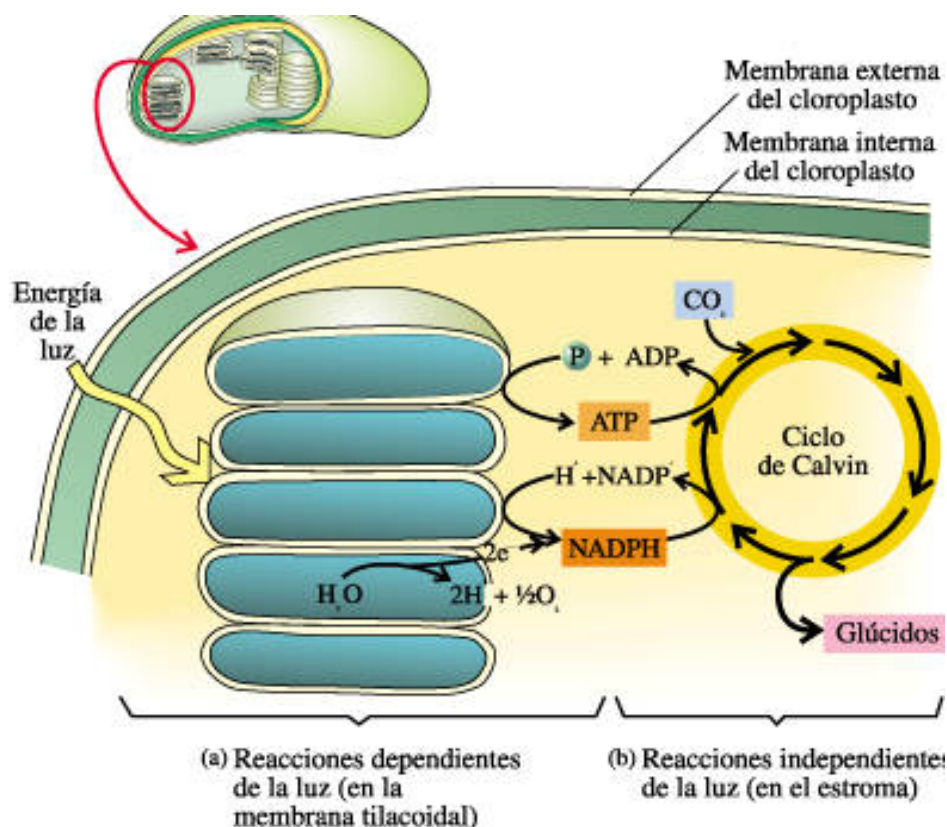


Figura 10-5. Esquema global de la fotosíntesis

10.2.1.3. La clorofila y otros pigmentos

Para que la energía lumínica pueda ser usada por los sistemas vivos, primero debe ser absorbida. Aquí entran en juego los pigmentos. Un pigmento es cualquier sustancia que absorbe luz en el rango visible. Algunos pigmentos absorben luz de todas las longitudes de onda y, por lo tanto, parecen negros. Otros solamente absorben ciertas longitudes de onda, transmitiendo o reflejando las longitudes de onda que no absorben.

La naturaleza de la luz

La luz blanca se separa en sus colores componentes cuando pasa a través de un prisma. A este fenómeno lo llamo Newton, el "célebre fenómeno de los colores". La luz visible es sólo una pequeña porción del vasto espectro electromagnético. De acuerdo con el llamado modelo corpuscular de la luz, un haz de luz está compuesto por pequeños paquetes de energía, denominados actualmente cuantos de luz o fotones.

La energía de un fotón no es la misma para todos los tipos de luz, sino que, en realidad, es inversamente proporcional a la longitud de onda: cuanto mayor sea la longitud de onda, menor será la energía. Los fotones de luz violeta (420nm), por ejemplo, tienen casi el doble de energía que los fotones de luz roja (700nm). Para el ojo humano, el espectro visible va desde la luz violeta -cuyos

rayos de longitudes de onda más cortos son de 380 nanómetros- a la luz roja, cuyos rayos visibles de mayor longitud son de 730 nanómetros (nm).

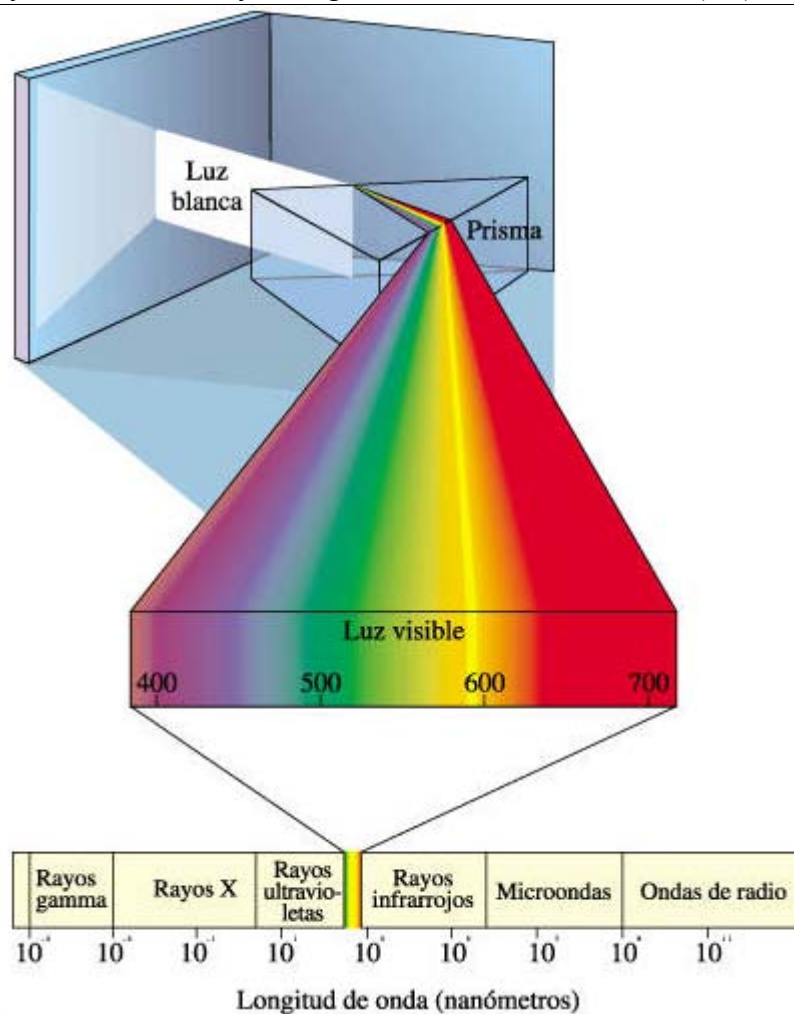


Figura 10-6. La luz blanca es una mezcla de colores diferentes, que van desde el violeta, en un extremo del espectro, hasta el rojo, en el otro.

Los pigmentos que intervienen en la fotosíntesis de los eucariotes incluyen las clorofilas y los carotenoides. Diferentes grupos de plantas y algas usan distintos pigmentos en la fotosíntesis. Hay varios tipos diferentes de clorofila que varían ligeramente en su estructura molecular.

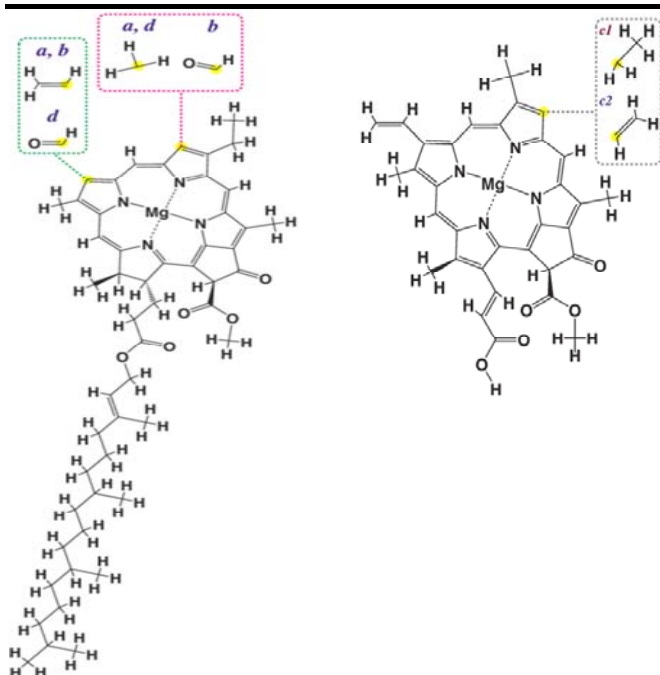


Figura 10-7. Estructura de la clorofila a, b, d, c1 y c2. De wikipedia

La **clorofila a** es el pigmento involucrado directamente en la transformación de la energía lumínica en energía química. La mayoría de las células fotosintéticas también contienen un segundo tipo de clorofila (**clorofila b y c**). Con respecto a los carotenoides, uno de los que se encuentran en las hojas es el beta-caroteno y otro es la luteína. Los carotenoides son pigmentos rojos, anaranjados o amarillos. En las hojas verdes su color está enmascarado por las clorofilas, que son más abundantes. En algunos tejidos, sin embargo, predominan los colores reflejados por los carotenoides, como el licopeno en los tomates maduros. Lo mismo ocurre en las células foliares cuando degradan su clorofila en el otoño.

La luz absorbida por los pigmentos lanza los electrones a niveles energéticos más altos. Dada la forma en que los pigmentos están compactados en las membranas, son capaces de transferir su energía a moléculas reactivas de clorofila a, empaquetadas en una forma particular.

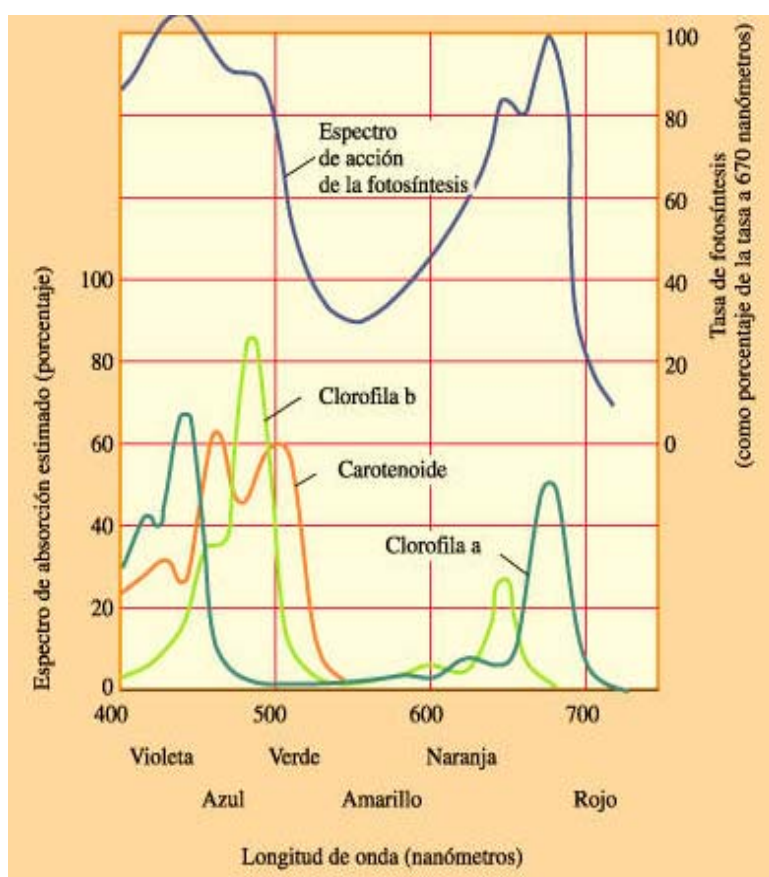


Figura 10-8. La curva superior muestra el espectro de acción de la fotosíntesis, y las curvas inferiores, los espectros de absorción para distintos pigmentos: la clorofila a, la clorofila b y los carotenoides que se encuentran dentro del cloroplasto.

10.2.1.4. Los órganos fotosintéticos

Los tejidos internos de las hojas están completamente encerrados por células epidérmicas transparentes, cubiertas con una capa cerosa, la cutícula. Esta capa la hace impermeable a los gases y los líquidos. El oxígeno, el dióxido de carbono y el vapor de agua salen o entran a la hoja principalmente a través de aberturas especiales: los estomas. Los gases y el vapor de agua llenan los espacios existentes entre las células de la capa esponjosa, entrando y saliendo de las células por difusión. El agua, absorbida por las raíces, entra en la hoja por medio de los vasos del xilema del haz conductor, en tanto que los azúcares, producto de la fotosíntesis, dejan la hoja a través de un tejido conductor conocido como floema, viajando a otras partes de la planta, entre ellas, los órganos no fotosintetizantes. La mayor parte de la fotosíntesis se realiza en las células del parénquima en empalizada, células alargadas que se encuentran directamente por debajo de la epidermis superior y que constituyen el mesófilo. Tienen una vacuola central grande y numerosos cloroplastos que se mueven dentro de la célula, orientándose con respecto a la luz. La luz es capturada en las membranas de los tilacoides, dentro de los cloroplastos.

10.2.1.5. Las membranas fotosintéticas

La unidad estructural de la fotosíntesis en las células vegetales es el cloroplasto. Dentro del cloroplasto se encuentran las **membranas tilacoides**, una serie de membranas internas que contienen los pigmentos fotosintéticos. Cada tilacoide tiene habitualmente la forma de un saco aplanado o vesícula.

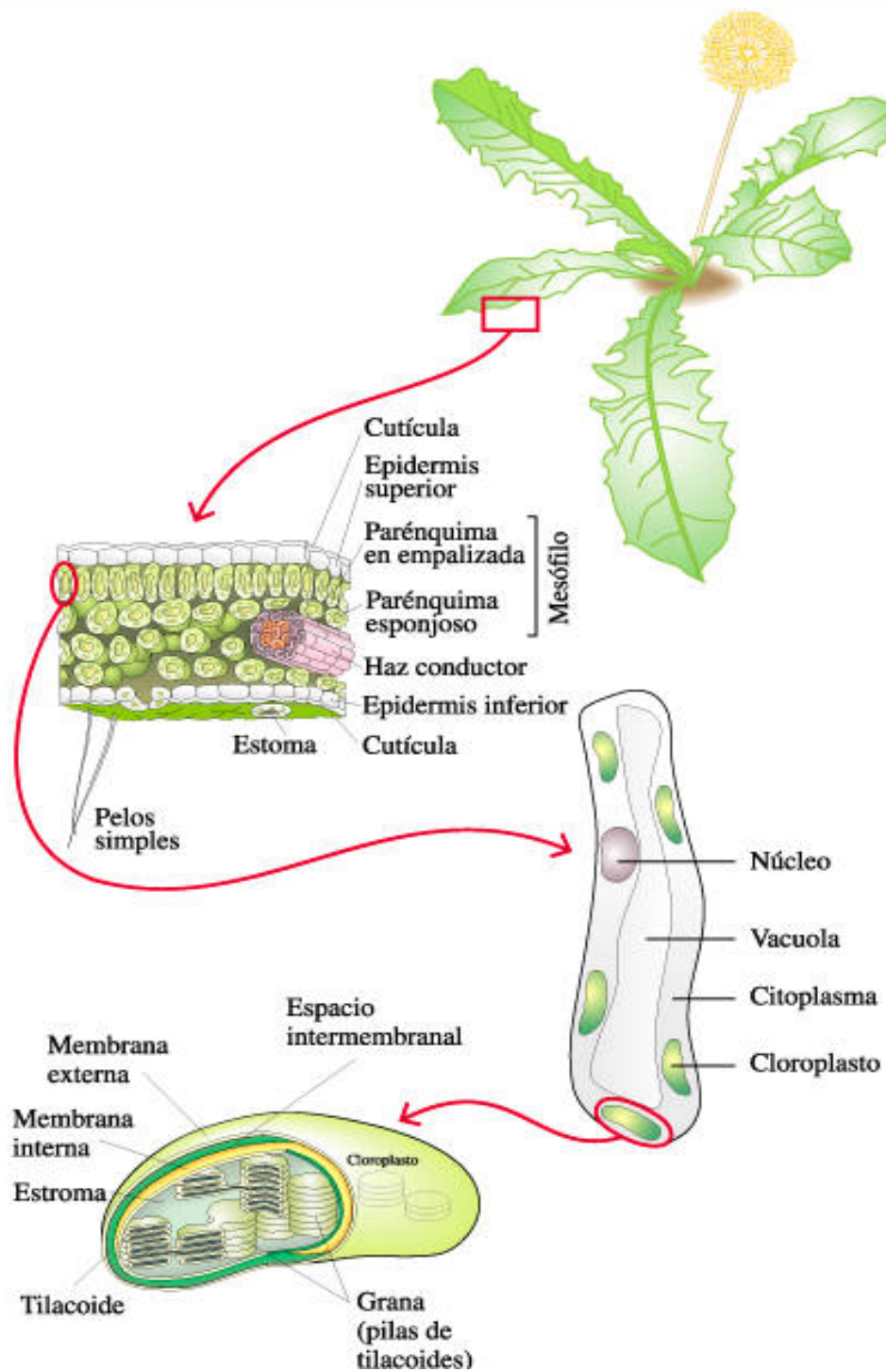


Figura 10-9. Ampliaciones de una hoja verde, y estructura membranal de un cloroplasto.

10.2.1.6. Las etapas de la fotosíntesis

La evidencia de que la fotosíntesis puede ser influenciada por distintos factores llevó a distinguir una etapa dependiente de la luz, la etapa llamada de reacciones "lumínicas", y una etapa enzimática, independiente de la luz, las reacciones "oscuras". Los términos reacciones "lumínicas" y "oscuras" han creado mucha confusión pues, aunque las reacciones "oscuras" no requieren de la luz como tal, sino solamente de los productos químicos de las reacciones

"lumínicas", pueden ocurrir tanto en la luz como en la oscuridad. Mas aún, trabajos recientes han mostrado que varias enzimas que controlan reacciones "oscuras" claves son reguladas indirectamente por la luz. Como resultado, estos términos han caído en desuso y están siendo reemplazados por vocablos que describen más precisamente los procesos que ocurren durante cada etapa de la fotosíntesis: las reacciones que capturan energía y las reacciones de fijación del carbono.

En la primera etapa de la fotosíntesis, la luz es absorbida por las moléculas de clorofila a, que están compactadas de un modo especial en las membranas tilacoides. Los electrones de las moléculas de clorofila a son lanzados a niveles energéticos superiores, y, en una serie de reacciones, su energía adicional es usada para formar ATP a partir de ADP y para reducir una molécula transportadora de electrones conocida como NADP^+ (Nicotinamide Adenine Dinucleotide Phosphate).

El NADP^+ es muy semejante al NAD^+ y también se reduce por la adición de dos electrones y de un protón, formando NADPH. Tanto el ADP, ATP como NAD^+ y NADP^+ son moléculas de azúcares fosforilados unidos a una base nitrogenada (nucleótidos). Sin embargo, los papeles biológicos de estas moléculas son notablemente distintos. El NADH generalmente transfiere sus electrones a otros transportadores de electrones, que continúan transfiriéndolos en pasos discretos a niveles de energía sucesivamente más bajos. El NADPH principalmente se usa para procesos de biosíntesis, mientras que el NADH se utiliza en la glicólisis y la respiración. En el curso de esta transferencia de electrones se forman moléculas de ATP. En contraste, el ATP y ADP proporcionan energía directamente a los procesos biosintéticos de la célula que requieren grandes ingresos de energía.

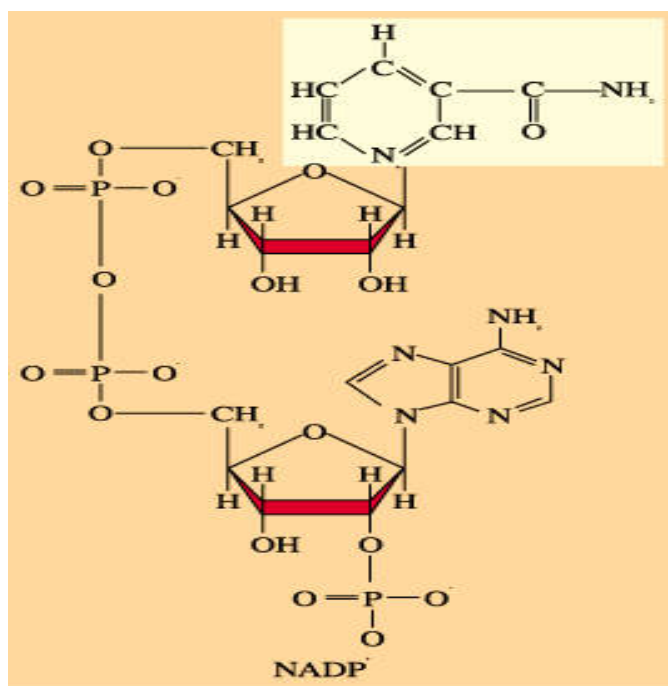


Figura 10-10. Esquema de la molécula de NADP^+ .

En esta primera etapa de la fotosíntesis, también se escinden moléculas de agua, suministrando electrones que reemplazan a los que han sido lanzados desde las moléculas de clorofila a. La escisión de las moléculas de agua es la causa de que se forme oxígeno libre, que difunde hacia el exterior. En la segunda etapa de la fotosíntesis, el ATP y el NADPH formados en la primera etapa se utilizan para reducir el carbono del dióxido de carbono a un azúcar simple. Así, la energía química almacenada temporalmente en las moléculas de ATP y de NADPH se transfiere a moléculas adecuadas para el transporte y el almacenamiento de energía en las células de las algas o en el cuerpo de las plantas. La resultante de este proceso es pues la formación de un esqueleto de carbono, a partir del cual pueden construirse luego otras moléculas orgánicas. La incorporación inicial de CO_2 en compuestos orgánicos se conoce como fijación del carbono. Los pasos por los cuales se lleva a cabo, llamados las reacciones de fijación del carbono, ocurren en el estroma del cloroplasto.

Resumen de las etapas de la fotosíntesis

	Condiciones	¿En dónde?	¿Qué ocurre aparentemente?	Resultados
Reacciones que Capturan energía	Luz	Tilacoides	La luz que incide sobre el Fotosistema II lanza electrones cuesta arriba. Estos electrones son reemplazados por electrones de moléculas de agua que, al escindirse, liberan O_2 . Los electrones luego pasan cuesta abajo, a lo largo de una cadena de transporte de electrones, al Fotosistema I y de éste -nuevamente cuesta abajo- al NADP, que se reduce formando NADPH. Como resultado de este proceso, se forma un gradiente de potencial electroquímico merced al cual se produce ATP a través de un mecanismo quimiosmótico	La energía de la luz se convierte en energía química que se almacena en enlaces de ATP y NADPH
Reacciones de fijación de carbono	No requieren luz, aunque algunas enzimas son reguladas por ella	Estroma	Ciclo de Calvin. El NADP y el ATP formados en las reacciones que capturan energía luminica se utilizan para reducir el dióxido de carbono. El ciclo produce gliceraldehído fosfato, a partir del cual puede formarse glucosa y otros compuestos orgánicos	La energía química del ATP y del NADPH se usa par incorporar carbono a moléculas orgánicas

10.2.1.7. Reacciones que capturan energía

En el siglo pasado se postuló un modelo para explicar cómo ocurren las reacciones que capturan energía. Según este modelo, la energía lumínica incide sobre pigmentos antena del Fotosistema II; luego, los electrones pasan cuesta abajo al Fotosistema I, a lo largo de una cadena de transportadores de electrones. Este pasaje genera un gradiente de protones que impulsa la síntesis de ATP a partir de ADP, proceso llamado fotofosforilación. Más adelante se explican estos procesos en detalle.

Al igual que la fosforilación oxidativa en las mitocondrias, la fotofosforilación en los cloroplastos es un proceso quimiosmótico. En las reacciones de fijación del carbono los productos de la primera etapa de la fotosíntesis se usan en la síntesis de moléculas orgánicas. Estas reacciones, que ocurren en el estroma, forman parte del ciclo reductivo de las pentosas fosfato (Ciclo de Calvin). Las moléculas orgánicas obtenidas en el Ciclo de Calvin -azúcares de tres, cuatro, cinco y seis carbonos- son usadas por las células vegetales para elaborar sacarosa, almidón, aminoácidos y ácidos grasos. Estas moléculas son utilizadas *in situ* para los propios fines de la planta o para exportar a otros tejidos a través del floema.

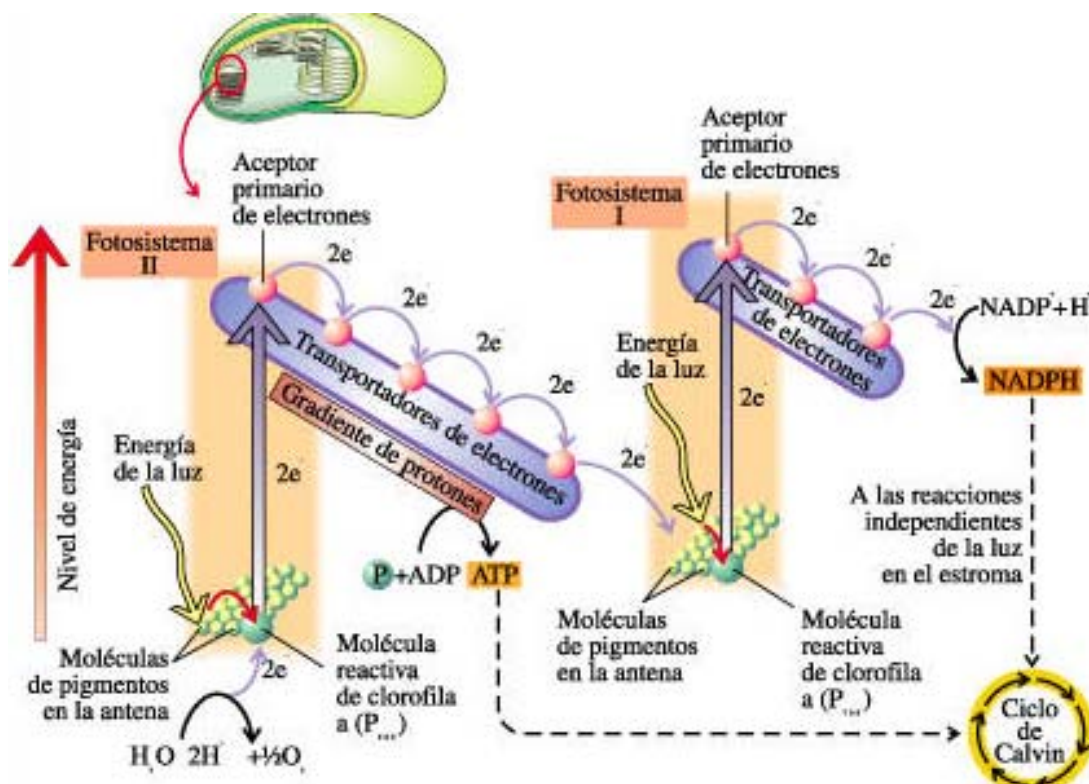


Figura 10-11. Esquema resumen del acoplamiento de los dos fotosistemas para romper molécula de agua, liberar oxígeno, generar energía química (ATP) y poder reductor (NADPH).

La energía lumínica incide sobre pigmentos antena del Fotosistema II, que contiene algunos cientos de moléculas de clorofila, a y b. Los electrones

son lanzados cuesta arriba desde la molécula reactiva P680 de la clorofila a un aceptor de electrones primario. Cuando se eliminan los electrones, ellos son reemplazados por electrones de las moléculas de agua, con la producción simultánea de O_2 libre y protones (iones H^+). Luego, los electrones pasan cuesta abajo al Fotosistema I a lo largo de una cadena de transportadores de electrones; este pasaje genera un gradiente de protones que impulsa la síntesis de ATP a partir de ADP, proceso denominado fotofosforilación. La energía lumínica absorbida en los pigmentos antena del Fotosistema I y transferida a la clorofila P700 da como resultado que se lancen electrones hacia otro aceptor primario de electrones. Los electrones eliminados del P700 son reemplazados por electrones del Fotosistema II y son finalmente aceptados por el transportador de electrones $NADP^+$. La energía proveniente de esta secuencia de reacciones está contenida en las moléculas de NADPH y en el ATP formado por fotofosforilación. Entre éstas se distinguen los pigmentos, los transportadores de electrones, los Fotosistemas I y II y enzimas necesarias, incluyendo las ATP sintetasas. La disposición de estas moléculas en la membrana tilacoidal hace posible la síntesis quimiosmótica del ATP durante la fotofosforilación.

Los electrones pasan desde el aceptor de electrones primario, a lo largo de una cadena de transporte de electrones, a un nivel de energía inferior, el centro de reacción del Fotosistema I. A medida que pasan a lo largo de esta cadena de transporte de electrones, parte de su energía se empaqueta en forma de ATP. La energía lumínica absorbida por el Fotosistema I lanza los electrones a otro aceptor primario de electrones. Desde este aceptor son transferidos mediante otros transportadores de electrones al $NADP^+$ para formar NADPH. Los electrones eliminados del Fotosistema I son reemplazados por los del Fotosistema II. El ATP y el NADPH representan la ganancia neta de las reacciones que capturan energía. Para generar una molécula de NADPH, deben ser lanzados dos electrones desde el Fotosistema II y dos del Fotosistema I. Se escinden dos moléculas de agua para formar protones y gas oxígeno, poniendo en disponibilidad los dos electrones de reemplazo necesarios para el Fotosistema II. Se regenera una molécula de agua en la formación de ATP.

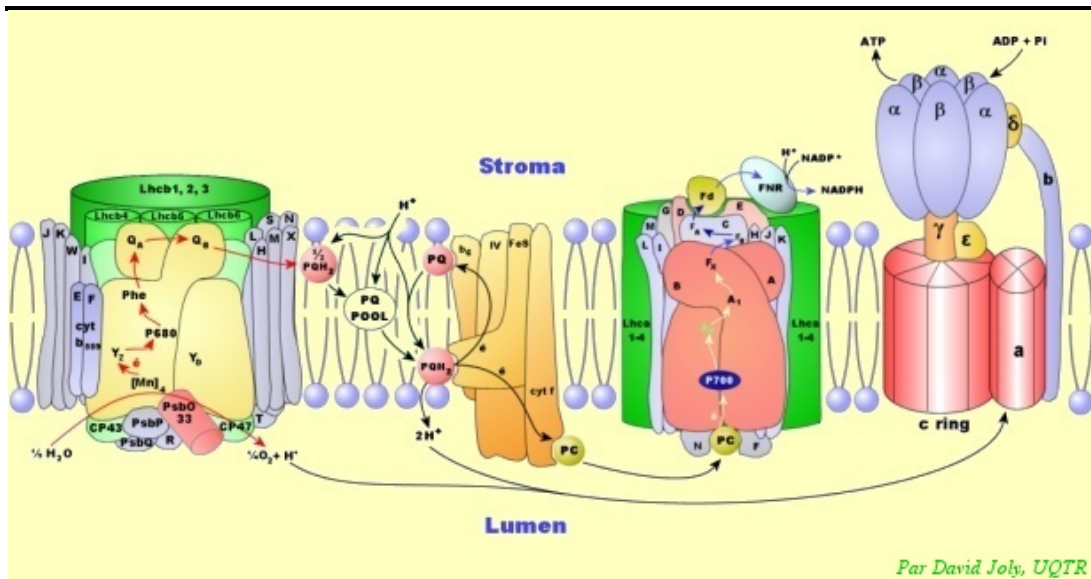


Figura 10-12. Complejos proteicos presentes en la membrana tilacoidal.

La fotofosforilación también ocurre como resultado del flujo cíclico de electrones, proceso en el que no participa el Fotosistema II. En el flujo cíclico de electrones, los electrones lanzados desde el P700 en el Fotosistema I no pasan al NADP⁺, sino que son desviados a la cadena de transporte de electrones que une al Fotosistema II con el Fotosistema I. A medida que fluyen a lo largo de esta cadena, nuevamente al P700, se transportan cargas a través de la membrana. Esto genera un potencial electroquímico en los tilacoides, que finalmente genera ATP. En un proceso quimiosmótico, como la fotofosforilación que ocurre en los cloroplastos, a medida que los electrones fluyen en la cadena de transporte de electrones desde el Fotosistema II al Fotosistema I, los protones son bombeados desde el estroma al espacio tilacoide, creando un gradiente electroquímico. A medida que los protones fluyen a favor de este gradiente desde el espacio tilacoide nuevamente al estroma, pasando a través de los complejos de ATP sintetasa, se forma ATP. Al igual que la fosforilación oxidativa en las mitocondrias, la fotofosforilación en los cloroplastos es un proceso de acoplamiento quimiosmótico.

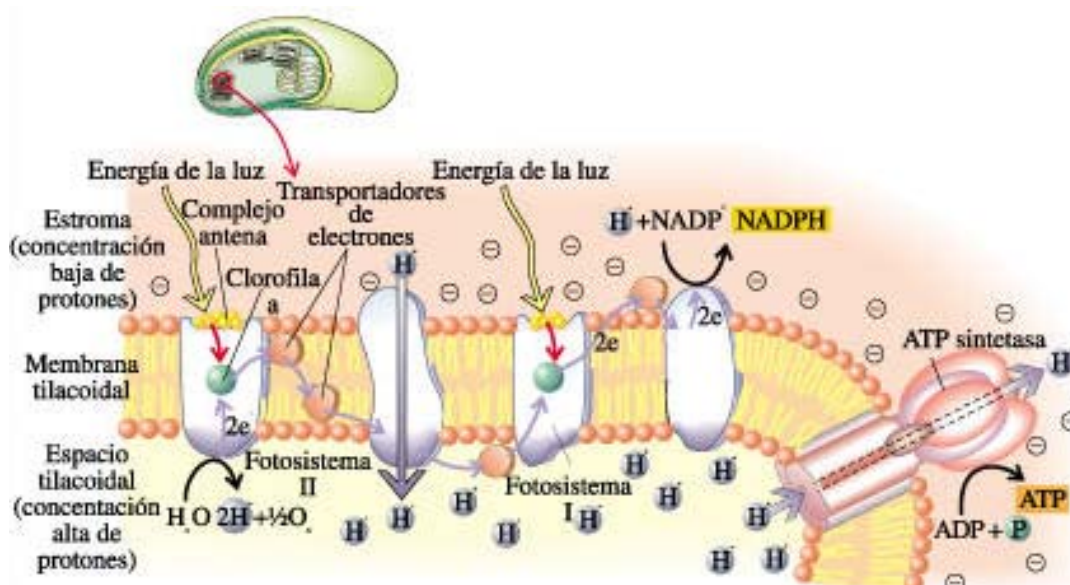


Figura 10-13. Los dos fotosistemas generan un gradiente de protones para la síntesis de ATP. En el interior del tilacoide hay un pH bajo, mientras que en el exterior predomina un pH 8.

En este proceso, los electrones de la molécula reactiva de clorofila a del Fotosistema II son impulsados a niveles energéticos superiores por la luz solar. A medida que descienden por una cadena de transportadores de electrones hacia la molécula reactiva de clorofila a del Fotosistema I, la energía que liberan es empleada para bombear protones (H^+). Los protones se bombean desde el estroma al espacio tilacoidal. Esto crea un gradiente electroquímico. Cuando los protones se mueven a favor del gradiente a través del complejo de la ATP sintetasa, desde el espacio tilacoidal al estroma del cloroplasto, el ADP se fosforila a ATP.

10.2.1.8. Reacciones de fijación de carbono

Las reacciones que fijan carbono son también conocidas como reacciones "oscuras" o reacciones "independientes de la luz". En organismos autótrofos acuáticos el dióxido de carbono (CO_2) penetra en los sin necesidad de estructuras especiales. Sin embargo, las plantas terrestres deben protegerse de la desecación y han desarrollado aberturas especiales denominadas estomas que regulan la entrada y salida del gas por las hojas. La fase oscura de la fotosíntesis también se denomina ruta reductiva de las pentosas fosfato (RPP) o mejor conocido como ciclo de Calvin (en honor a su descubridor Melvin Calvin). Las reacciones tienen lugar en el estroma dentro de los cloroplastos.

A fines de la segunda guerra mundial, en los laboratorios de Berkeley (California), Calvin y sus colaboradores, usando Carbono-14 y las entonces nuevas técnicas de intercambio iónico, cromatografía en papel y radiografía "mapearon" completamente el ciclo del carbono en la fotosíntesis. Por estos

trabajos resultó laureado con el premio Nobel en 1961, y el ciclo del carbono se conoce comúnmente como ciclo de Calvin, o de Calvin-Benson.

En las reacciones de fijación del carbono que ocurren en el estroma, el NADPH y el ATP, producidos en las reacciones de captura de energía, se usan para reducir un compuesto de tres carbonos, el gliceraldehído-3-fosfato (GAP). A esta vía en la que el carbono se fija para formar 3-fosfoglicerato (3PGA) mediante ribulosa-1,5- bifosfato y CO_2 , se le denomina vía de tres carbonos o C3. Esta reacción es catalizada por medio de la enzima Ribulosa 1-5-bifosfato carboxilasa oxidasa, RUBISCO.

Rubisco presenta 8 subunidades grandes y 8 subunidades pequeñas ensambladas en 4 dímeros (figura 10-8). Los genes que codifican para las unidades pequeñas son nucleares, mientras que los genes de las subunidades grandes son plastídicos. La subunidad pequeña se ensambla en los ribosomas del citosol y se dirige al cloroplasto por medio de un péptido señal. La enzima es ensamblada por completo en este organelo.

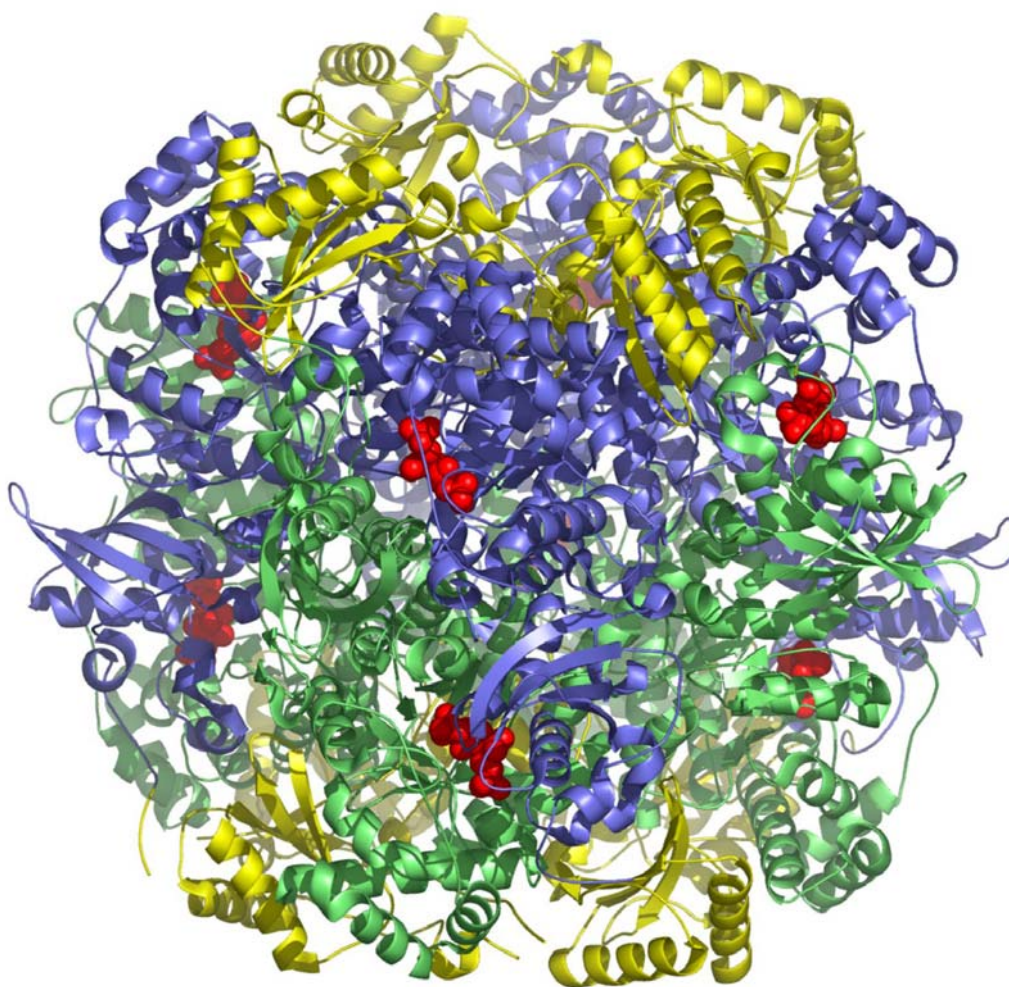


Figura 10-14. Estructura secundaria y terciaria de la holoenzima Rubisco. En rojo se muestran las moléculas de ribulosa-1-5-bisfosfato unidas a los sitios activos de la enzima.

Además de la reacción de carboxilación, rubisco también reacciona con O_2 , dando lugar al proceso de fotorrespiración. Como el CO_2 y el O_2 son sustratos que compiten por la enzima, la carboxilación de la ribulosa -1, 5-bifosfato depende de la relación CO_2/O_2 del medio.

Durante la carboxilación, rubisco ingresa tres moléculas de dióxido de carbono por cada ciclo. (figura 10-9).

Se combinan seis moléculas de ribulosa bifosfato (RuBP), un compuesto de cinco carbonos, con seis moléculas de dióxido de carbono, produciendo seis moléculas de un intermediario inestable que pronto se escinde en doce moléculas de fosfoglicerato, un compuesto de tres carbonos. Estos últimos se reducen a doce moléculas de gliceraldehído fosfato. Diez de estas moléculas de tres carbonos se combinan y se regeneran para formar seis moléculas de cinco carbonos de RuBP. Las dos moléculas "extra" de gliceraldehído fosfato representan la ganancia neta del ciclo de Calvin. Estas moléculas son el punto de partida de numerosas reacciones que pueden implicar, por ejemplo, la síntesis de glúcidos, aminoácidos y ácidos grasos.

La energía que impulsa al ciclo de Calvin son el ATP y el NADPH producidos por las reacciones de captura de energía en la primera etapa de la fotosíntesis.

El gliceraldehído-3-fosfato también puede ser utilizado como material de partida para otros compuestos orgánicos necesarios para la célula. Otras plantas que viven en ambientes secos y cálidos tienen mecanismos que les permiten fijar inicialmente el CO_2 por una de dos vías, y así logran minimizar la pérdida de agua. Estas vías se conocen como la vía de cuatro carbonos, o C4 y el camino de las plantas CAM, y preceden al Ciclo de Calvin.

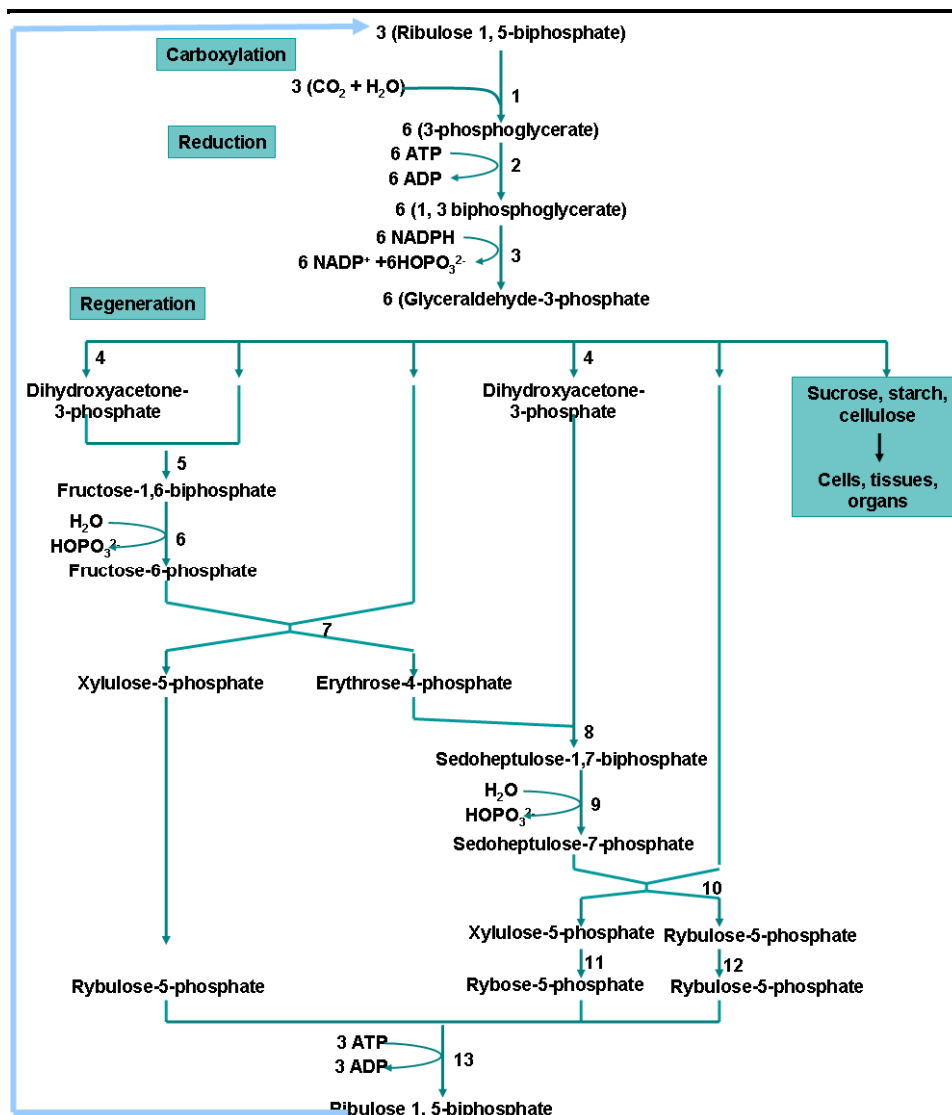


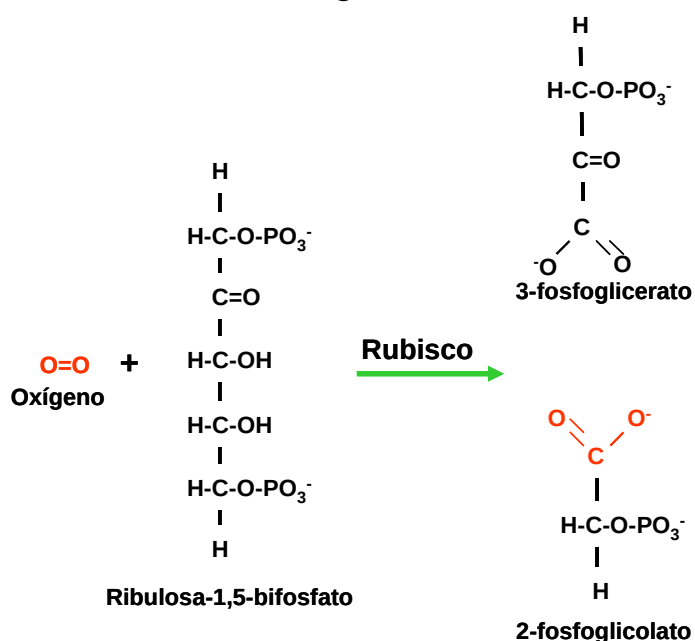
Figura 10-15. Ciclo de Calvin. Las enzimas están representadas por números: 1 Rubisco; 2 3-phosphoglycerate kinase; 3 glyceraldehyde-3-phosphate kinase; 4 triose phosphate isomerase; Aldolase; 6 fructose-1,6-bisphosphate phosphatase; 7 transketolase; 8 Aldolase; 9 Sedoheptulose-1,7-bisphosphate phosphatase; 10 transketolase; 11 Ribulose-5-phosphate epimerase; 12 ribose-5-phosphate isomerase; 13 Ribulose-5-phosphate kinase (modificada de Taiz y Zeiger 1998).

10.2.1.9. Fotorrespiración

Como ya se mencionó, la fotorrespiración es el proceso por el cual Rubisco fija O_2 en lugar de fijar CO_2 . Las tasas relativas de carboxilación/oxigenación están determinadas por la concentración de ambos gases y por la preferencia de Rubisco hacia cada uno de ellos (Foyer *et al.*, 2009). Dado que la reacción de carboxilación de Rubisco evolucionó en condiciones de alta concentración de CO_2 , las consecuencias de la oxigenación eran mínimas (Holland, 2006), pero en las condiciones atmosféricas actuales la reacción de oxigenación se lleva a cabo mucho más rápido en las plantas C_3 , por lo que la fotorrespiración es considerada

uno de los procesos biológicos más ineficientes, por lo que, desde hace décadas se han realizado varios esfuerzos por erradicarlo (Foyer 2009). Aunque, por otro lado, se ha demostrado que este proceso es necesario en algunas plantas C3. Por ejemplo, se ha propuesto que procesos tan importantes como la fijación de nitrógeno pueden depender de la fotorrespiración (Rachmilevitch et al., 2004)

El oxígeno, un inhibidor de la fotosíntesis, puede estar presente en alta concentración en las hojas de las plantas. La elevada concentración de oxígeno en los cloroplastos puede inducir fotorrespiración. En la fotorrespiración, el oxígeno sustituye al CO_2 como sustrato de la RUBISCO, provocando la liberación de CO_2 y la oxidación de la RuBP como se muestra en la siguiente ecuación:



La elevada concentración de oxígeno en el lugar de reacción de la RUBISCO es la causa de la fotorrespiración. Las plantas C4 evitan la fotorrespiración mediante una extensión del ciclo de Calvin-Benson que consigue concentrar el CO_2 , y no el oxígeno, en las células de la vaina del haz, donde tiene lugar la reacción catalizada por la RUBISCO. Las plantas C4 pueden mantener alta la concentración local del CO_2 para la actividad de la RUBISCO sin incremento simultáneo de la concentración de oxígeno.

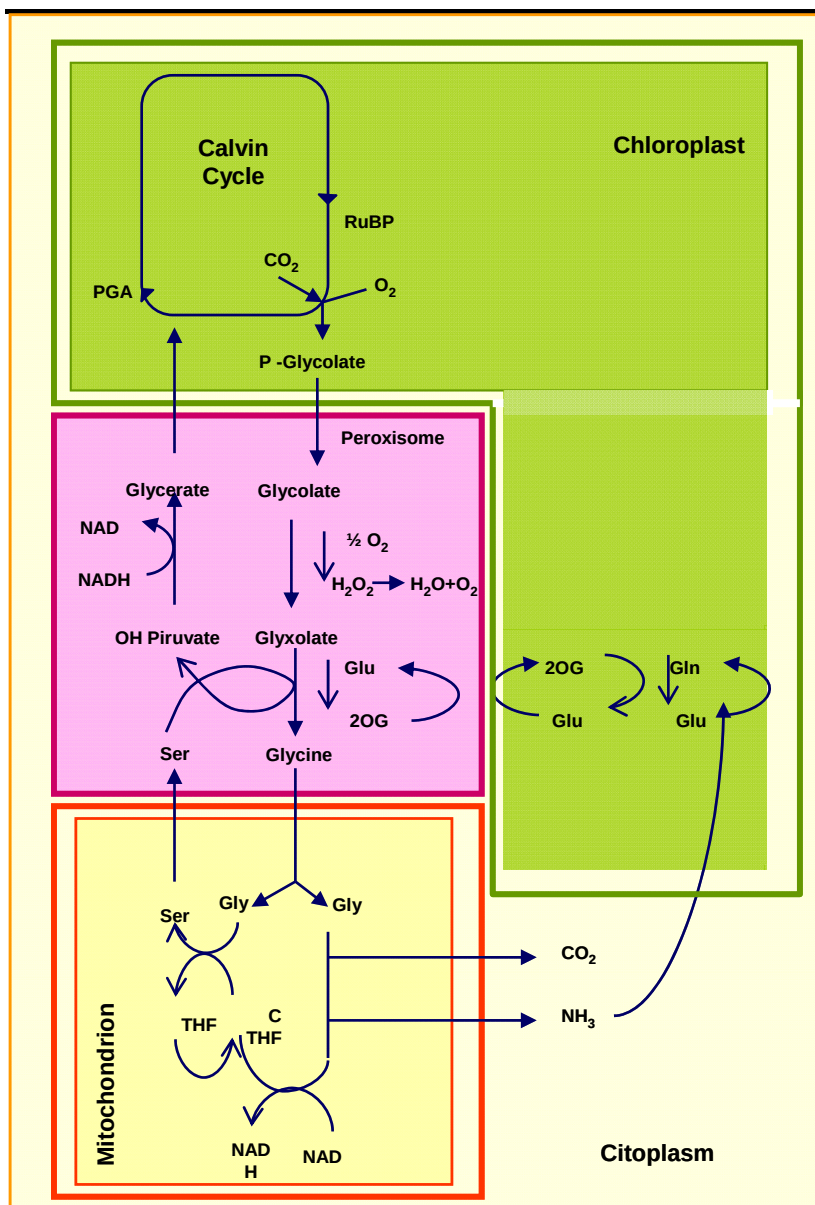


Figura 10-16. Esquema metabólico de la fotorrespiración.

An abbreviated scheme of the photorespiratory pathway. Phosphoglycolate produced by ribulose biphosphate (RuBP) oxygenase activity is converted to glycolate by phosphoglycolate phosphatase in the chloroplast. Glycolate enters peroxisomes and is converted to glyoxylate by glycolate oxidase. Glyoxylate is transaminated to Gly by either Ser:glyoxylate aminotransferase or Glu:glyoxylate aminotransferase. In mitochondria, Gly is converted to CO₂, ammonia and the methylene group of methylene tetrahydrofolate (C1-THF). Gly and C1-THF condense to produce Ser. Peroxisomal Ser is deaminated to hydroxypyruvate, which is reduced to glycerate by hydroxypyruvate reductase. Glycerate enters the chloroplast and is phosphorylated to 3-phosphoglycerate, an intermediate of the Calvin cycle. Ammonia released during Gly decarboxylation is used by Gln synthetase to produce Gln. Glu synthase condenses 2-oxoglutarate (2-OG) and Gln to produce two molecules of Glu. A dicarboxylate transporter in the chloroplast envelope transfers oxoglutarate, Glu, and Gln across the chloroplast envelope. Overall, two molecules of phosphoglycolate are converted to one molecule of phosphoglycerate and one molecule of CO₂ (Modificada de Somerville C R., 2001).

Tarea 10-1

Traduzca el anterior recuadro en inglés, al castellano, y dibuje su propio esquema de fotorrespiración considerando los diferentes compartimentos subcelulares.

10.2.2. Plantas C3, C4 y CAM

10.2.2.1. Plantas C3

Se denominan así a las plantas en las que el primer producto de la fijación de CO_2 es una molécula de 3 carbonos, el 3 fosfoglicerato. La ruta metabólica C3 se encuentra en los organismos fotosintéticos como las cianobacterias, algas verdes y en la mayoría de las plantas vasculares. Estas últimas, son plantas que crecen en zonas templadas ($15-25^\circ\text{C}$) y se saturan de luz con $200-300 \text{ J m}^{-2} \text{ s}^{-1}$. Como la intensidad de la luz solar en el verano es de $800 \text{ a } 1000 \text{ J m}^{-2} \text{ s}^{-1}$, la mayoría de esa luz no la pueden aprovechar esas plantas.

10.2.2.2. Plantas C4

Existe otro grupo denominadas plantas C4. En estas el primer producto de la fijación de CO_2 es un compuesto de 4 carbonos como el malato y el aspartato. En estas plantas la captura del dióxido de carbono se da mediante la enzima fosfoenolpiruvato carboxilasa (PEPC). Este proceso ocurre en el mesófilo, mientras que la asimilación de carbono, ya en azúcares se realiza en el haz de vaina, tras la descarboxilación de la molécula de 4 carbonos, se continúa con el ciclo que presentan las plantas C3. Por lo anterior las plantas C4 presentan otro tipo de anatomía foliar, denominada anatomía tipo Kranz (figura 10-11).

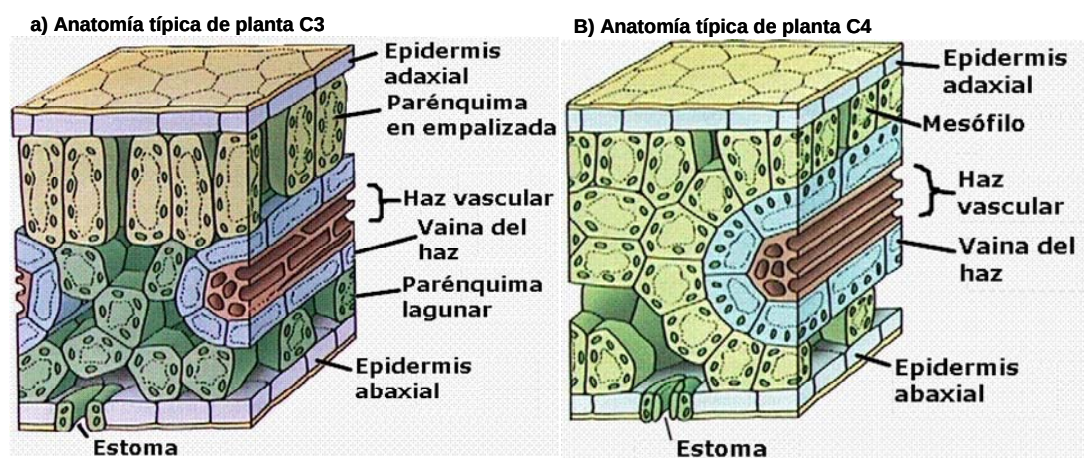


Figura 10-11. a) típica anatomía de una planta C3; b) anatomía tipo Kranz, típica de plantas C4. (http://www.etsmre.upv.es/varios/biologia/Temas/tema_11.htm)

Una característica de la anatomía Kranz es la presencia de cloroplastos de distinto tipo. Los cloroplastos del mesófilo solo tienen fotosistema I, no acumulan almidón ni tampoco tienen proteína Rubisco. Mientras que los cloroplastos del haz vascular sí tienen los dos fotosistemas, sí tienen actividad de Rubisco y sí acumulan almidón. El hecho de que la proteína Rubisco solo se concentre en ciertas células permite a las plantas C₄ tener una mayor eficiencia de uso de agua, y también de nitrógeno. La proteína Rubisco representa casi el 30% del nitrógeno que requiere la planta para crecer, por lo que cualquier ahorro en Rubisco, permite ahorrar nitrógeno como fertilizante.

Las especies C₄, que incluyen, la caña de azúcar, el sorgo, el maíz y otros pastos, crecen en climas tropicales, pues presentan temperaturas óptimas de crecimiento entre 30 y 47°C; no exhiben prácticamente ningún síntoma de saturación de luz. Aún con intensidades de 880 J m⁻² s⁻¹ por lo cual pueden hacer mejor uso de las intensidades de luz altas y crecen bien en condiciones de escasez de agua.

10.2.2.3. Plantas CAM

Otras especies realizan la fotosíntesis denominada CAM (*Crassulacean Acid Metabolism*). En estas plantas los procesos C₃ y C₄ se llevan a cabo en la misma célula, pero están separados temporalmente. Durante la noche se fija el CO₂ atmosférico como malato (C₄) mediante la PEPC en el citosol. El malato se guarda en la vacuola. Durante el día, el malato sale de la vacuola y se descarboxila. El CO₂ liberado se asimila mediante el ciclo C₃ (Dodd et al., 2002). Las plantas que realizan este proceso crecen en climas cálidos, pues su temperatura óptima de crecimiento es 35°C o mayor. Entre estas se encuentra la piña, el agave y el nopal.

Figura 10-17. Planta CAM del género *Opuntia*

10.2.3. Descripción de las distintas rutas fotosintéticas

10.2.3.1. Fotosíntesis C₃

Tarea: Explique la fotosíntesis C₃

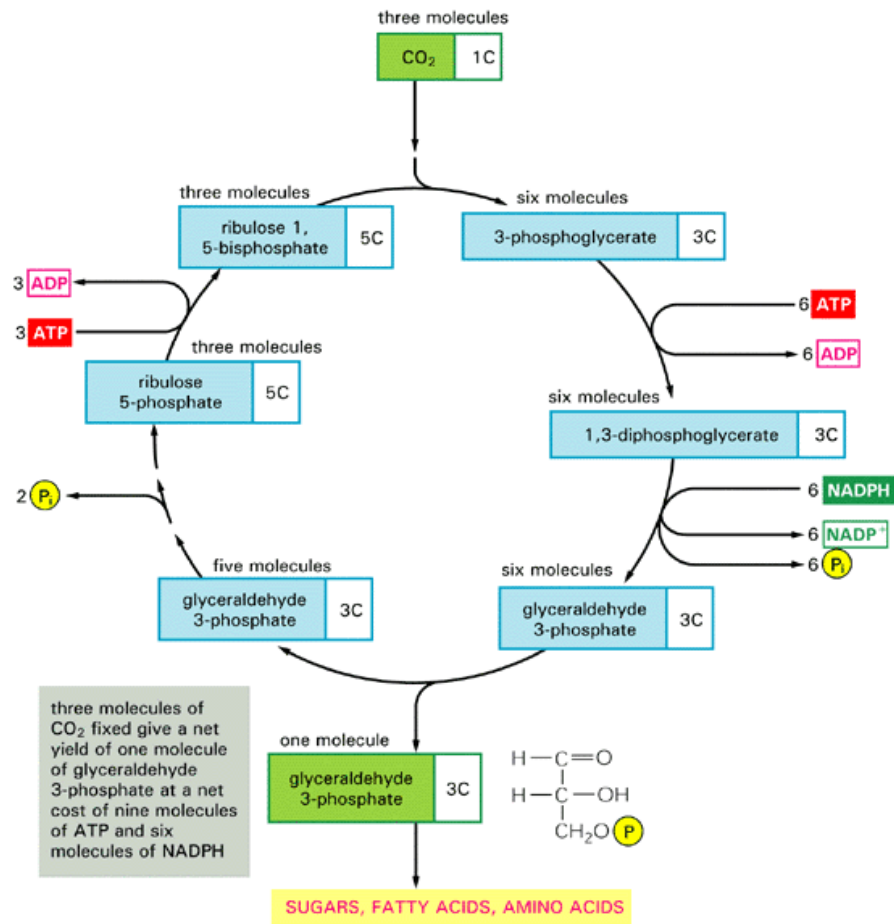


Figura 10-18. Ciclo de Calvin simplificado.

Calvin Cycle (Dark Reactions of Photosynthesis)

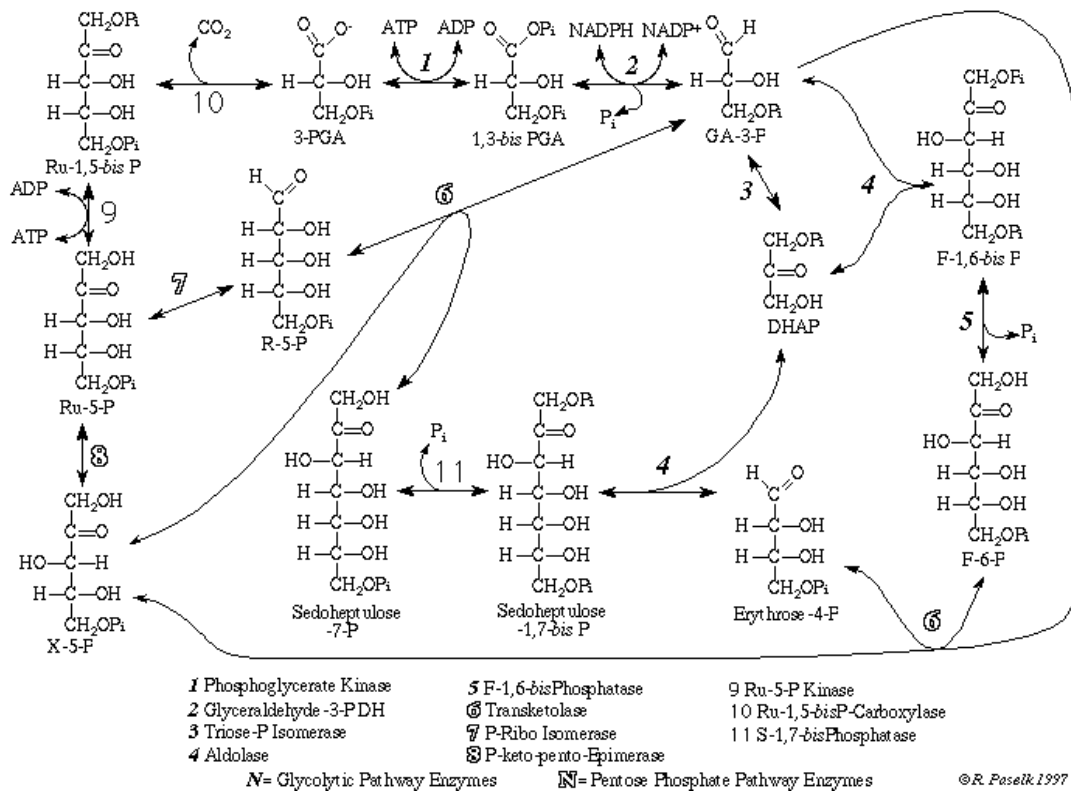


Figura 10-19. Ruta reductiva de las pentosas fosfato.

10.2.3.2. Fotosíntesis C4

Tarea: Explique la fotosíntesis C4



Figura 10-20. La escasez de agua es un problema cada vez mas serio en el mundo.

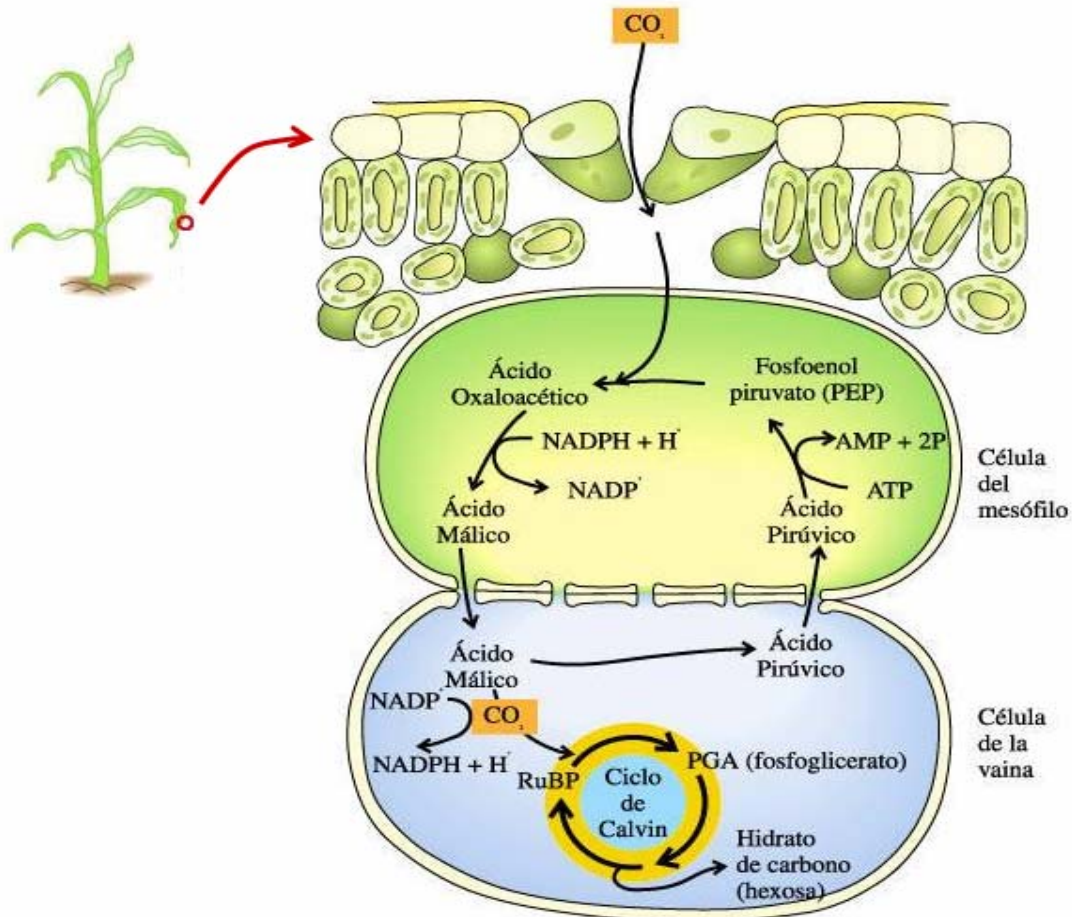


Figura 10-21. Fotosíntesis tipo C4. El CO_2 es fijado en dos compartimientos diferentes: en el mesófilo el CO_2 es fijado como HCO_3^- por la anhidrasa carbónica para ser tomado a continuación por la PEPCase que incorpora el carbono en un ácido C4. Este ácido C4 es transportado hacia el haz de vaina por la acción de acarreadores específicos ATP dependientes, en donde es descarboxilado para liberar CO_2 que es fijado por RUBISCO e incorporado en el ciclo de Calvin-Benson.

Por otro lado, dado que la actividad estomática responde finamente al balance entre CO_2 ganado/ H_2O perdida, aquellas condiciones que lleven a un balance desfavorable como alta temperatura e irradiancia, alto déficit de presión de vapor entre mesófilo y atmósfera, aporte limitado de agua por el suelo o conductividad eléctrica muy alta en la solución de agua del suelo, tenderán al incremento en la restricción difusiva del agua con el cierre estomático parcial o total. Dicho cierre estomático también impacta negativamente en la difusión de CO_2 , lo cual se traduce en aumento en la actividad fotorrespiratoria de la planta, cosa que no ocurre en las plantas C4 o CAM.

En aquellos ambientes con restricciones hídricas constantes, estacionales o diarias como son las zonas áridas, semiáridas y ambientes epifíticos las plantas C4 y CAM funcionan como especialistas de gran éxito con mayor

WUE en comparación con las plantas C3. Las modificaciones bioquímicas con lo cual se consigue esto se relacionan con el aumento en la cantidad y eficiencia de acción de la anhidrasa carbónica (AC), enzima que tiene importancia marginal en las plantas C3, y con la acción de un sistema de bombeo del CO_2 conseguido a través de la acción de la fosfoenolpiruvatocarboxilasa (PEPCasa) y ATPasas de membrana.

Para las plantas C4 el resultado de las modificaciones evolutivas es que el CO_2 es fijado en dos compartimientos diferentes: en el mesófilo el CO_2 es fijado como HCO_3^- por la AC para ser tomado a continuación por la PEPCasa que incorpora el carbono en un ácido C4. Este ácido C4 es transportado hacia la vaina del haz vascular por la acción de acarreadores específicos ATP dependientes en donde es descarboxilado para liberar CO_2 que es fijado por RUBISCO e incorporado en el ciclo de Calvin-Benson. Con la acción de este mecanismo de concentración y bombeo de CO_2 hacia los sitios de fijación por RUBISCO la planta es capaz mantener tasas altas de asimilación de CO_2 en presencia de baja concentración intercelular de dicho gas. A pesar de estas adaptaciones las plantas C4 no son más tolerantes al estrés hídrico severo que las C3; esto significa que el mecanismo C4 es una adaptación encaminada al uso eficiente del agua, no a la tolerancia al estrés hídrico.

Algunas plantas que viven en climas secos y calurosos mantienen una baja concentración de oxígeno en sus hojas pues mantienen los estomas cerrados para evitar la pérdida de agua. Para alcanzar concentraciones de CO_2 adecuadas para la fotosíntesis, las plantas C4 han adaptado la fotorrespiración con una modificación del ciclo de Calvin-Benson. Las plantas C4 tienen una anatomía especial, con prominentes células de vaina de haz alrededor de los vasos de las hojas. La fotorrespiración es mínima en plantas C4 comparada con plantas C3, y el CO_2 es concentrado activamente en las células de la vaina del haz.

10.2.3.3. Fotosíntesis CAM

En las plantas CAM, sin embargo, el resultado de las modificaciones evolutivas es que el CO_2 es fijado en dos etapas separadas temporalmente, y no físicamente como ocurre en las C4. Durante la noche la apertura de los estomas permite la difusión de CO_2 que es fijado como HCO_3^- por la AC y es tomado por la PEPC que lo incorpora en ácidos C4 que se acumulan en las vacuolas vía una bomba de membrana ATP dependiente. Durante el día, los estomas cierran y los ácidos C4 son llevados al citoplasma, a través de un mecanismo aparentemente pasivo, en donde son descarboxilados. El CO_2 liberado, que alcanza concentraciones internas muy altas, es fijado en los cloroplastos por RUBISCO para incorporarlo al ciclo de Calvin-Benson. La respuesta se relaciona con la presión selectiva que ejercen

ciertos ambientes en cuanto a la relación CO_2 fijado vs. H_2O transpirada o Eficiencia en el Uso del Agua (WUE). Puede demostrarse que incluso bajo condiciones ambientales favorables, una planta C_3 pierde aproximadamente 100 moléculas de H_2O por molécula de CO_2 que entra por los estomas. En zonas con aporte constante de agua este hecho no representa un problema pero en regiones áridas y semiáridas sí llega a serlo.

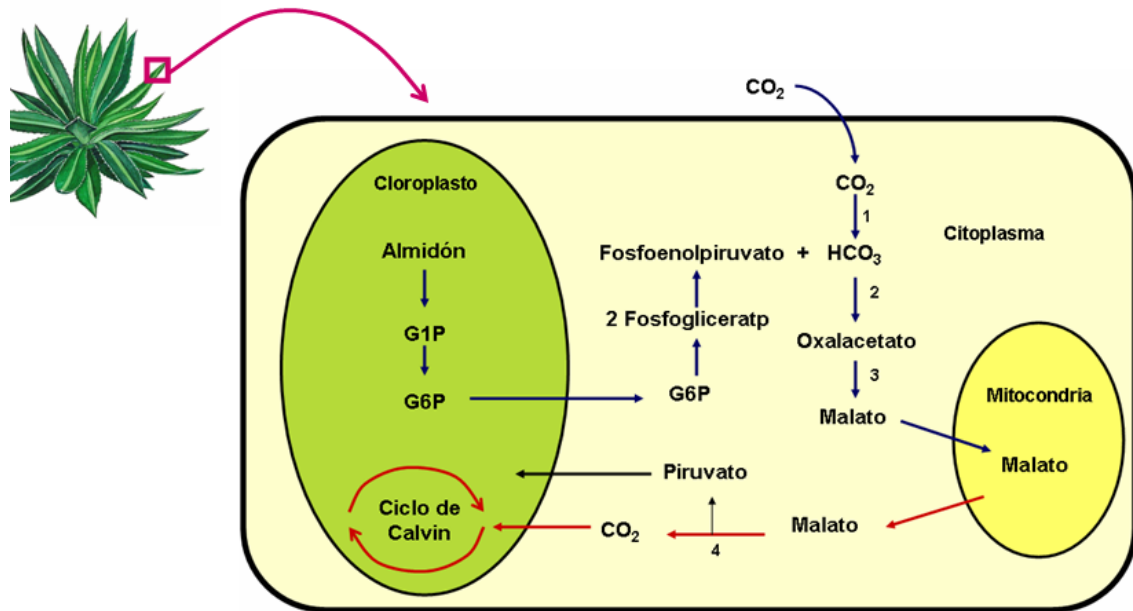


Figura 10-22. Metabolismo tipo CAM. Con las flechas azules se indican las reacciones nocturnas del ciclo C_4 y con rojas, las reacciones que ocurren durante el día. Algunas enzimas están indicadas con números: 1 anhidrasa carbónica; 2 Fosfoenolpiruvato carboxilasa; 3 malato deshidrogenasa; 4 enzima málica.

Tarea 10-2

¿Cuál de los siguientes enunciados sobre la fotosíntesis en plantas C_4 NO es cierto?

- A. El primer producto de la fijación del dióxido de carbono es un compuesto con 4 átomos de carbono.
- B. La fotosíntesis C_4 es una adaptación para las plantas que viven en climas cálidos y áridos.
- C. El dióxido de carbono es fijado inicialmente en células mesófilas, pero el ciclo de Calvin es activo en las células de la vaina del haz en las hojas de las plantas C_4 .
- D. El ATP total que usan las plantas C_4 para la biosíntesis de azúcares es mayor que el usado en las plantas C_3 .

E. La fotorrespiración es mínima en plantas C4 comparada con plantas C3.

10.2.3.4. Consideraciones ecofisiológicas

Mientras que las plantas C3 transpiran 500-700 g de agua por cada gramo de materia seca, las plantas C4 pierden solamente 250-400 g de agua. Entre las plantas que alcanzan altas eficiencias fotosintéticas están el maíz, sorgo y caña de azúcar.

Las diferencias metabólicas y de gasto energético entre plantas C3, C4 y CAM son debidas a una respuesta ambiental. Cada uno de estos tipos se desarrolla en climas diferentes, y cada uno representa una adaptación a ese clima. Esto hace que el mayor gasto energético para la fijación de CO₂ que existe en plantas CAM y c4 tenga sentido. Las plantas C3 para fijar una molécula de O₂ gastan 3 moléculas de ATP y dos moléculas de NADPH, mientras que las plantas C4 y CAM gastan para lo mismo 5 o 6.5 moléculas de ATP respectivamente, y 2 de poder reductor. La conversión diurna de málico para formar almidón requiere ATP y justifica la diferencia en consumo energético.

CAM y C4 son tipos de plantas adaptadas a vivir en ambientes cálidos y áridos las primeras y cálidos pero más húmedos las segundas. En estos ambientes la apertura de estomas para dejar circular el aire y así poder fijar el CO₂ les supondría perdidas de agua, de ahí que las C4 y CAM utilicen mecanismos de acumulación de CO₂ que les permitan evitar esas perdida de agua.. Pero no sólo presenta esa ventaja la acumulación de CO₂ en la planta para después ser utilizado en el ciclo de Calvin. La Rubisco a altas temperaturas (a partir de los 30° C más o menos) pierde afinidad por el CO₂ con lo que el mayor gasto energético que se utiliza para acumularlo dentro de la planta queda compensado, ya que en el caso de una planta c3 tuviese humedad suficiente para realizar la fijación sin desecarse pasaría mucho tiempo en fase oxigenativa, lo cual representa un gasto energético extra (gasta ATP y poder reductor) y además no estaría fijando CO₂ con lo que el rendimiento sería inferior {Foyer, 1994 #530}.

Por tanto el mayor gasto energético de plantas C4 y CAM queda compensado en los ambientes en los que viven, ya que en esos ambientes un metabolismo tipo C3 sería menos rentable y en algunos casos inviable debido a la desecación.

Si por el contrario atendemos a la producción de biomasa entre estos 3 tipos de plantas nos encontramos con una serie de diferencias.

Biomasa bruta producirían menos las plantas C3, ya que estas alternan fase oxigenativa con fase carboxilativa. Por el contrario C4 y CAM prácticamente no presentan fotorrespiración debido a que acumulan CO₂ con lo que estas plantas producen más biomasa ya que aprovechan todo el CO₂ en formación de fotosintatos {Pietrini, 1999 #430}.

Pero si atendemos a biomasa neta en relación con la energía que se invierte, entonces son las plantas C3 las que más producen y esto es debido a que gastan menos energía (pese a alternar fase oxigenativa con fase carboxilativa) que las plantas CAM y C4, las cuales hacen un mayor gasto de ATP y poder reductor para acumular CO₂.

|

10.2.3.5. Resumen de diferencias entre los diferentes tipos de fotosíntesis

- **Diferencias en el gasto energético.** Las plantas C3 para fijar una molécula de O₂ gastan 3 moléculas de ATP y dos moléculas de NADPH, mientras que las plantas C4 y CAM gastan para lo mismo 5 o 6.5 moléculas de ATP respectivamente, y 2 de poder reductor. La conversión diurna de malato para formar almidón requiere ATP.
- **Uso eficiente del agua.** Mientras que las plantas C3 transpiran entre 450 y 950 g gramos de agua por cada gramo de materia seca que producen las plantas C4 y CAM pierden entre 250 -350 y 18-125, respectivamente. (referencias)
- **Uso eficiente del nitrógeno.** Las plantas C4 requieren de menos Rubisco para alcanzar la misma tasa fotosintética. Esto se traduce en un menor consumo de nitrógeno como fertilizante. (referencias)
- **Producción de biomasa.** Tomando en cuenta la biomasa bruta, las plantas C3 producen menos que C4 y CAM, pero si de eficiencia neta se trata, las plantas C3 producen más biomasa con menos energía, pues gastan menos ATP que las C4 y las CAM.

Para entender por qué existen otros procesos de fijación y asimilación de CO₂ con diferencias energéticas y metabólicas, es importante tomar en cuenta que:

- El CO₂ es menos soluble conforme aumenta la temperatura (Taiz y Zeiger, Plant Physiology on line, <http://4e.plantphys.net>), por lo que el costo de la fotorrespiración aumenta, al disminuir la concentración de CO₂
- Tanto Plantas C4 como CAM evitan en gran medida la fotorrespiración, por medio de mecanismos de concentración de

CO₂. En las plantas C4 la fijación y asimilación de CO₂ están separados espacialmente, permitiendo concentrar el CO₂ en el sitio de catálisis de Rubisco y evitando el ciclo de oxigenación (Leegood 2002). En las plantas CAM los procesos se encuentran separados temporalmente; los estomas se abren generalmente por la noche cuando la temperatura ha descendido, lo que además de hacer el CO₂ mas soluble, evita la pérdida de agua y la fotorrespiración (Dodd et al., 2002),

- Por lo anterior, los gastos energéticos y metabólicos extraordinarios que parecen presentar los mecanismos C4 y CAM, se ven compensados y llegan a ser menores conforme aumenta la temperatura

Tabla 10-1. Diferencias entre plantas C3, C4 y CAM

* Sin embargo, bajo condiciones de riego las plantas CAM se encuentran entre las más productivas conocidas.

10.2.3.6. Discriminación isotópica

Las plantas C3 y C4 se llaman así por la molécula que primero se detectaron Calvin y sus colaboradores en experimentos con radioisótopos de carbono 14. Como se mencionó anteriormente, en las plantas C3 el primer compuesto orgánico fabricado en la fotosíntesis tiene tres átomos de carbono y en el tipo C4 tiene 4. La biomasa que se produce por medio de estos tipos de fotosíntesis C3 o C4 se puede distinguir por su composición isotópica de sus átomos de carbono.

Si midiéramos el carbono atmosférico, casi el 99 % del CO₂ será del tipo que contiene el carbono ligero ¹²C. Una pequeña parte, el 1,1 % del CO₂, es algo más pesado, ya que contiene ¹³C. Y finalmente existe también en la atmósfera, en muy pequeña proporción, un tipo de CO₂ que contiene ¹⁴C, que es radiactivo e inestable, y cuyas aplicaciones han solidó ser fundamentalmente paleocronológicas.

Las plantas C3 y C4 tienen valores $\delta^{13}\text{C}$ muy diferentes. Esto se debe a que la enzima Rubisco discrimina el dióxido de carbono de una manera más fuerte que la enzima PEPCase. El 85 % de las plantas superiores son del tipo C3 (casi todas las arbóreas) y tienen unos valores de $\delta^{13}\text{C}$ muy bajos, entre -22 % y -30 %. El otro 15 % de las plantas son del tipo C4. En su mayoría son hierbas tropicales y tienen unos valores de $\delta^{13}\text{C}$ más altos, entre -10 % y -14 %.

Por lo tanto, el valor $\delta^{13}\text{C}$ del carbono de los paleosuelos depende en gran parte del tipo de planta que ha crecido en ellos. Es menor cuando han

dominado las plantas C3 y mayor cuando han proliferado las del tipo C4. Por eso, el estudio de las variaciones de $\delta^{13}\text{C}$ en los paleosuelos continentales nos puede dar indicaciones del tipo de plantas, C3 o C4, que han predominado en determinados períodos.

Indirectamente, el valor $\delta^{13}\text{C}$ de los paleosuelos puede también indicarnos la evolución de la concentración de CO_2 atmosférico. Ocurre que con concentraciones elevadas de CO_2 , las plantas de tipo C3 se ven favorecidas con respecto a las plantas de tipo C4, ya que las plantas de tipo C3 requieren menos energía para realizar la fotosíntesis. Por el contrario, cuando la concentración de CO_2 es baja, aumentan las del tipo C4, ya que poseen un mecanismo de concentración de CO_2 que las favorece. Por lo tanto, cuanto menor sea $\delta^{13}\text{C}$ en el paleosuelo analizado, más probabilidad hay que la concentración de CO_2 haya sido alta. Y viceversa.

La concentración de carbono 13 de las plantas también puede indicarnos la existencia de los períodos de sequía. Durante las sequías algunas plantas tienden a cerrar sus estomas para perder menos agua. Entonces, al haber disponible menos CO_2 entrante, las plantas discriminan menos al carbono 13 y su concentración en los azúcares aumenta. Al parecer, antes del Mioceno (hace 15 millones de años), las plantas C4 eran casi inexistentes. De ahí que se piense que la disminución de CO_2 en el Mioceno, causada quizás por una mayor meteorización ligada a la emersión del Tibet, pueda haber originado el desarrollo de las plantas C4, y que el avance de las hierbas tropicales, que suelen ser de tipo C4, favoreció la evolución de los grandes mamíferos de la sabana {Foyer, 1998 #445}.

Especies típicas de Importancia económica	C3 cebada, frijol, tomate	Trigo, papa, arroz,	C4 Maíz, caña de azúcar, perla	sorgo, mijo	CAM Piña, nopal
% de la flora mundial en numero de especies	89%		<1%		10%
Hábitat típico	Distribución amplia		Sitios cálidos y praderas		Sitios xéricos y epifíticos
Primer producto estable de la fijación de CO_2	PGA		Malato Aspartato		Malato
Anatomía	Vaina del haz vascular no presente o sin cloroplastos		Vaina del haz vascular con cloroplastos (Kranz)		Suculencia celular o de los tejidos

Fotorrespiración	Hasta 40% de la fotosíntesis	No detectable	No detectable
Punto de compensación para la asimilación de CO ₂	40-100 µl/L	0-10 µl/L	0-10 µl/L
[CO ₂] intracelular en luz de día (µl/L)	200	100	10000
Frecuencia estomática (estomas mm ⁻²)	40 - 300	100 - 160	1 - 8
EUA (g CO ₂ fijado por kg H ₂ O transpirada)	1 - 3	2 - 5	10 - 40
Tasa máxima de crecimiento (g m ⁻² d ⁻¹)	5-20	40-50	0.2
Productividad máxima. (ton ha ⁻¹ año ⁻¹)	10-30	40-80	Generalmente menor a 10*

10.3. Tejidos vegetales

10.3.1. Monocotiledóneas y dicotiledóneas

Los nombres dicotiledóneas y monocotiledóneas se refieren al hecho de que el embrión de una dicotiledónea tiene dos cotiledones (hojas seminales) y el embrión de una monocotiledónea tiene uno. Es decir, la clasificación taxonómica se basa en diferencias en la estructura de algunos órganos y tejidos. Las diferencias características se hacen visibles en el cuerpo de la planta. Por ejemplo, los haces vasculares principales de las hojas de las dicotiledóneas habitualmente se disponen en red; los haces vasculares de las hojas de las monocotiledóneas habitualmente son paralelos. En los tallos de las dicotiledóneas, los haces de tejido vascular se disponen alrededor de un núcleo central de tejido fundamental; en los tallos de las monocotiledóneas, los haces vasculares están dispersos en el tejido fundamental. Existen también diferencias características en el número de piezas florales y en el número de hendiduras de los granos de polen entre ambas clases de angiospermas.

Embriones	Hojas	Tallos	Piezas florales	Granos de polen
Dicotiledónea				
 Dos cotiledones	 Nervadura normalmente ramificada	 Haces vasculares dispuestos radialmente	 Normalmente cuatro o cinco (o múltiplos)	 Tres poros o hendiduras
Monocotiledónea				
 Un cotiledón	 Nervadura paralela	 Haces vasculares esparcidos	 Normalmente tres o múltiplos de tres	 Un poro o hendidura

Figura 10-23. Las dos clases de angiospermas son las dicotiledóneas y las monocotiledóneas.

Las células del embrión de una angiosperma se diferencian tempranamente en su desarrollo en tres tejidos distintos: la protodermis, el procambio y el meristema fundamental. Estos tejidos embrionarios, conocidos como meristemas primarios, producen los tres sistemas de tejidos que son continuos en todo el cuerpo de la planta. La protodermis, el primer tejido que se diferencia, es el origen del sistema de tejido dérmico, que proporciona una cubierta externa protectora para todo el cuerpo de la planta. El procambio, el tejido que se diferencia luego, origina el sistema de tejido vascular, compuesto por xilema y floema. El xilema transporta agua y minerales disueltos, mientras que el floema transporta azúcares disueltos y otros compuestos orgánicos desde las células fotosintéticas (autótrofas) de las hojas y de los tallos verdes a las células no fotosintéticas (heterótrofas) de la planta. Los tejidos vasculares están rodeados por el sistema de tejido fundamental, derivado del meristema embrionario fundamental. Las principales diferencias en la estructura de las hojas, tallos y raíces residen en la distribución relativa de los sistemas de tejido vascular y fundamental.

Las células que se encuentran más frecuentemente en el cuerpo de la planta son del tipo conocido como parénquima. Estas células, que aparecen en los tres sistemas de tejidos y predominan en los tejidos fundamentales, son típicamente poliédricas y tienen paredes delgadas y flexibles. Además de

un núcleo, mitocondrias y otras organelas, las células parenquimáticas generalmente contienen plástidos que, según la ubicación de la célula, pueden ser cloroplastos, leucoplastos o cromoplastos.

Además de la fotosíntesis, las células parenquimáticas desempeñan una variedad de funciones esenciales en la planta, que incluyen la respiración y el almacenamiento de alimento y agua. Cada uno de los sistemas de tejidos contiene también tipos celulares adicionales, especializados en funciones particulares de cada órgano.

10.3.2. Semillas

Los embriones de muchas angiospermas pasan por un período de latencia antes de que germine la semilla. Con la germinación se reinicia el crecimiento, se rompe la cubierta de la semilla y surge el esporofito joven. Las primeras hojas de follaje se abren hacia el sol y comienzan a hacer fotosíntesis, mientras que, internamente, continúa el proceso de crecimiento que da origen al cuerpo de la planta. En los primeros estadios del desarrollo, el crecimiento del joven esporofito depende de las reservas acumuladas en la semilla.

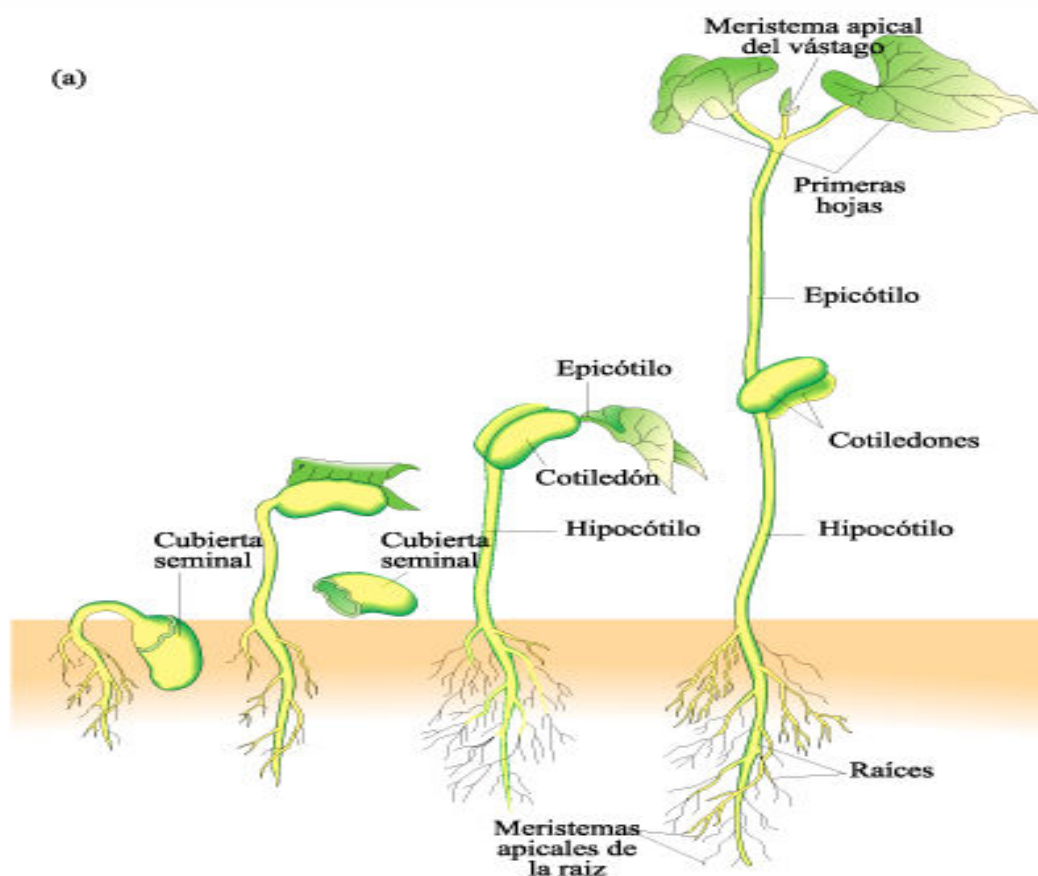


Figura 10-24. Desarrollo de una plántula dicotiledónea (frijol).

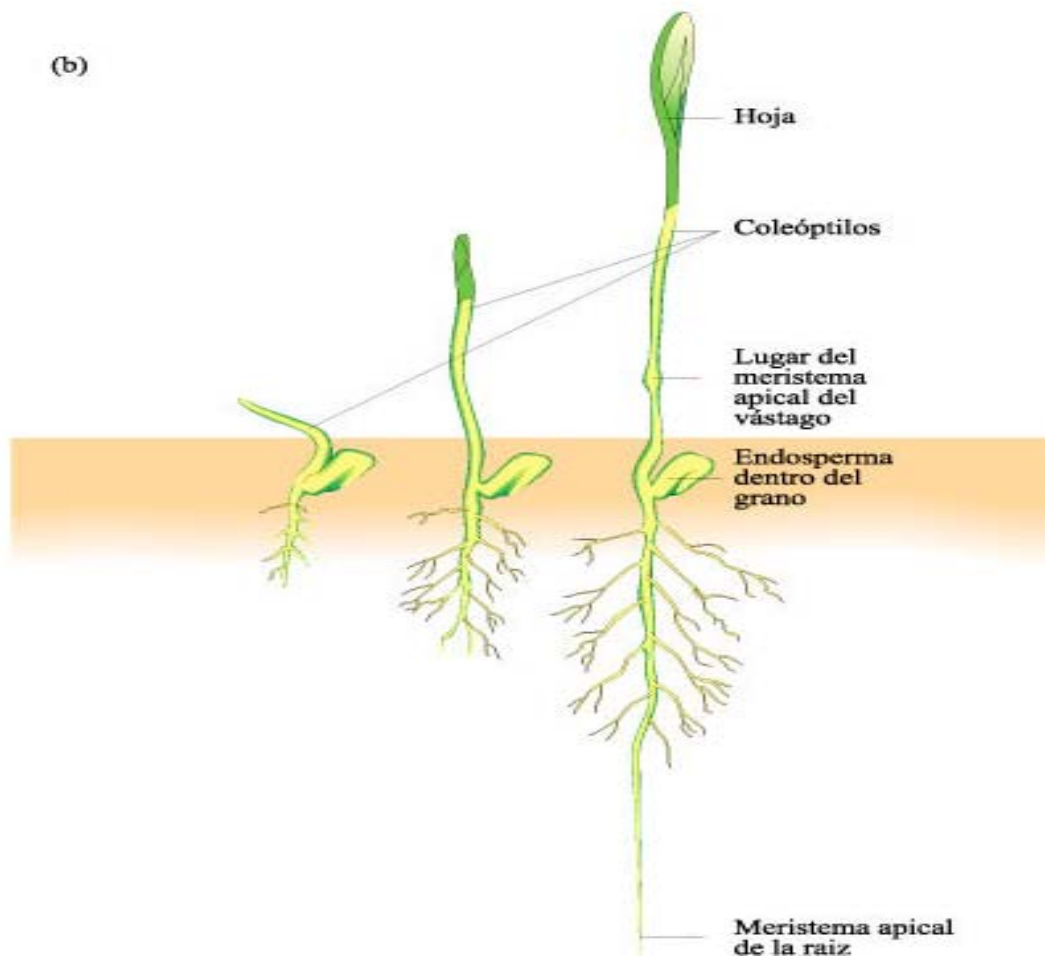


Figura 10-25. Desarrollo de una plántula monocotiledónea (maíz).

Antes de la germinación, la semilla absorbe agua y se hincha, rompiendo la cubierta seminal. Primero emerge la raíz joven, seguida del hipocótilo ("debajo de los cotiledones"). Los cotiledones finalmente se marchitarán y caerán. Luego emerge el epicótilo que se encuentra entre cotiledón y haz de hojas.

La primera estructura que aparece por encima del suelo es el coleóptilo, que forma una vaina cilíndrica sobre el vástago en crecimiento de la planta. Típicamente, el resto de endosperma, con el escudete (el cotiledón único) en su interior, está presente en la joven plántula.

El crecimiento primario de la planta implica la diferenciación de los tres sistemas de tejido, el alargamiento de las raíces y tallos, y la formación de las raíces laterales y de las ramas. Después de completarse el desarrollo del embrión, el crecimiento primario posterior se origina en los meristemos apicales de la raíz y del vástago. Las plantas, a diferencia de los animales, continúan creciendo durante todo su ciclo de vida.

10.3.3. Hojas

Las hojas son las principales áreas fotosintéticas de una planta. Las células fotosintéticas de las hojas son células parenquimáticas que forman dos

tipos de tejidos: parénquima en empalizada, constituido por células alargadas y densamente empaquetadas ubicadas justo por debajo de la superficie superior de la hoja, y parénquima esponjoso, que consiste en células de contorno irregular situadas en el interior de la hoja y con grandes espacios intercelulares. Estos espacios están llenos de gases, que incluyen vapor de agua, oxígeno y dióxido de carbono. La mayor parte de la fotosíntesis ocurre en las células en empalizada, que están especializadas en la captación de la luz.

El parénquima en empalizada y el parénquima esponjoso constituyen el tejido fundamental de la hoja, conocido como mesófilo. El mesófilo está envuelto casi herméticamente por las células epidérmicas, que secretan una sustancia cerosa llamada cutina, formando una cubierta denominada cutícula, sobre la superficie externa de la epidermis. Las células epidérmicas y la cutícula son transparentes, lo que permite que la luz las atraviese y penetre en las células fotosintéticas.

Las sustancias entran y salen de las hojas a través de dos estructuras completamente diferentes: los haces vasculares y los estomas. El agua y los minerales disueltos son transportados a las hojas, y los productos de la fotosíntesis son transportados fuera de ellas, por medio de los haces vasculares. Los haces vasculares atraviesan los pecíolos y se continúan con los tejidos vasculares del tallo y la raíz.

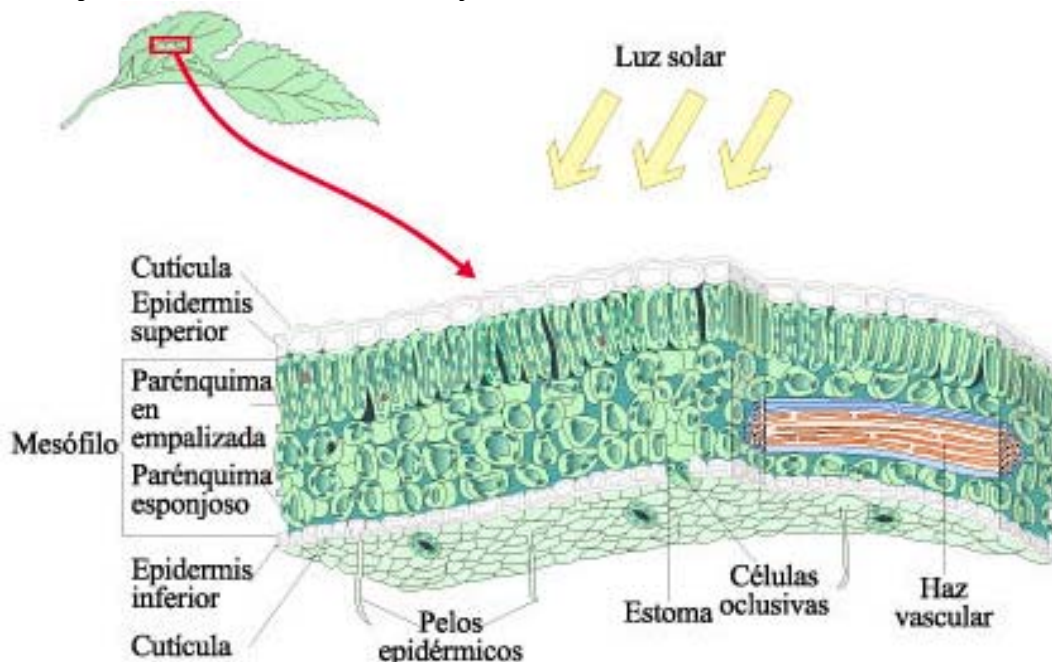


Figura 10-26. Estructura celular de una hoja.

Ejemplos de hojas: Las hojas presentan una variedad de formas y tamaños, que van desde frondes grandes a escamas diminutas. Estas diferencias en la morfología y tamaño guardan una estrecha relación con los ambientes en los cuales vive la planta.

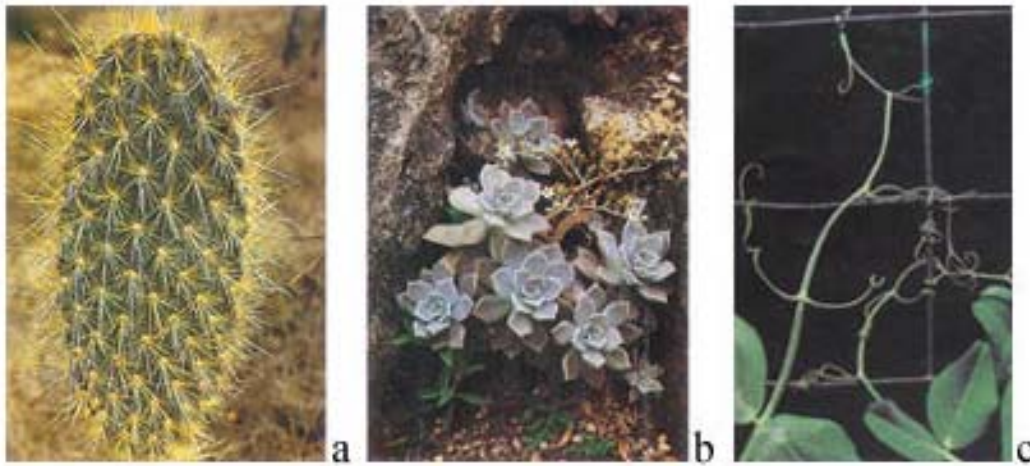


Figura 10-27. Hojas modificadas. a) Espinas de un nopal. b) Hojas succulentas adaptadas al almacenamiento de agua (Sedum). c) Zarcillo de una planta de guisante. En la planta de guisante, que tiene hojas compuestas, sólo los folíolos individuales se modifican como zarcillos; otros folíolos de hojas compuestas son aplanados, lo que proporciona una superficie amplia para la fotosíntesis.

10.3.4. Raíces

Las raíces son estructuras especializadas que fijan la planta al suelo e incorporan agua y minerales esenciales. La raíz embrionaria, o radícula, es la primera estructura que rompe la cubierta seminal y se elonga rápidamente.

La estructura interna de la raíz de las angiospermas es comparativamente simple. En las dicotiledóneas y en la mayoría de las monocotiledóneas los tres sistemas de tejidos (dérmico, fundamental y vascular) están dispuestos en tres capas concéntricas: la epidermis, la corteza y el cilindro central.

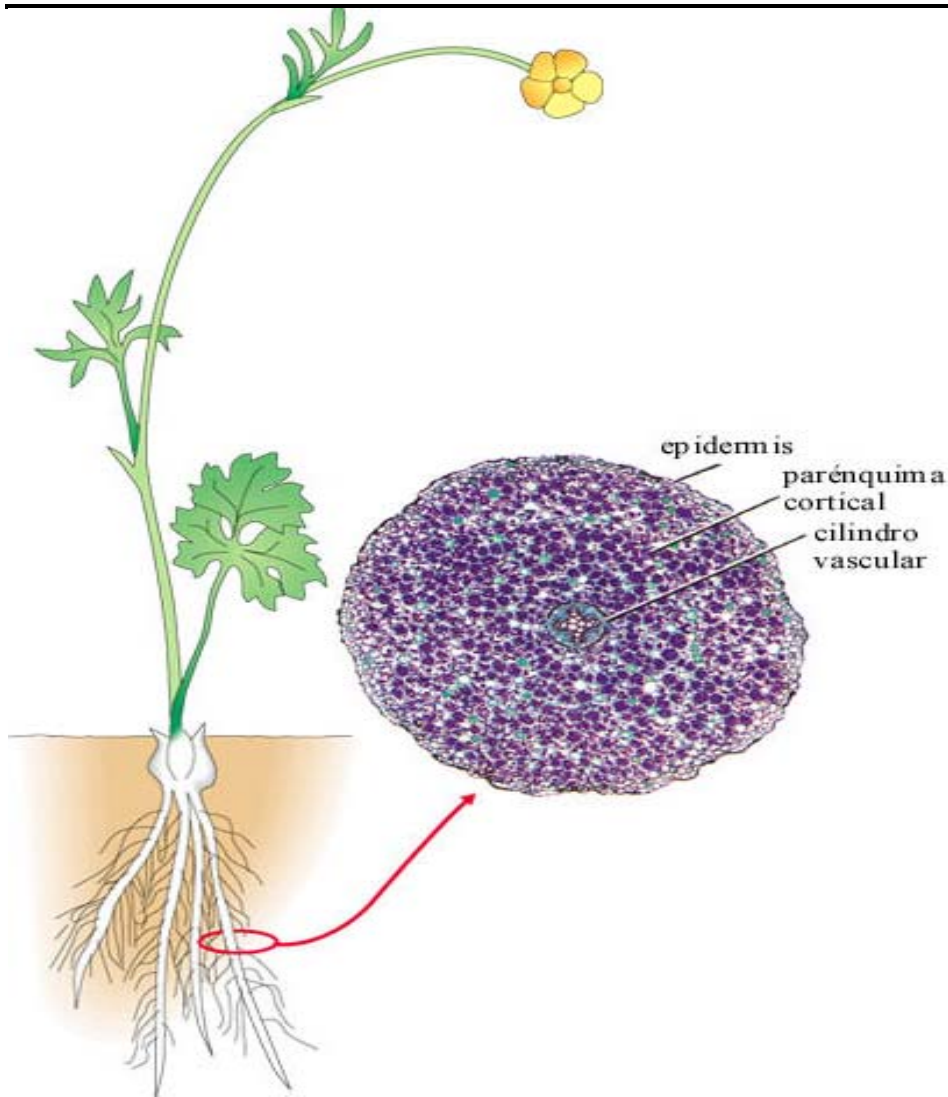


Figura 10-28. Raíz de un ranúnculo, una dicotiledónea, en corte transversal. Los plástidos en las células parenquimáticas de la corteza contienen granos de almidón, teñidos en púrpura en esta preparación.

Las raíces jóvenes tienen una capa externa de epidermis y, a lo sumo, una cutícula muy delgada. Prolongaciones de las células epidérmicas forman los pelos radicales, que incrementan en sumo grado la superficie absorbente de la raíz. Debajo de la epidermis está el tejido fundamental de la raíz, la corteza, compuesta principalmente por células parenquimáticas, frecuentemente especializadas en el almacenamiento. La capa más interna de la corteza es la endodermis, una sola capa de células especializadas cuyas paredes contiguas, en las que se ha depositado suberina, conforman la banda de Caspary.

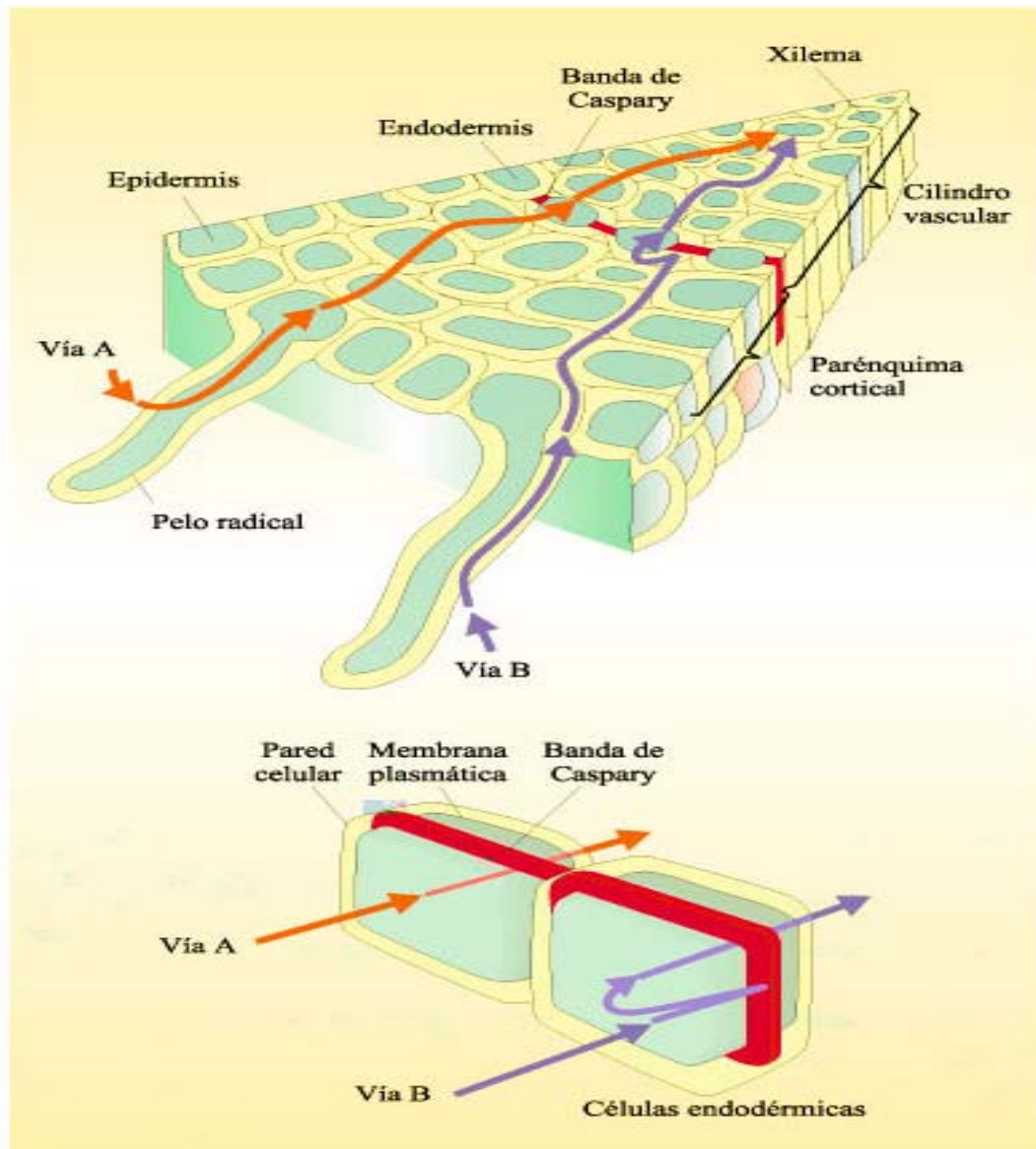


Figura 10-29. Diagrama de un corte transversal de una raíz que muestra las dos vías de absorción de agua y sustancias disueltas.

La mayor parte de los solutos y parte del agua que entra en la raíz siguen la vía A, que se indica en color rojo; los solutos penetran a las células de la epidermis de la raíz por transporte activo o por difusión facilitada; en tanto que el agua se mueve por diferencia de gradiente de potencial hídrico. Luego de atravesar las membranas celulares, ya en el simplasma de la raíz, el agua y los solutos se desplazan de una a otra célula a través de los plasmodesmos. El simplasma es la continuidad del protoplasma a través de los plasmodesmos. Otra parte del agua y algunos de los solutos que entran a la raíz siguen la vía B, indicada en color azul, moviéndose a través de las paredes celulares y a lo largo de sus superficies (por el apoplasto). Nótese, sin embargo, la ubicación de la banda de Caspary y de qué manera bloquea la vía B en torno a todo el cilindro vascular de la raíz.

Para atravesar la banda de Caspary, tanto el agua como los solutos tienen que ser transportados a través de las membranas celulares de las células

endodérmicas o de otras células situadas más externamente, a través de la vía A. Después que el agua y los solutos han cruzado la endodermis, la mayoría de los solutos continúan a lo largo de la vía A hasta las células conductoras del xilema, y la mayor parte del agua retorna a la vía B para cubrir la distancia que resta hasta las células del xilema. En las zonas más jóvenes de la raíz, donde aún no se ha formado la banda de Caspary, el agua y los nutrientes esenciales pueden alcanzar el xilema a través de la vía B.

Inmediatamente por dentro de la endodermis hay otra capa de células, el periciclo, del cual surgen las ramificaciones de la raíz. Por dentro del periciclo están el xilema y el floema.

Las Figuras 9.17 y 9.18 muestran las diferencias de las estructuras celulares entre una raíz de una planta monocotiledónea y dicotiledónea.

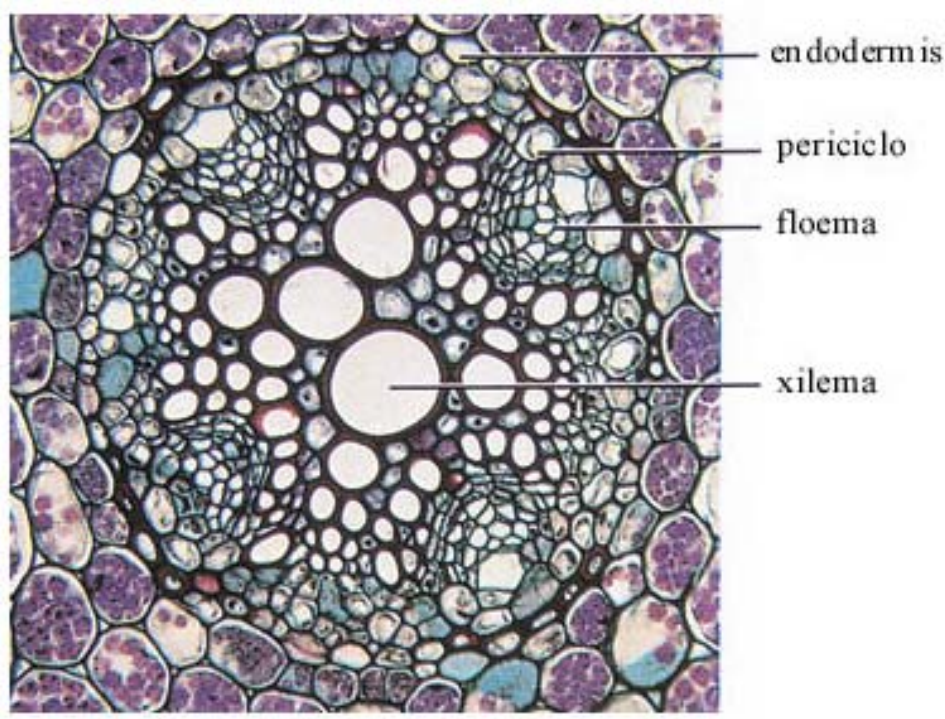


Figura 9.17 Detalle del cilindro vascular de la raíz del ranúnculo (una dicotiledónea).

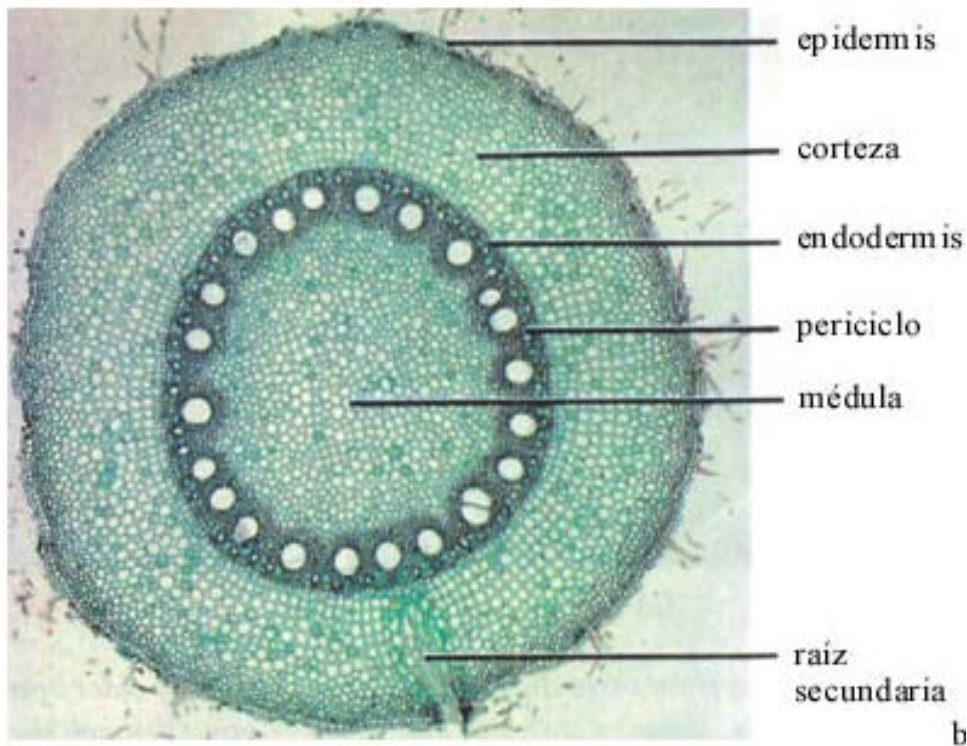


Figura 9.18 Corte transversal de la raíz de una planta de maíz (una monocotiledónea), que muestra el cilindro vascular que rodea la médula. La endodermis, que está por fuera del periciclo, se considera parte de la corteza. La endodermis contiene las bandas de Caspary

10.3.5. Tallos

Los tallos portan las hojas y flores de la planta. Son la vía por la cual las sustancias se transportan desde las raíces hacia las hojas, y viceversa.

Los tallos verdes, al igual que las hojas, tienen una capa externa de células epidérmicas cubiertas con una cutícula. La masa del tallo joven es tejido fundamental, que puede dividirse en un cilindro externo (la corteza) y un núcleo interno (la médula). El tejido fundamental está compuesto principalmente por células parenquimáticas, pero también puede contener células colenquimáticas y esclerenquimáticas (fibras y esclereidas). Las figuras 9.19 y 9.20 muestran algunos ejemplos de cortes tallos de diferentes especies.

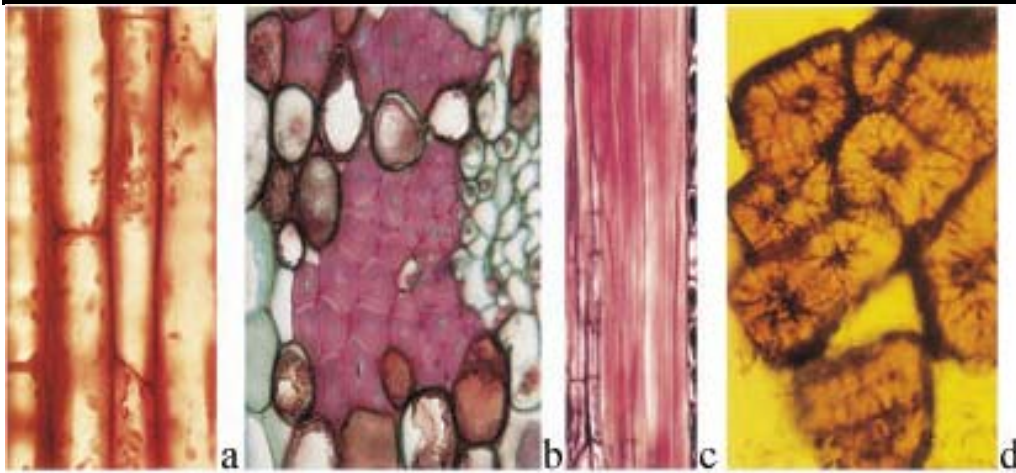
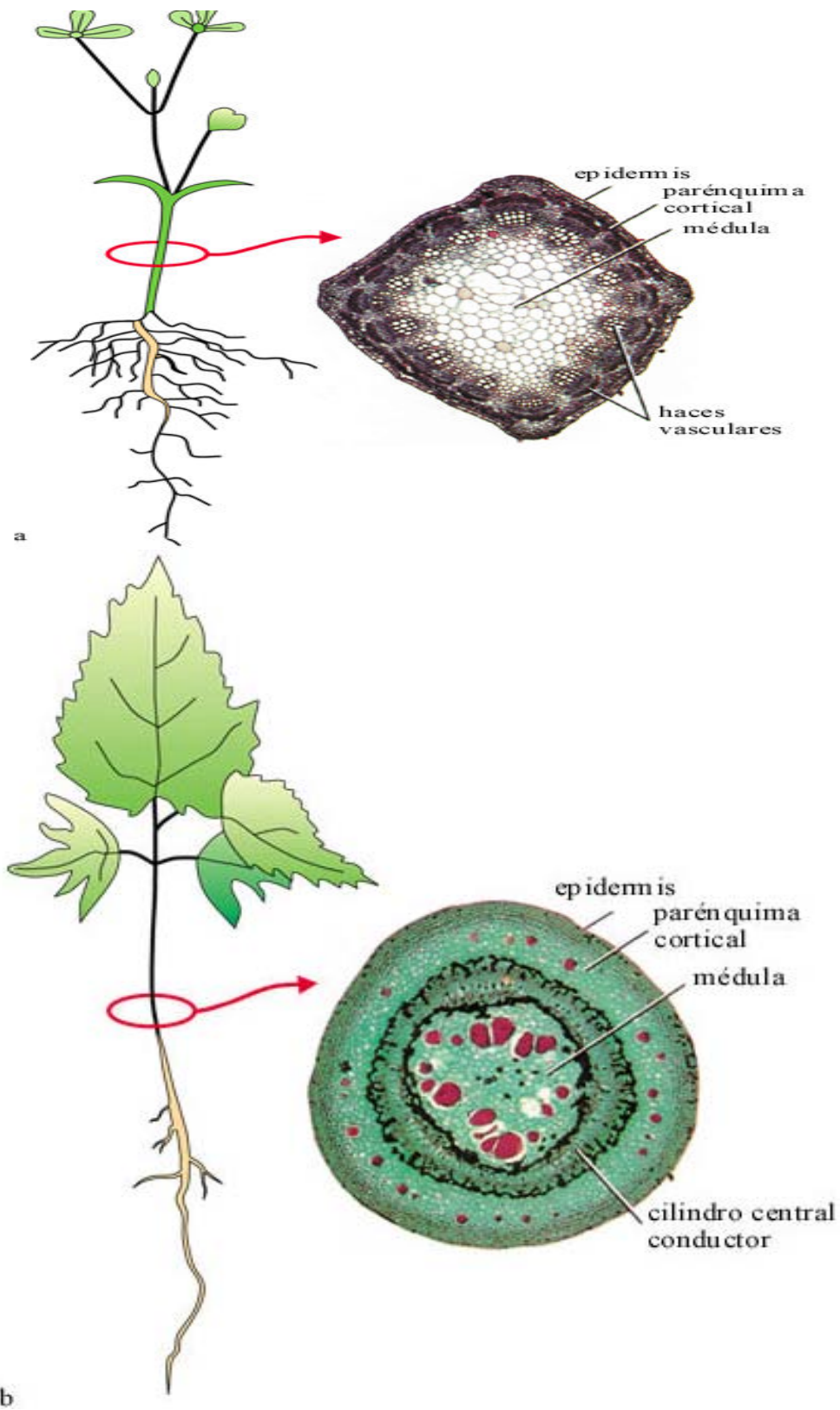


Figura 10-30. Algunos tipos de células del tejido fundamental de los tallos.

a) Células colenquimáticas vistas en corte longitudinal. b) Corte transversal y c) corte longitudinal de fibras del floema del tallo de un tilo (*Tilia americana*). Sólo una porción de la longitud de las fibras puede verse en c). d) Las esclereidas, otro tipo de células esclerenquimáticas, tienen paredes lignificadas muy gruesas.

Diferencias de cortes transversales de plantas dicotiledóneas y monocotiledóneas:



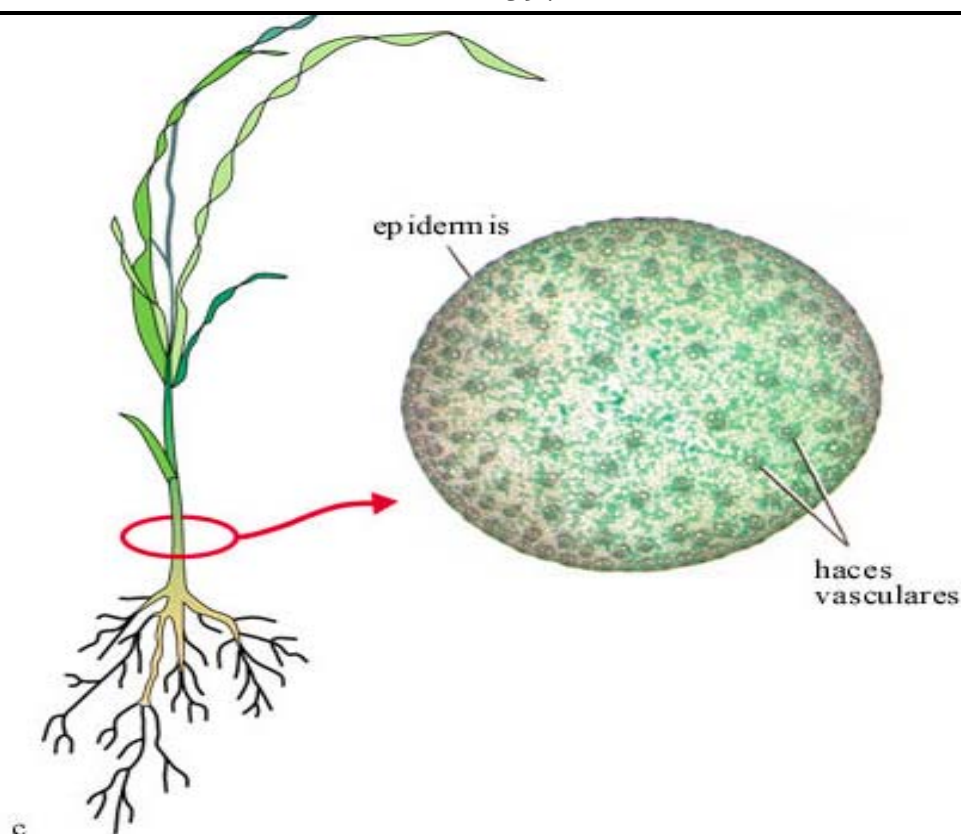


Figura 9.22 Cortes transversales de dos tallos de dicotiledónea y de un tallo de monocotiledónea. a) En la alfalfa, una dicotiledónea, el cilindro vascular está formado por haces vasculares separados. b) En este tallo joven del tilo, también una dicotiledónea, el tejido vascular forma un cilindro continuo. Este tallo contiene conductos de mucílago, que se tiñen de rojo. c) En el maíz, una monocotiledónea, numerosos haces vasculares están dispersos en el tejido fundamental.

10.3.6. Tejido Vascular

10.3.6.1. Floema

El floema es un tejido especializado en transportar azúcares a larga distancia. Está formado por dos tipos celulares relacionados ontogenéticamente: elementos cribosos (ECs) y células acompañantes (CAs). Los ECs son células altamente modificadas que durante la maduración pierden el núcleo y las mitocondrias. Poseen abundantes perforaciones en las paredes celulares de la zona donde se unen a otros ECs para formar conductos conocidos como tubos cribosos por donde viajan los azúcares. Las CAs se caracterizan por tener un protoplasma denso, núcleo y numerosas mitocondrias, se encuentran conectadas a los ECs por una red de plasmodesmos, de tal forma que la funcionalidad metabólica de los ECs depende de los procesos que ocurren en las CAs. Figura 9.20

Los ECs de la mayoría de las plantas contienen proteínas específicas conocidas como proteínas-P, las cuales pueden formar filamentos, túbulos

y agregados cristalinos. Las proteínas-P de la familia Cucurbitaceae son las que mejor se han caracterizado, pues se les puede purificar fácilmente a partir del exudado del floema. En *Cucurbita máxima* se han encontrado dos polipéptidos: PP1 de 96 KDa que forma filamentos y la PP2 de 48 KDa que funciona como lectina. Las proteínas PP1 y PP2 son responsables de formar un sello cuando los ECs sufren heridas.

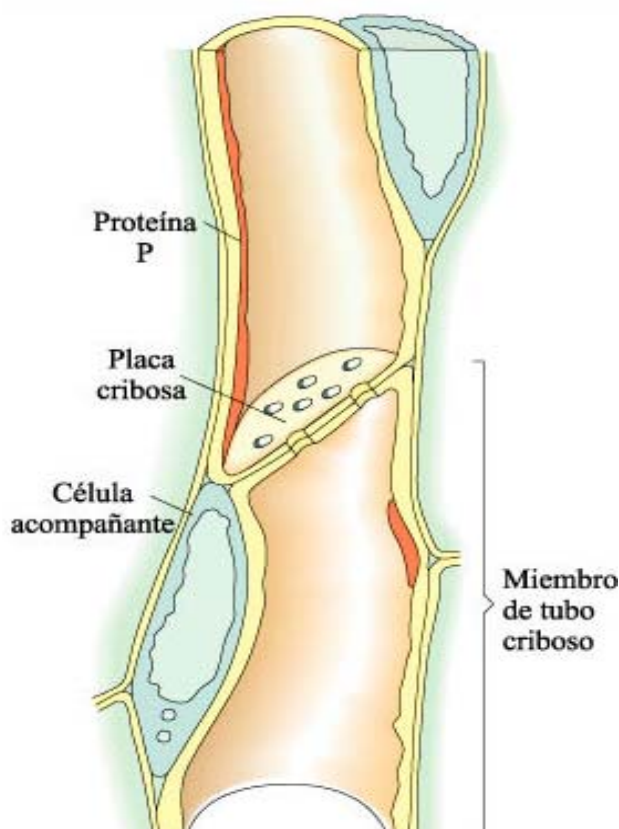


Figura 9.20 En las angiospermas, las células de conducción del floema son los miembros de tubo criboso, células vivas con paredes terminales perforadas, que forman tubos cribosos continuos. Asociada íntimamente con cada miembro de tubo criboso hay una célula acompañante.

10.3.6.2. Simplasto y apoplasto

Un aspecto de gran importancia en la distribución de fotosintatos, es la forma en que tanto las células del mesófilo de las hojas como las del tejido demanda están comunicadas con el floema, pues ello define la ruta que seguirá el movimiento de los fotosintatos. Al movimiento de los azúcares por medio de plasmodesmos se denomina vía simplástica y el gradiente de concentración determina la magnitud y la dirección del flujo de fotosintatos (Figura 9.22). Existe otra vía denominada apoplástica, la cual se caracteriza porque los azúcares que viajan por el floema se liberan al espacio intercelular de las células que forman los órganos de demanda. En este caso, la cantidad de azúcares que dichas células asimilen dependerá de la actividad de transportadores ubicados en la membrana plasmática.

La sacarosa que se descarga en el apoplasto puede ser incorporada por las células del tejido demanda mediante un transportador específico y/o puede

ser hidrolizada por la invertasa de pared celular. Las membranas plasmáticas de las células de los órganos de demanda también expresan transportadores de hexosas que permiten la entrada de la glucosa y fructosa, producidas por la invertasa.

Algunos de los transportadores de azúcares han sido caracterizados en sistemas heterólogos. Se ha determinado que funcionan como simporter (H^+ /sacarosa ó H^+ /hexosas) y su actividad permite el movimiento de azúcares en contra del gradiente de concentración. Se ha sugerido que la predominancia de las rutas de descarga de fotosintatos simplástica y apoplástica está relacionada con la forma en que se usan los fotosintatos. En tejidos que acumulan almidón la descarga es por vía simplástica, pues el almidón además de ser osmóticamente menos activo que la sacarosa, se compartamentaliza en los amiloplastos, lo cual ayuda a que se mantenga el gradiente de concentración, que favorece la asimilación de fotosintatos. Por otro lado, en tejidos con descarga apoplástica generalmente se acumulan azúcares libres, algunos de los cuales se almacenan en la vacuola, si bien esto favorece la formación de un gradiente de concentración que facilita a la asimilación de azúcares la participación de los transportadores ubicados en la membrana celular es fundamental. Figura 9.22

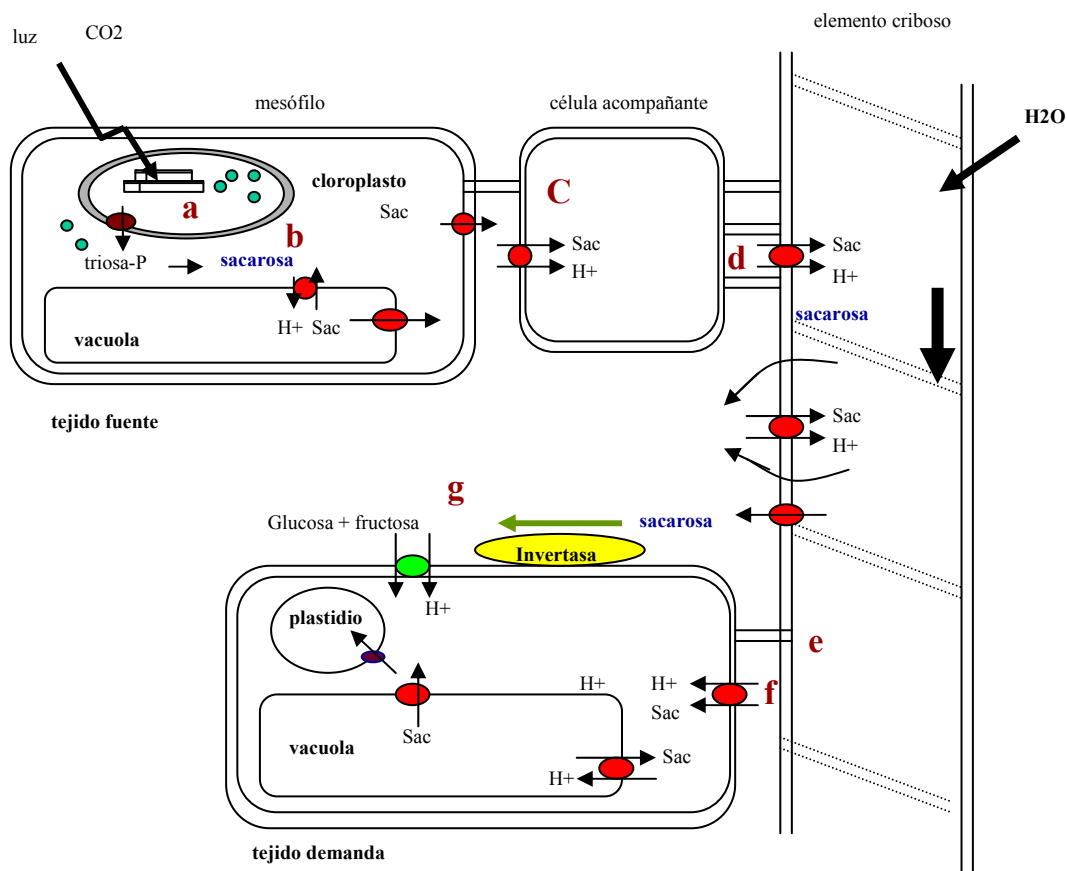


Figura 9.21. Ruta de fotosintatos de tejido fuente hacia tejido demanda . a. La planta realiza fotosíntesis en el cloroplasto. b. Mediante en ciclo de Calvin se forman triosas fosfato que son exportadas el citosol y se forma

sacarosa, donde puede ser almacenada en la vacuola o ser exportada hacia tejidos demanda. c. Mecanismo de vía simplasto, la sacarosa viaja a través de conductos llamados plasmodemata hasta el floema. d. Mecanismo vía apoplásto, la sacarosa es liberada al apoplásto y transportadores de sacarosa H^+ /sacarosa cargan la sacarosa hacia el floema. e y f. Descarga del floema vía simplasto y apoplásto respectivamente. g. La sacarosa puede ser hidrolizada por una invertasa de pared cuyos productos; glucosa y fructosa y son incorporados por transportadores de hexosas H^+ /hexosa. Tomado de Lalonde et al., 1999.

10.3.6.3. Xilema

El xilema se encarga de trasladar la savia bruta desde la raíz hacia la parte proximal de la planta. La savia bruta se compone en su mayor parte de agua e iones inorgánicos, aunque algunos compuestos orgánicos pueden estar presentes. La energía para este transporte no la proporcionan los mismos elementos traquearios, que en el tejido desarrollado están de hecho muertos, sino por dos fenómenos físicos:

La [ósmosis](#), que desplaza hacia arriba el agua acumulada en la raíz gracias a la diferencia en potencial soluble del tejido radical y la humedad del suelo; al absorber agua, la raíz impulsa hacia arriba parte de la misma. Este fenómeno, sin embargo, no basta para llevarla hasta las hojas, y su intensidad varía enormemente entre especies; en [Vitis riparia](#) alcanza los 145 kPa, mientras que en [Celastrus orbiculatus](#) es virtualmente cero.

La succión, que atrae hacia las hojas el agua contenida en el tejido vascular para compensar la pérdida de la misma por la [transpiración](#) a través de las hojas.

El tejido de conducción del xilema está constituido por una serie de traqueidas o vasos. Se trata de un tejido leñoso de los vegetales superiores que conduce agua y sales inorgánicas en forma ascendente por toda la planta y proporciona también soporte mecánico. En las hojas, las flores y los tallos jóvenes, el xilema se presenta combinado con floema en forma de haces vasculares conductores. Las raíces tienen un cilindro central de xilema. El xilema formado a partir de los puntos de crecimiento de tallos y raíces se llama primario. Pero además, la división de las células del cámbium, situado entre el xilema y el floema, puede producir nuevo xilema o xilema secundario; esta división da lugar a nuevas células de xilema hacia el interior en las raíces y hacia el exterior en casi todos los tallos. Algunas plantas tienen muy poco xilema secundario o ninguno, en contraste con las especies leñosas; el término botánico xilema significa madera.

Por otro lado, el xilema puede contener tres tipos de células alargadas: traqueidas, elementos vasculares o vasos (tráqueas) y fibras. En la madurez, cuando desempeñan funciones de transporte, todas estas células están muertas. Las traqueidas son células alargadas con paredes gruesas caracterizadas por la presencia de zonas delgadas muy bien definidas

llamadas punteaduras. Los elementos vasculares o vasos son traqueidas especializadas cuyas paredes terminales están atravesadas por uno o varios poros; una serie vertical de elementos vasculares que forman un tubo continuo se llama vaso. Las fibras son traqueidas especializadas de pared muy engrosada que apenas realizan funciones de transporte y que sirven para aumentar la resistencia mecánica del xilema.

El xilema de las especies más antiguas desde el punto de vista de la evolución, como los helechos y las coníferas, está formado por traqueidas. En casi todas las **angiospermas** (plantas con flor), el xilema contiene también vasos y fibras bien desarrollados. Como las secuencias de especialización de todos estos elementos tisulares se observan con bastante claridad, el estudio del xilema aporta importantes claves para dilucidar la evolución de las plantas superiores.

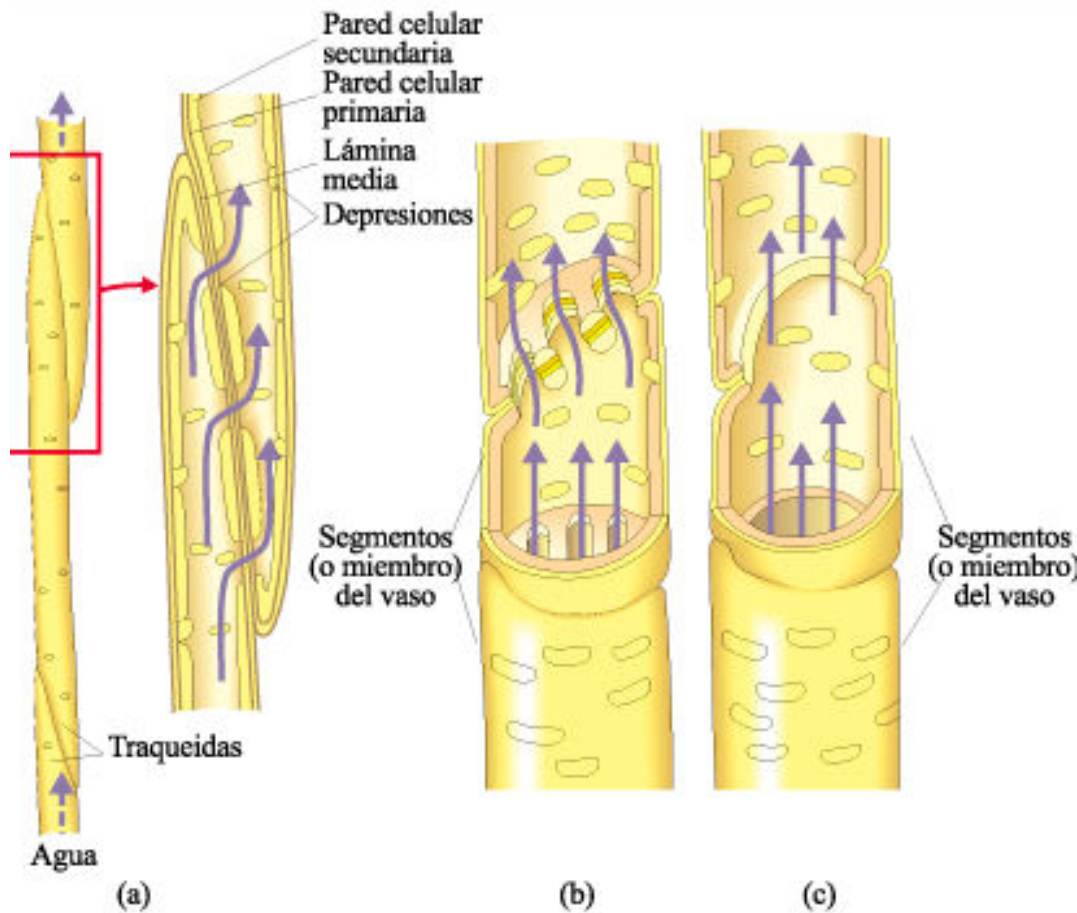


Figura 10-31. Elementos del xilema

Las traqueidas y los vasos, las células que componen el xilema en las angiospermas. a) Las traqueidas son un tipo más primitivo y menos eficiente de célula conductora que predominan en las gimnospermas. El agua que se mueve de una traqueida a otra pasa a través de depresiones. Estas depresiones no son perforaciones, sino simplemente áreas en las cuales no hay una pared celular secundaria. Por lo tanto, el agua que se mueve de una traqueida a otra pasa a través de dos paredes celulares

primarias y de la laminilla media. Los vasos, característicos de las angiospermas, difieren de las traqueidas en que las paredes primarias y las laminillas medias de los vasos están perforadas en los extremos, donde se unen con otros vasos. b) Puede haber numerosas perforaciones en células contiguas de miembros de vaso, o c) las paredes contiguas pueden disolverse por completo cuando las células maduran formando una sola abertura. Los vasos se caracterizan también por ser más cortos y más anchos que las traqueidas y sus paredes contiguas son menos oblicuas. Los vasos se conectan con otros vasos y también con otras células por depresiones de las paredes laterales.

10.4. Sistemas de transporte vascular

10.4.1. Transporte por el xilema

Figura 10-32. a) formas de vaso del xilema. Fig. B: miembro de vaso en el xilema del quebracho blanco, MEB 700x. Fig c: Las células del floema conducen alimento desde las hojas al resto de la planta. Las células del floema están ubicadas por fuera del xilema. Fig. D: Esquema de células de colénquima en corte transversal. Fig. E: A la derecha esquema de braquisclereidas de pera (*Malus sylvestris*). Al centro esquema de las astroesclereidas del pecíolo de *Nymphaea* sp. (Planta acuática) y macrosclereidas del la cubierta seminal de la arveja (*Pisum sativum*). A la izquierda esquema de las fibras, en vista longitudinal y en corte transversal. Fig. F: esquema de las células parenquimáticas; Fig. G: imagen de microscopía electrónica de barrido (MEB) de las células del parénquima medular de un tallo de amor seco (*Bidens pilosa*) 430x.

10.4.1.1. Apoplasto y Simplasto

En el siglo 19, el fisiólogo alemán E. Munch introdujo un concepto de valor en la discusión del camino del agua en la planta. El sugirió que las paredes celulares interconectadas y los elementos del xilema llenos de agua deberían considerarse como un único sistema que denominó apoplasto. Es decir en cierto sentido, la parte no viva del vegetal. Consiste de todas las paredes celulares en la corteza de la raíz y técnicamente las paredes de la endodermis, aunque para nuestros propósitos no incluiría las **bandas de Caspary** de las células endodérmicas que tienen suberina que es hidrófoba. Todas las traqueidas y vasos en el xilema serían también parte del apoplasto como también lo son todas las paredes celulares del resto de la planta, incluyendo aquellas de tejidos tales como floema, otras células en la corteza y en las hojas. Se entiende que el ELA no sería equivalente al apoplasto, puesto que termina en las bandas de Caspary. Excepto por el engrosamiento de la endodermis, el ascenso del agua en el vegetal tendría lugar por el apoplasto, particularmente la porción de xilema, pero

incluyendo las paredes celulares de la corteza y aún las de las células vivas en las hojas. Munch también denominó al resto de la planta, la porción viva, como simplasto. Está incluiría el citoplasma de todas las células en la planta. Se ha encontrado que el citoplasma de células adyacentes se encuentra conectado por pequeños poros en las paredes celulares. Esas conexiones citoplasmáticas reciben el nombre de plasmodesmos.

La hipótesis formulada por Dixon a principios de siglo sigue teniendo el mayor atractivo en la explicación del ascenso del agua. Uno de sus elementos es la propiedad de adhesión del agua. Hay muchos datos en la actualidad que apoyan esta teoría. En ella hay tres elementos básicos, a saber:

- la fuerza impulsora,
- la hidratación en el camino por adhesión a las sustancias del camino y
- la cohesión de las moléculas del agua.

La fuerza impulsora es, como se vio, el gradiente decreciente en el potencial agua desde el suelo hacia la atmósfera a través de la planta. El agua se mueve en el camino desde el suelo a través de la epidermis, corteza y endodermis, dentro del tejido vascular de las raíces, hacia arriba a través del xilema del leño, al interior de las hojas y finalmente por transpiración a través de los estomas hacia la atmósfera. Y es la estructura especial de este camino, es decir el relativamente pequeño diámetro y espesor de las paredes de los tubos, y la propiedad de imbibición de las paredes celulares del parénquima foliar los que hacen que el sistema sea funcional. Las fuerzas de imbibición o hidratación entre las moléculas de agua y las paredes celulares, se deben a uniones hidrógeno y es llamada adhesión, la cual es una fuerza atractiva entre moléculas disímiles. La tercera clave es la cohesión, esto es, la fuerza de atracción mutua entre moléculas de agua que también se debe a sus uniones H^+ . En el ambiente del camino del agua, las fuerzas cohesivas son tan grandes que el agua puede ser tirada desde la punta de los árboles altos por la evaporación en la atmósfera, y este estiramiento puede extenderse por todo el trayecto hacia abajo a través del tronco y las raíces y hasta el suelo mismo.

Figura 10-33. Absorción por los pelos radiculares. Ruta del agua y nutrientes absorbidos hasta las raíces. Apoplasto y simplasto. Papel de la Endodermis. Banda de Caspary.

Figura 10-34. Capilaridad y presión radicular. Teoría de la tensión cohesión. Evidencias.

Estado metaestable de la columna de agua en el xilema. Tensión o presión negativa de la columna de agua en el xilema. Cavitación. Protección.

10.4.1.2. Movimientos de agua en el tallo

El agua se mueve a través de las plantas por medio de los tejidos vasculares. La fuerza motriz para que se den movimientos de agua dentro del xilema se deriva de la transpiración en las hojas. La fuerza motriz para que se de el movimiento dentro del floema se deriva del metabolismo de los carbohidratos en los tejidos *source* y tejidos *sink*. El agua penetra en la planta a través del sistema radicular y pasa por un filtro selectivo en las células de la endodermis. El agua asciende por los tubos del xilema, desde las raíces al resto de la planta. El agua es un disolvente universal y por eso lleva consigo sales minerales y algunos productos de las reacciones químicas de la planta como metabolitos y hormonas.

El agua también circula por los tubos cribosos del floema, para transportar el azúcar por las hojas durante la fotosíntesis. Las raíces no disponen de otro medio de nutrición y morirían sin la presencia de estos azúcares translocados por las hojas.

Por medio de la translocación, el agua aporta turgencia y rigidez a las células. Estas características favorecen el anclaje de las hojas y tejidos nuevos, también es el medio por el que se realiza la transpiración, amortigua los cambios de temperatura, estabiliza el pH de su metabolismo. Mantiene el volumen del citoplasma que está formado mayoritariamente por agua.

La transpiración de la planta es la evaporación del agua efectuada desde la hoja y desde el resto de las superficies de la planta. La transpiración es el último paso del recorrido del agua, habiéndose iniciado en el suelo y continuado por toda la planta.

Tanto la respiración como la fotosíntesis aportan agua y calor. La primera enfría el tejido vegetal, haciendo que las plantas transpiren. Alrededor del 90% de toda el agua que penetra en la planta por medio de las estructuras llamadas estomas. Por la acción de las “células guarda”, los estomas se abren durante las horas diurnas para permitir el libre intercambio y la liberación de vapor de agua.

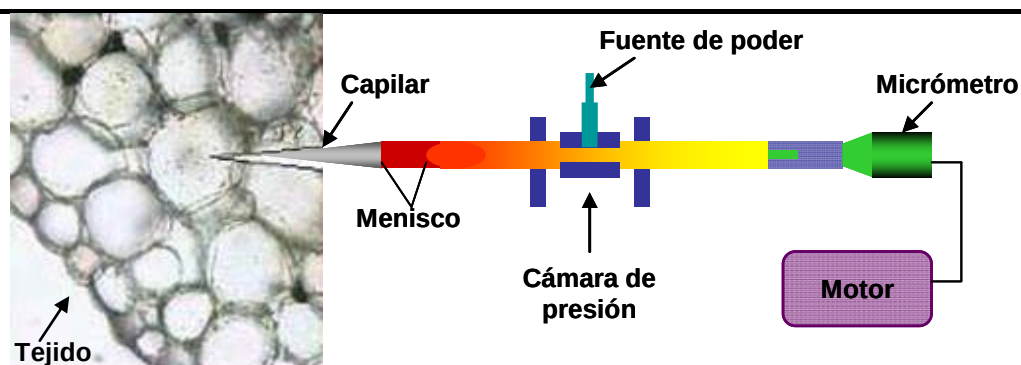


Figura 10-35. Prueba para determinar la presión positiva del floema

El agua procedente de las precipitaciones o el riego penetra en el suelo, donde toma contacto con el sistema radicular. La evapotranspiración es el proceso mediante el cual se expulsa agua de los suelos mediante la evaporación y la traspiración de las plantas.

El estrés hídrico es más significativo en una planta que en otra, dependiendo de donde se produce dicho estrés. No solamente las plantas, que dejan caer las yemas florales, las flores e incluso los frutos pequeños sin madurar, sino las semillas germinales pueden morir con gran facilidad por falta de agua. El maíz, por ejemplo, puede manifestar una pérdida de grano en la mazorca si no se aplica la cantidad adecuada de agua durante la polinización. La mayoría de las plantas muestran indicios de necesidades hídricas ante la aparición de trastornos irreparables. Las hojas se decoloran y pueden llegar a marchitarse o retorcerse, como en el caso del maíz. Otras plantas como la yuca tienen la capacidad de tirar sus hojas un día (marchitarse) y recobrarse cuando vuelve a llover.

El exceso de agua es tan perjudicial como la escasez. Las semillas y las plantas también necesitan oxígeno para germinar y crecer. Las semillas y las raíces de las plantas en suelos encharcados, no tienen oxígeno suficiente para el mantenimiento de sus procesos vitales. El exceso de humedad en las raíces produce la aparición de enfermedades, por lo que esta parte de la planta se puede pudrir y morir. Las plantas se amarillean y se deforman. Entre las formas de agua no disponible en el suelo para las plantas se incluyen:

- El agua gravitacional que se pierde en el drenaje
- El agua higroscópica la que se fija a las partículas del suelo.

10.4.1.3. Movimientos de agua en las hojas

La pérdida de agua por parte de las plantas en forma de vapor se conoce como transpiración. Es una consecuencia inevitable de la apertura de los estomas. Esta apertura es necesaria pues a través de los estomas ingresa el dióxido de carbono que se utiliza en la fotosíntesis. Si se cierran los

estomas entonces se puede ahorrar agua, pero por otro lado, no se puede llevar a cabo la fijación de dióxido de carbono.

A medida que el dióxido de carbono, esencial para la fotosíntesis, penetra en las hojas por los estomas se pierde vapor de agua a través de éstos. Aunque esta pérdida de agua plantea problemas serios para las plantas, suministra la fuerza motriz mediante la que se absorbe agua por las raíces. Además provee un mecanismo que enfría las hojas. Esto ocurre porque el agua, al evaporarse, lleva consigo calor. La temperatura de una hoja puede ser hasta 10 ó 15 ° C inferior a la del aire circundante. Esto se puede medir por medio de un termómetro de infrarojo. De hecho, algunos mejoradores usan estos aparatos de medición de la temperatura de las hojas para detectar diferencias fisiológicas entre diferentes variedades.

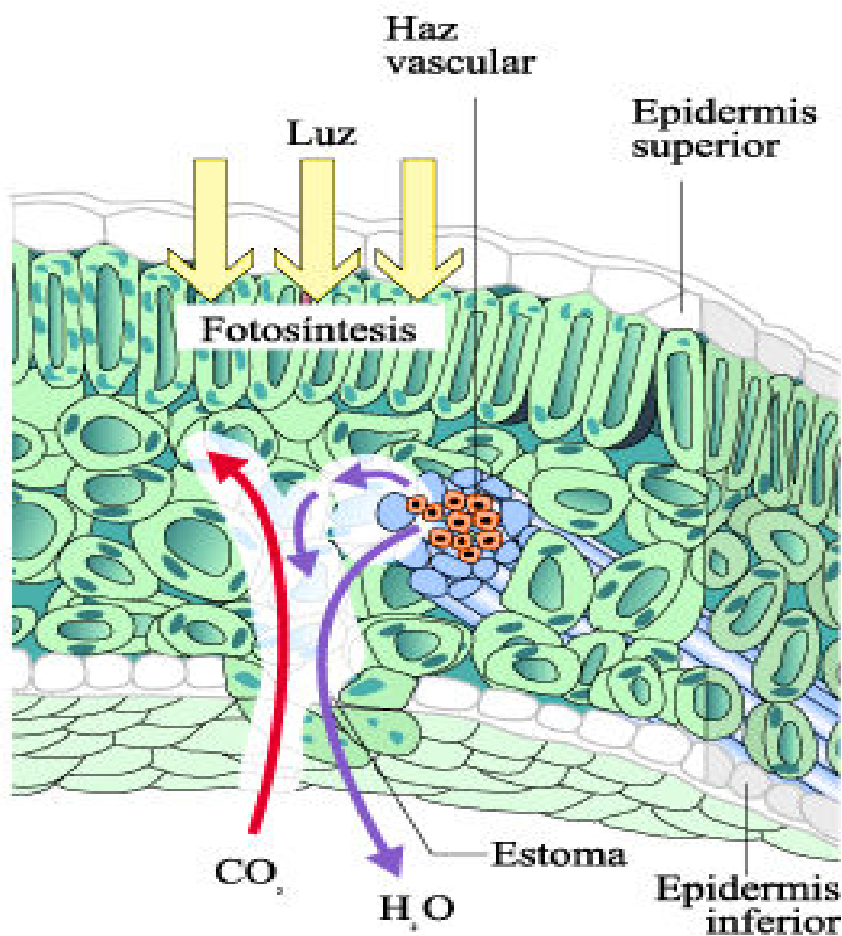


Figura 10-36. Intercambio de gases por los estomas.

El agua entra en la planta desde el suelo por las raíces. El movimiento del agua hacia las células de la raíz sólo es posible cuando el potencial hídrico en el suelo es mayor al potencial hídrico en las raíces. Los datos sugieren que la pérdida de agua genera fuerzas que permiten su absorción. Los procesos que conducen a la entrada de agua a las células de la raíz son capaces por sí solos -y bajo ciertas condiciones- de generar una presión

positiva. Tal presión, conocida como presión de raíz es, sin embargo, sólo suficiente para que el agua ascienda un corto trecho en el tallo.

El agua viaja a través del cuerpo vegetal en las células conductoras del xilema (vasos y traqueidas).

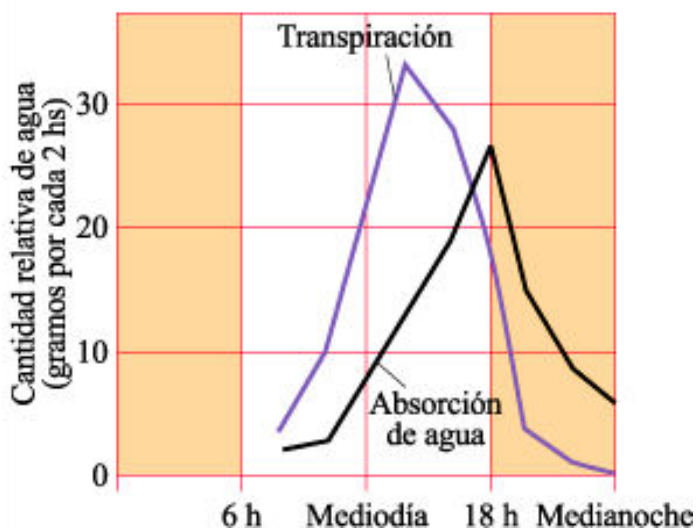


Figura 10-37. Mediciones efectuadas en árboles de fresno muestran que un aumento en la transpiración es seguido por un aumento en la absorción de agua.

De acuerdo con la teoría de cohesión-tensión, el agua se mueve en las traqueidas y vasos bajo presión negativa (presión menor a la atmosférica, también denominada tensión). Dado que las moléculas de agua se mantienen juntas (cohesión), hay una columna continua de moléculas de agua que es arrastrada por tracción, desde la solución que se encuentra en el suelo al interior de la raíz, molécula por molécula, debido a la evaporación del agua en la parte superior.

10.4.1.4. Transpiración y estomas

La pérdida de agua por parte de las plantas en forma de vapor es una consecuencia inevitable de la apertura de los estomas. La transpiración es un proceso fundamental para que las sales minerales puedan moverse desde el suelo hasta la hoja y para controlar la temperatura del vegetal. Esta apertura es necesaria pues a través de los estomas ingresa el dióxido de carbono que se utiliza en la fotosíntesis (Figura 10-38.). A medida que el dióxido de carbono, esencial para la fotosíntesis, penetra en las hojas por los estomas se pierde vapor de agua a través de estos. Esta pérdida de agua plantea problemas serios para las plantas. La transpiración también suministra la fuerza motriz mediante la que se absorbe agua por las raíces. Además provee un mecanismo que enfría las hojas. La temperatura de una hoja puede ser hasta 10 ó 15 ° C inferior a la del aire circundante. Esto ocurre porque el agua, al evaporarse, lleva consigo calor.

El proceso de transpiración de las plantas produce la presión negativa que jala el agua hacia arriba. Este proceso continúa hacia las raíces, donde el agua en los espacios extra celulares que rodean al xilema es jalada hacia adentro por las perforaciones de las paredes de los elementos de los vasos y las traqueidas. Este movimiento del agua hacia arriba y hacia adentro finalmente causa que el agua presente en el suelo se mueva hacia el cilindro vascular por ósmosis a través de las células endodérmicas. La fuerza generada por la evaporación del agua desde las hojas, transmitida hacia abajo por el xilema hacia las raíces, es tan fuerte que se puede absorber agua de los suelos bastantes secos.

La transpiración tiene efectos positivos y negativos. Los positivos le proporcionan la energía capaz de transportar agua, minerales y nutrientes a las hojas en la parte superior de la planta. Los negativos son la mayor fuente de pérdida de agua, pérdida que puede amenazar la supervivencia de la planta, especialmente en climas muy secos y calientes.

Casi toda el agua se transpira por los estomas de las hojas y del tallo, por lo tanto una planta al abrir y cerrar sus estomas debe lograr un equilibrio entre la absorción de CO_2 para la fotosíntesis y la pérdida de agua de la transpiración. El flujo de agua es unidireccional desde la raíz hasta el brote porque sólo éste puede transpirar (Figura 10-X)

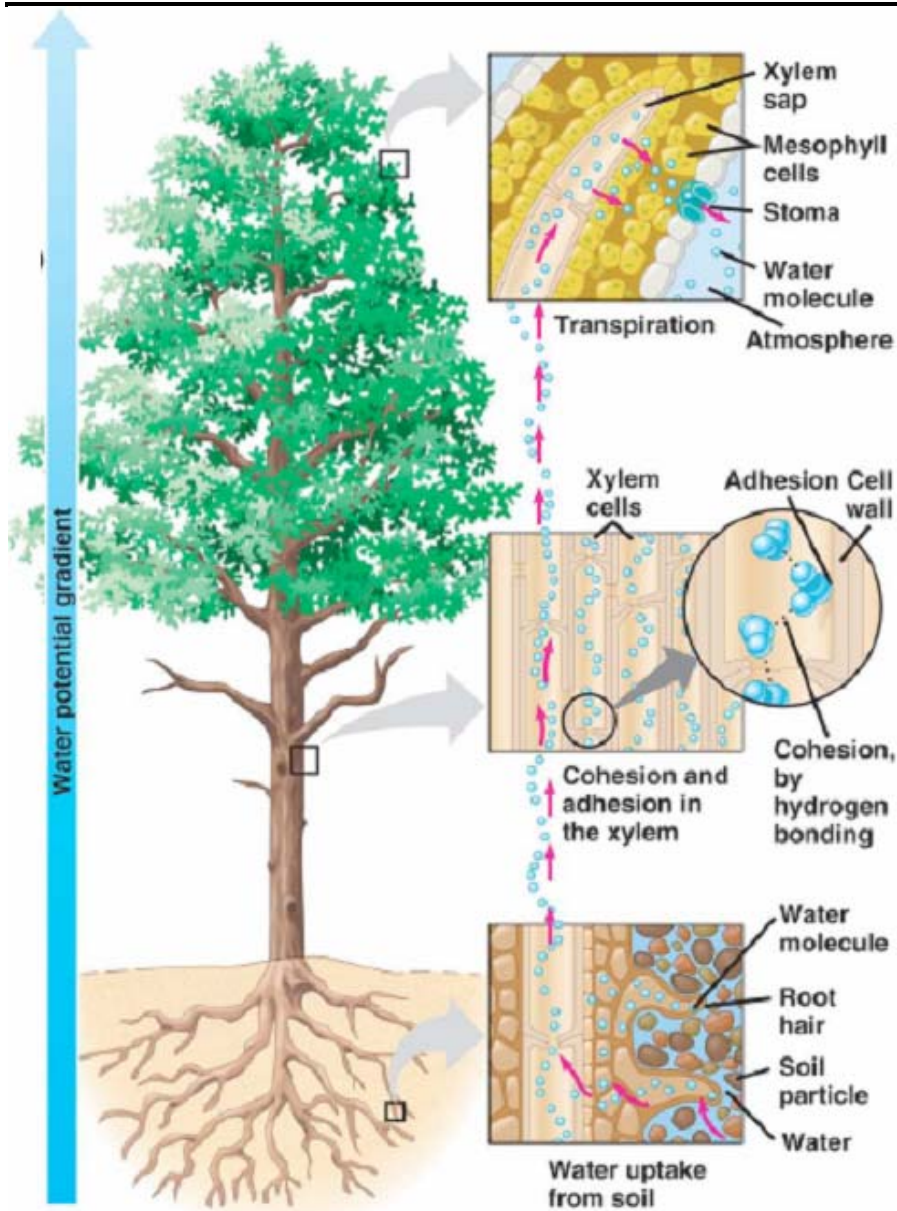


Figura 10-X. La transpiración provoca la evaporación en las paredes celulares del mesófilo, generando tensión en el xilema. La cohesión entre las moléculas de agua en el xilema transmite la tensión desde las hojas a la raíz, provocando así, el movimiento de agua desde el suelo hacia la atmósfera (tomado de <http://www.bufordcityschools.org/bhs/teachers/melissagreen/documents/34Ch36PlantTransport2005a.pdf>).

Figura 10-38. Mecanismo de movimiento estomático.

10.4.1.4.1. Estomas

El oxígeno y el dióxido de carbono entran y salen de las hojas a través de estomas. La difusión de los gases, incluyendo al vapor de agua, hacia el

interior y exterior de la hoja es regulada por los estomas. Para ello los estomas se abren y se cierran por la acción de las células oclusivas, debido a cambios en la turgencia. La turgencia de estas células aumenta o disminuye por el movimiento del agua, que sigue al movimiento de iones potasio hacia adentro o hacia afuera de las células oclusivas. Diversos factores concurren a regular la apertura y cierre de estomas, los cuales incluyen el estrés hídrico, la concentración de dióxido de carbono, la temperatura y la luz.

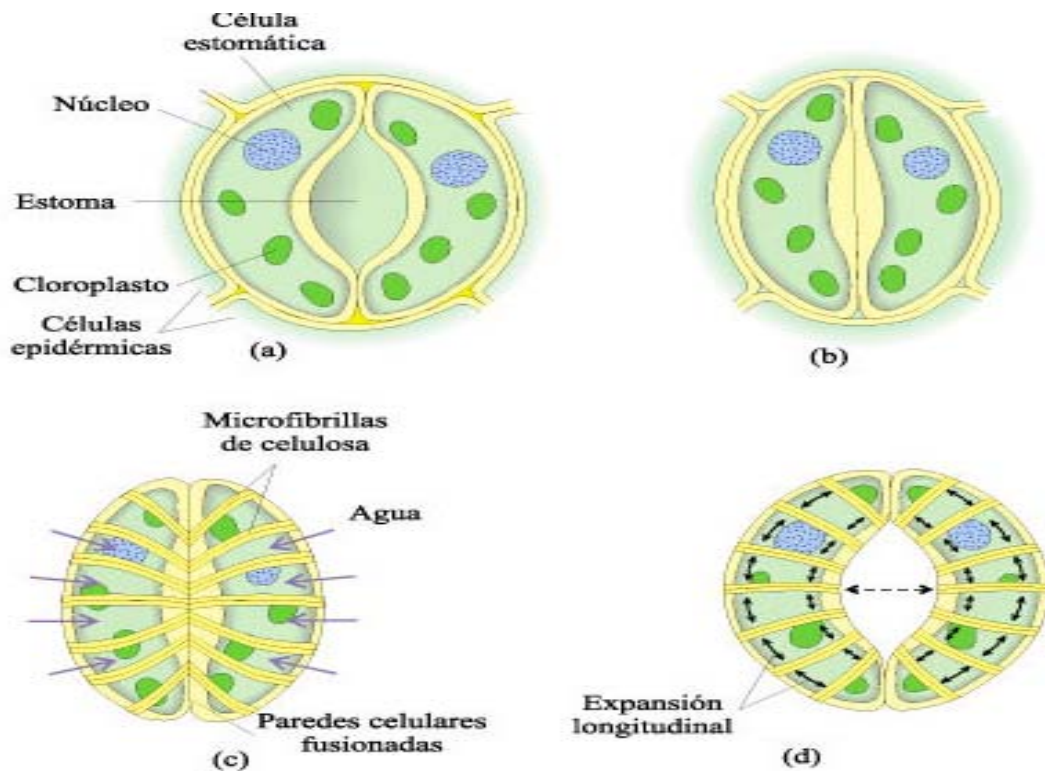


Figura 9.25 Mecanismo de movimiento estomático.

Un estoma está bordeado por dos células oclusivas que: a. abren el estoma cuando están turgentes b. lo cierran cuando pierden turgencia. La clave de la apertura de los estomas reside en las microfibrillas de celulosa dispuestas alrededor de las células oclusivas. c. Cuando el agua entra a las células oclusivas, las células sólo pueden expandirse en dirección longitudinal. d. Como las dos células están unidas por los extremos, esta expansión longitudinal las obliga a arquearse y al estoma a abrirse.

10.4.2. Transporte por el floema

En las plantas superiores, el floema es un tejido vascular que conduce azúcares y otros nutrientes sintetizados desde los órganos que los producen hacia aquéllos en que se consumen y almacenan (en forma ascendente y descendente). El floema está organizado en haces vasculares, que son los filamentos longitudinales del tejido conductor, asociados con el tejido

conductor de agua o xilema. Los haces vasculares constituyen importantes órganos estructurales de los tallos herbáceos y los nervios de las hojas. En el cilindro vascular que atraviesa el centro de la raíz del ranúnculo, por ejemplo, el xilema forma un núcleo central estrellado en cuyas ranuras se insertan los haces de floema. De forma típica, el xilema ocupa el lado del haz vascular más próximo a la médula, aunque no son raras disposiciones distintas. En las partes más viejas de la planta, las células blandas del floema son aplastadas y empujadas hacia afuera por el floema nuevo que se va formando en el proceso de crecimiento. El floema nuevo se crea por la acción del cámbium o zona de crecimiento, una capa celular que separa el xilema del floema y produce células de este segundo tipo hacia el exterior de la planta.

El floema consta de dos tipos de células conductoras: tubos cribosos, que son los elementos más característicos, y células anexas. Los tubos cribosos son células alargadas con las paredes de los extremos perforadas por numerosos poros diminutos; a través de ellos pueden pasar las sustancias disueltas. Estos elementos están conectados en series verticales. Las células están vivas cuando llegan a la madurez, pero los núcleos se desintegran antes de iniciar la función conductora. Las células anexas, más pequeñas, conservan los núcleos durante la madurez y también están vivas; se forman junto a los tubos cribosos y se cree que controlan el proceso de conducción.

Varios tipos celulares forman el floema, tejido vegetal encargado del transporte de una solución de sacarosa y otras muchas sustancias orgánicas conocidas tradicionalmente como savia elaborada. Estos tipos celulares son: los elementos de los tubos cribosos con sus células anexas, las células cribosas, las fibras y las células parenquimáticas (Figura 10-39).

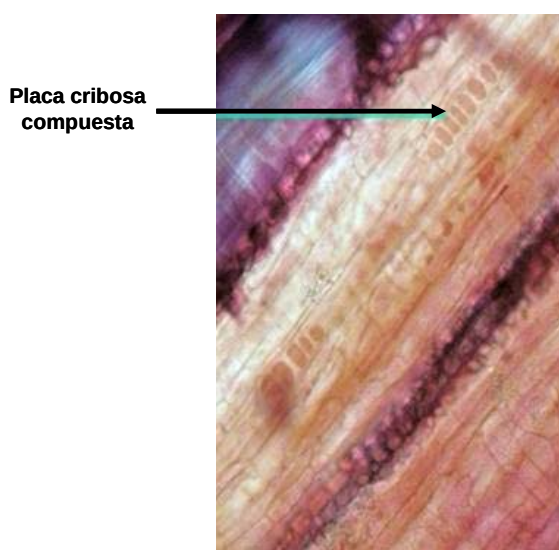


Figura 10-39 .Corte longitudinal de tallo de chopo (*Populus boleana*) (200x) mostrando una región del floema secundario donde pueden

observarse algunas placas cribosas compuestas y algunas células del parénquima floemático axial. <http://www.inea.uva.es/servicios/hist>

El floema puede tener fibras de líber, que son muy fuertes, y en algunas especies constituyen la materia prima de la que se obtienen fibras comerciales, como lino y yute, utilizadas en la confección de tejidos, arpillera y sacos o costales.

10.4.2.1. Transporte de fotoasimilados

La actividad metabólica de los diferentes órganos (o partes de órganos vegetales) requiere el aporte de fotoasimilados en cantidades diversas. En algunos casos, los procedentes de la actividad fotosintética de ese órgano, o bien de la hidrólisis de reservas acumuladas previamente en él, pueden satisfacer y sobrepasar los niveles señalados por estas necesidades; el órgano se autoabastece y está en condiciones de exportar fotoasimilados. En otros casos, el órgano puede ser claramente deficitario y debe importar fotoasimilados. El transporte de fotoasimilados a larga distancia, de un órgano a otro, se denomina translocación y se lleva a cabo, en general, por el floema.

Así como el agua y los solutos inorgánicos ascienden a través del xilema, o corriente de transpiración, los azúcares manufacturados durante la fotosíntesis salen de la hoja a través del floema, o corriente de asimilables (Figura 10-40) hacia lugares donde se utilizan, como el vástago en crecimiento y la caliptra de la raíz, así como a lugares de almacenamiento como frutos, semillas y el parénquima de almacenamiento de tallos y raíces.

Figura 10-40. Diagrama de la hoja que muestra los caminos seguidos por las moléculas de agua de la corriente de transpiración a medida que se mueven desde el xilema de un vaso menor hacia las células mesofíticas, se evaporan de las superficies de las paredes de las células mesofíticas, y se difunden después fuera de la hoja a través de un estoma abierto (líneas continuas). También se muestran los caminos seguidos por las moléculas de azúcar producidas durante la fotosíntesis a medida que se mueven desde las células mesofíticas al floema del mismo vaso y entran en la corriente de asimilación. Se cree que las moléculas de azúcar producidas en el parénquima en empalizada se dirigen al parénquima esponjoso y después lateralmente al floema a través de las células esponjosas (líneas discontinuas).

Las células conductoras del floema de las Angiospermas son los elementos cribosos que carecen de núcleo y de la mayoría de los orgánulos, pero son ricos en una proteína filamentosa específica del floema, llamada proteína P.

Los **elementos cribosos** forman series longitudinales llamadas tubos cribosos. Los elementos cribosos presentan poros, que forman áreas cribosas en las paredes laterales, y placas cribosas en las paredes transversales. Las placas cribosas posibilitan la comunicación y amplia continuidad citoplasmática entre elementos cribosos de un mismo tubo criboso (Figura 10-42).

10.4.2.2. Translocación intercelular por los Plasmodesmas

Figura 10-41. Diagrama que muestra los elementos básicos en la circulación del agua, iones inorgánicos, y fotoasimilados en la planta.

El agua y los iones inorgánicos que absorbe la raíz se mueven hacia arriba por el xilema en la corriente de transpiración. Parte se mueve lateralmente hacia los tejidos de la raíz y del tallo, mientras que otra parte es transportada hacia zonas de la planta en crecimiento y hojas maduras. En las hojas, cantidades sustanciales de agua e iones inorgánicos son transferidos al floema y son exportados con sacarosa y la corriente de asimilación. Las partes de la planta en crecimiento, que son relativamente inefectivas capturando agua a través de la transpiración, reciben muchos de sus nutrientes y agua vía el floema. El agua y los solutos que entran en las raíces en el floema se pueden transferir al xilema y ser recirculados en la corriente de transpiración. En la figura 10-35 la letra A indica sitios especializados en la absorción y asimilación de materias primas del entorno. C y D designan sitios de carga y descarga, respectivamente, e I, puntos principales de intercambio entre el xilema y el floema. (La figura anterior y está modificadas de Raven, P.H. & Eichhorn, S.E. (1999). *Biology of Plants*. 6th ed., W.H. Freeman and Company. Worth Pub en.).

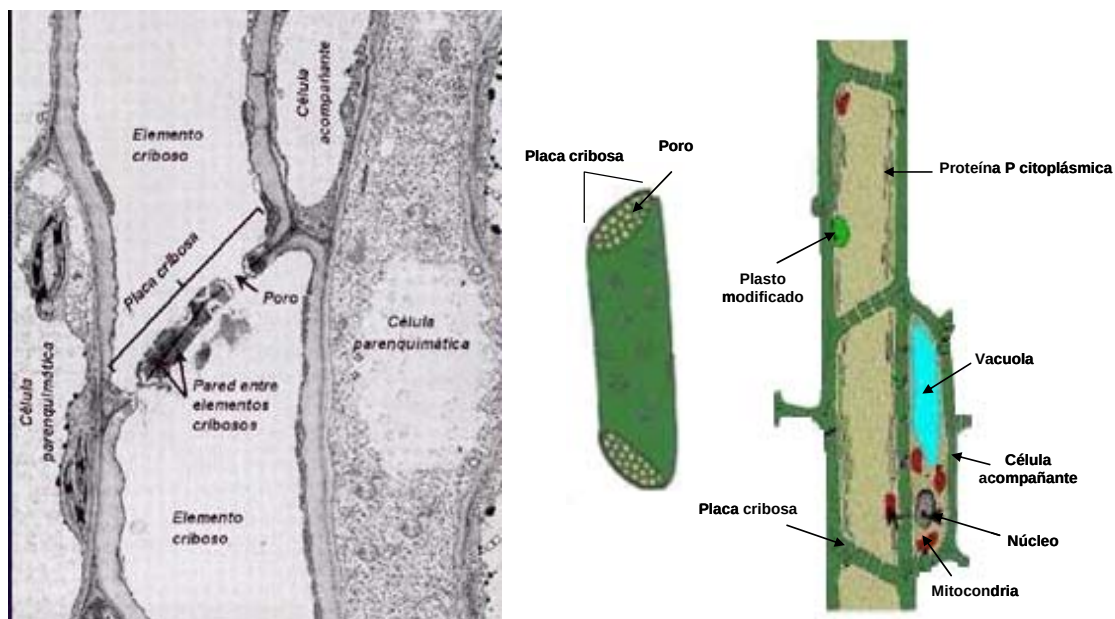


Figura 10-42. Estructura interna de los tubos cribosos. En la izquierda podemos observar un esquema de los elementos de los tubos cribosos. En la derecha se observa una micrografía electrónica de transmisión de una placa cribosa en sección longitudinal.

El floema es el tejido conductor especializado en la translocación de fotoasimilados. El movimiento de este contenido puede ser tanto ascendente como descendente y sus diferentes componentes pueden moverse en sentidos contrarios, aún dentro de un mismo haz conductor.

Del movimiento de los fotoasimilados, se dice que sigue un modelo de tejido fuente a tejido sumidero (consumidor). Las principales fuentes de solutos asimilables son las hojas fotosintetizantes, pero los tejidos de almacenamiento pueden servir también como importantes fuentes. Todas las partes de las plantas incapaces de satisfacer sus propias necesidades nutricionales, pueden actuar como sumideros, esto es, pueden importar productos asimilables. Así, los tejidos de almacenamiento actúan como sumideros cuando están importando productos asimilables y como fuentes cuando los exportan (**¡Error! No se encuentra el origen de la referencia.**).

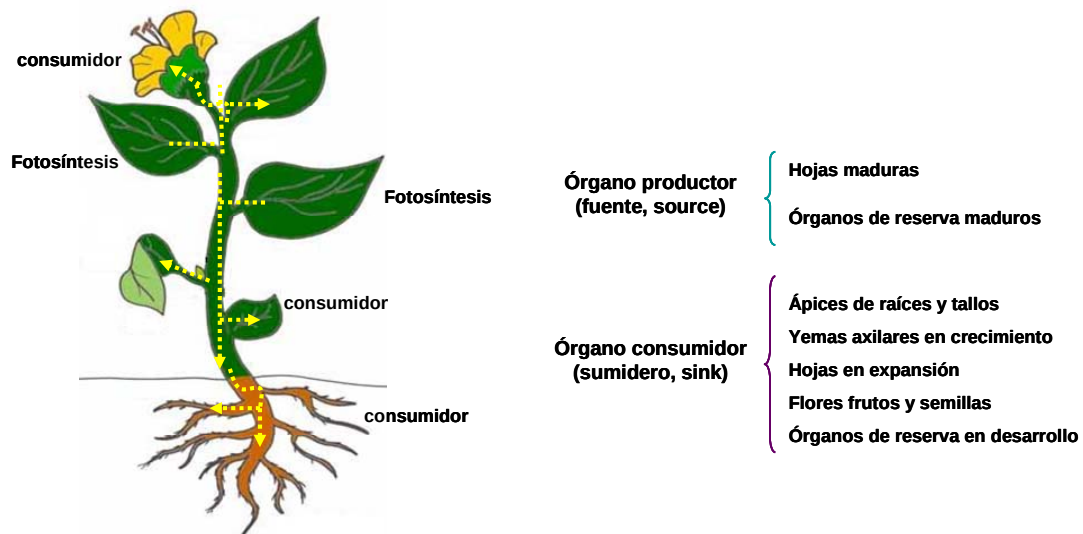


Figura 10-43. Esquema donde se muestra la diferencia de función entre los órganos fuente y los sumideros. También se indican los principales órganos de la planta que actúan como fuente o como sumidero.

Las primeras evidencias que apoyan el papel del floema en el transporte de productos asimilables llegaron por las observaciones de árboles a los que se les había quitado un anillo completo de corteza. La corteza de los tallos más viejos está compuesta principalmente de floema y no contiene xilema. Cuando a un árbol que está fotosintetizando se le quita una tira de corteza o se le hace una incisión circular alrededor de él, la corteza por encima de la manipulación se hincha, indicando la acumulación de productos que se

mueven hacia abajo por el floema desde las hojas fotosintetizadoras (Figura 10-44).

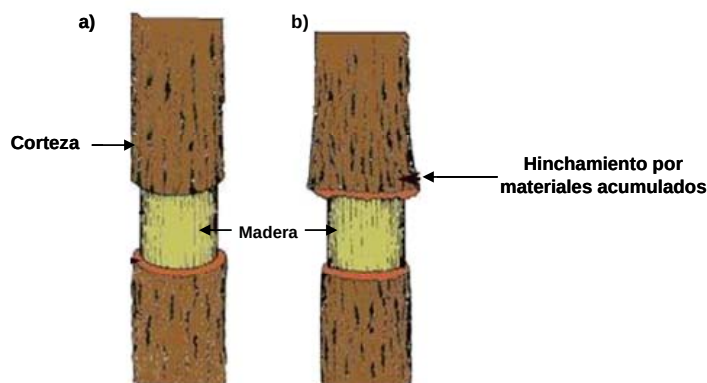


Figura 10-44. Cuando de un tallo se extrae un trozo de corteza en forma de anillo a), los tejidos por encima del corte se abultaban b). Este fenómeno se debe a un crecimiento nuevo de la madera y tejidos de la corteza estimulado por una acumulación de alimento que se desplaza hacia abajo desde las hojas y que queda interceptado en el anillo. Este fenómeno no se produce durante los meses de invierno. (Figura modificada de Raven, P.H. & Eichhorn, S.E. (1999). *Biology of Plants*. 6th ed., W.H. Freeman and Company. Worth Pub.).

La savia elaborada es una solución muy concentrada con un contenido de materia seca de 50 a 300 g/l. El 90 por 100 de la materia seca de la savia elaborada corresponde a azúcares, particularmente sacarosa. En otras plantas también se encuentran otros oligosacáridos, por ejemplo, rafinosa y estaquiosa, así como alditoles. Los monosacáridos (por ejemplo glucosa, fructosa) no se transportan. En cambio, la savia elaborada contiene también aminoácidos, amidas, nucleótidos, ácidos orgánicos e iones inorgánicos (aunque no Ca^{2+}). Estas sustancias, no obstante, aparecen en concentraciones mucho más pequeñas, comparadas con las de azúcares. Los pulgones y otros parásitos requieren los azúcares y otros compuestos acompañantes, como compuestos nitrogenados, en cantidades equilibradas. Es por eso cuando toman la alimentación descrita el exceso de azúcar lo exudan en forma de “melaza”.

A lo largo de los años se han propuesto diferentes mecanismos para explicar el transporte de productos asimilables en los tubos cribosos del floema. Probablemente el primero fue el de difusión, seguido del de corriente citoplasmática. La difusión y la corriente citoplasmática normales, del tipo que se encuentra en las células de las plantas superiores, fueron en gran parte abandonados como posibles mecanismos de translocación cuando se supo que las velocidades del transporte de asimilables (típicamente 50 a 100 centímetros por hora) eran demasiado

altas para que cualquiera de estos fenómenos justificara el transporte a grandes distancias vía los tubos cribosos.

Se han propuesto hipótesis alternativas para explicar el mecanismo de transporte en el floema, pero sólo una, la hipótesis de flujo de presión, justifica satisfactoria y prácticamente todos los datos obtenidos en estudios experimentales y estructurales del floema. Todas las otras hipótesis tienen serias deficiencias.

En la planta, la sacarosa producida por la fotosíntesis en una hoja es secretada activamente a los tubos cribosos menores. Este proceso activo, llamado carga de floema, disminuye el potencial hídrico en el tubo criboso y hace que el agua que está entrando a la hoja por la corriente de transpiración penetre en el tubo criboso por ósmosis. Con el movimiento de agua al tubo criboso de esta fuente, la sacarosa es transportada pasivamente por el agua a un sumidero, como una raíz de almacenamiento donde la sacarosa es extraída (descargada) del tubo criboso (descarga floemática). La extracción de sacarosa provoca un aumento del potencial hídrico en el tubo criboso del sumidero y el movimiento subsiguiente del agua fuera de él en ese lugar. La sacarosa puede ser utilizada o almacenada en el sumidero, pero la mayor parte del agua regresa al xilema y recircula en la corriente de transpiración.

Tarea 10-3.

Realice un mapa conceptual sobre lo aprendido. Compare el sistema vascular de las plantas con el sistema vascular de los animales. Indique las similitudes y las diferencias.

10.4.2.3. Transporte de proteínas y ácidos nucleicos

Tarea: consiga un artículo donde se haya reportado el transporte de RNA en el phloema.

11. Metodologías de mejoramiento

11.1. Metodologías clásicas

11.1.1. Introducción

El mejoramiento de plantas es un término utilizado para describir la creación, selección y establecimiento de fenotipos superiores de especies vegetales. El desarrollo de cultivos mejorados debe responder a las necesidades tanto de los productores como de los consumidores. Las principales metas del mejoramiento de plantas, incluyendo a los cultivos agrícolas y hortícolas, se han enfocado a mejorar el rendimiento, la calidad nutrimental y otras características de valor comercial. Dentro de la escala global, el paradigma del mejoramiento de plantas ha sido extraordinariamente exitoso, encontrando ejemplos importantes como la introducción de variedades de trigo (*Triticum aestivum*) y arroz (*Oryza sativa*) que generaron la revolución verde (Everson y Golin, 2003). También están los ejemplos sobre el desarrollo de híbridos de maíz (*Zea mays*; Duvick, 2001) y la reciente comercialización de cultivos transgénicos (James 2007). Estos y muchos otros productos que resultaron del mejoramiento de plantas, han generado múltiples beneficios para la sociedad en general, en términos de un mayor y sustentable suministro de biomasa, la cual es cosechada y aprovechada en forma de alimentos, vitaminas, medicinas y diversas materias primas como fibras, madera, papel, lubricantes y combustibles.

El mejoramiento de plantas tiene una larga historia integrando las últimas innovaciones en biología y genética para optimizar el desarrollo de cultivos optimizados. La selección prehistórica de fenotipos visibles que facilitaban la cosecha e incrementaban la productividad, llevaron a la domesticación de las primeras variedades de cultivos (Harlan, 1992). Esta domesticación puede ser considerada como uno de los primeros ejemplos de biotecnología, entendida en un sentido amplio, como el uso de herramientas tecnológicas para facilitar procesos biológicos.

11.1.2. Importancia del ambiente

La información genética determina el potencial de una especie. Pero eso no significa que el potencial que tienen los genes realmente se expresen a nivel de fenotipo. La interacción inherente de los genes, los transcritos, las proteínas y los metabolitos con el ambiente es la que finalmente determina el fenotipo observable de un individuo. La evaluación correcta y precisa de la variación fenotípica es el elemento crucial para una selección efectiva. No es suficiente con medir solo la variación genotípica, ya que esta no incluye los otros parámetros que son muy importantes. La variación

expresada depende tanto del potencial genético como de las condiciones del ambiente (factores bióticos y abióticos), así como de las variaciones circunstanciales del azar. La magnitud del efecto de cada uno de estos componentes puede variar de caso en caso. Las metodologías clásicas de mejoramiento genético parten precisamente de este punto. Las metodologías modernas deben tomar en cuenta esta complejidad. La conexión entre genotipo y fenotipo no es tan simple como la conexión entre una secuencia nucleotídica y su secuencia proteínica.

11.1.3. Importancia de los modos de reproducción

El conocimiento detallado del modo de reproducción de una especie es esencial para su mejoramiento. La forma de reproducción de un cultivo es un factor categórico para el desarrollo adecuado de esquemas de cruza, incluyendo los métodos de selección que se deben aplicar para el mejoramiento. Lo que se entiende es por ejemplo: la auto-fertilización de cultivos como en el trigo, implica formas de trabajo muy diferentes a los que se usan en una especie de polinización cruzada como el maíz, que además muestra un fuerte efecto de heterosis. Por lo regular se usan métodos de selección para un cultivo que toman en cuenta el estado genético en que van a ser usados en campo. Es difícil escoger lo que se quiere a nivel de híbrido, si solo se evalúa el fenotipo a nivel de línea. Por ejemplo, la autopolinización de cultivos de polinización cruzada conduce a la drástica reducción en su rendimiento. A esto se le llama **depresión endogámica**. Un investigador que no esté acostumbrado a trabajar con híbridos tal vez pueda subestimar estos detalles al inicio. Es por ello que muchas veces los investigadores deben aprender las particularidades de una especie antes de convertirse en buenos mejoradores. Un mejorador que ha trabajado con trigo tarda de 2 a 5 años en convertirse en un buen mejorador de maíz. Por otro lado, los mejoradores de arroz y de trigo pueden transferir sus prácticas más rápidamente debido a que estos dos cultivos son igualmente autógamos. Lo mismo sucede entre sorgo y maíz, que son ambos cultivos heterógamos.

Por otro lado, los métodos de mejoramiento de cultivos **sexuales** son totalmente diferentes de los cultivos de reproducción **asexual** (por ejemplo, cultivos clonales como el agave o el plátano). La propagación asexual es un tipo de reproducción en donde alguna planta o parte de ella pueden ser utilizadas para la multiplicación, sin siquiera implicar un cambio genético de generación en generación. Es por ello que deben tomarse muy en cuenta los parámetros sexuales y las condiciones genéticas en que se hacen las evaluaciones de campo y la selección de los genotipos.

La selección de vegetales se puede dividir en cuatro grandes categorías:

- la cría de líneas de cultivos autógamos (trigo, arroz, etc)
- la cría de poblaciones de cultivos alógamos (
- la cría de híbridos en cultivos alogamos que presentan una fuerte heterosis (maíz, sorgo, etc)
- la cría de clones de cultivos de propagación vegetativa (por ejemplo agave, henequén, plátano, papa, fresa, aguacate, etc).

Dentro de estas categorías, hay varios métodos de cultivos posibles. Además, existe un amplio espectro de técnicas y tecnologías para realizar pasos más fáciles de estos métodos, más rápidos, más precisos, o permitir a todos ellos, como: El cultivo de de microsporas, el rescate de embriones, mutagénesis, la transferencia de genes mediada por *Agrobacterium*, el análisis de QTLs, los marcadores moleculares, etc.

11.1.4. Autopolinización o polinización libre

Algunas de las razones por las cuales el método de auto fertilización es tan efectiva, es por la eficacia de la reproducción, así como la disminución de la variación genética. La mayoría de los loci se fija y se reduce así el número de variantes. Lo que se pretende es la fijación de alelos superiores correspondientes a genotipos con características deseadas (rendimiento, adaptabilidad, propiedades agronomicas e industriales, etc). Con cada generación de autopolinización la tasa de heterocigotos disminuye en un 50%, por lo tanto, una homocigosis de 96.9% se obtiene después de 5 autofecundaciones mientras que una homocigosis de >99.9% solo se obtiene después de 10 autofecundaciones.

En las plantas de polinización libre no se da la endogamia. Este modo de reproducción también permite la selección de las poblaciones silvestres heterogeneas. Los individuos fenotípicamente inferiores se eliminan poco a poco de la población durante los ciclos evolutivos. Es decir, en cada generación se van cambiando las frecuencias alélicas del pool genético.

Los pasos críticos en la mejora de los cultivos, son la elección de los padres y la identificación de las mejores variantes que se generan por cruzamientos, segregación y recombinación en cada generación.

Algunos métodos de selección que se usan son por ejemplo: la selección masal, la selección por pedigrí y la selección recurrente. Una forma comun de derivar líneas homocigoticas es la selección de familias, que se derivan de descendientes de sola semilla (SSD). También existen formas de derivar líneas homocigoticas de forma mas rápida como es el método de los dobles haploides (DH). En la mayoría de los programas de mejoramiento genético se aplica una combinación de algunos de estos métodos. En las siguientes secciones explicaremos algunas estrategias de selección de forma resumida.

11.1.5. Selección Masal

Este método de selección depende principalmente de la selección masiva de plantas dentro de una población, de acuerdo a criterios del fenotipo de plantas individuales. Se siembra una población en una parcela aislada, y se fomenta la polinización cruzada entre todas las plantas (recombinación por polinización libre). Por lo regular se usan densidades de plantación altas. Este tipo de selección se lleva a cabo sobre todo con cultivos semilla pequeña como la alfalfa, que generalmente son plantadas en altísimas densidades. También se hacen selección Masal para diversas poblaciones criollas de maíz.

Se seleccionan plantas en base a alguna característica de interés. Por ejemplo: aspecto de planta, tamaño de mazorca, forma de grano, color de semillas, etc. Las plantas, mazorcas y semillas seleccionadas se mezclan y se siembran a granel para la próxima generación. Este método se utiliza para mejorar la población en general por medio de un criterio o varios criterios fenotípicos o genotípicos. La selección puede ser positiva o negativa. Es un método sencillo y a la vez eficaz y rápido para la mejora de razas o poblaciones criollas.

Un inconveniente de la selección masal es la gran influencia que el ambiente tiene sobre el desarrollo del fenotipo y el rendimiento de las plantas individuales. A menudo no está claro si las plantas fenotípicamente superiores son también genotípicamente superiores. Las diferencias de ambiente pueden conducir a que la selección masal tenga muy baja eficacia, y por consiguiente, el progreso genético por unidad de tiempo sea menor al que se puede obtener con otros esquemas de mejoramiento.

En la selección en masa hay algunos factores que deben tenerse en cuenta al seleccionar las plantas. Es mejor seleccionar por bloques. El gradiente de campo puede ser un factor muy importante en el momento de la selección. La selección se debe hacer cuando la mayoría de las características de la planta se presenten claramente. En algunas plantas es durante la floración, mientras que en otras es durante la cosecha.

El método de selección masal sólo será efectivo para los rasgos altamente hereditarios. Es decir, solo funcionará para los rasgos fenotípicos que sean altamente visibles a nivel de planta o semilla individual. Esto solo se da para características con alta heredabilidad, por ejemplo, el color azul o amarillo de las semillas de maíz.

Sin embargo, este método es simple y barato y supone menos trabajo que la selección pedigrí en las generaciones primordiales. Es muy recomendable tener grandes poblaciones de plantas para garantizar que suficiente segregación para que los mejores combinaciones de alelos se presenten en algunas plantas individuales cuando empieza la selección. Durante los avances generacionales, un número menor de registros se mantienen a comparación que la selección por pedigrí.

La selección masal es el método más simple, más fácil y más antiguo de selección de plantas. Las plantas individuales son seleccionadas en base a su desempeño fenotípico. Se utilizan semillas a granel para producir la próxima generación. La selección en masa ha demostrado ser muy eficaz en el mejoramiento de maíz en las etapas iniciales. Pero su eficacia para la mejora del rendimiento fue objeto de severas críticas en los años 40 en Estados Unidos. Esto finalmente culminó en el perfeccionamiento de los métodos de selección para maíz. Esta selección no prevé ningún control sobre la matriz de polen durante la polinización. La fuente de polen no se conoce en una polinización cruzada. Como consecuencia, la selección se limita sólo a las madres. Las estimaciones de heredabilidad se reducen a la mitad, ya que sólo los padres se utilizan para la cosecha de semillas.

Sin embargo, los avances generacionales son de un solo ciclo, por lo que se pueden llevar mas rondas de selección a comparación con otros métodos de selección que requieren de 2 a 3 años para cada ciclo de selección.

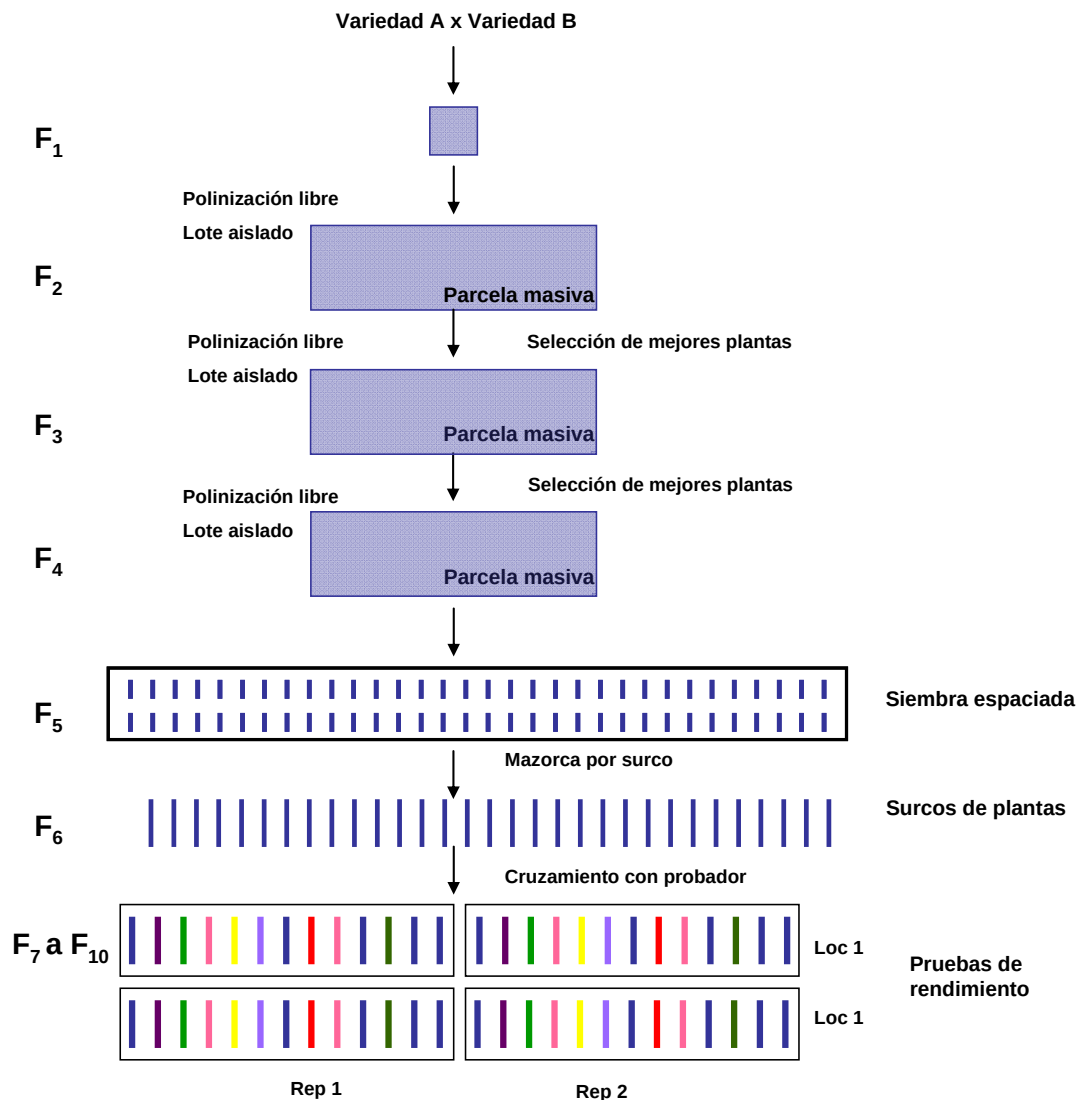


Figura 11-1. Método de selección de población masiva (Masal).

Método de selección Masal: La progenie de la cruce se cultiva en forma masiva hasta la generación F5. Se seleccionan plantas o cabezas de línea y se siembran en surcos de plantas o cabezas de línea en la F6. Se seleccionan los surcos superiores y se cultivan en un ensayo preliminar de rendimiento. Las cepas superiores se hacen crecer en pruebas de rendimiento de la F8 a la F10. Pueden hacerse varias modificaciones a este procedimiento, por ejemplo la selección podría comenzar en la F3 o F4, purificando las líneas que tengan un rendimiento superior en las siguientes generaciones; o bien, las parcelas masivas podrían repetirse y cosecharse para determinar el rendimiento y desechar cruces completas con base en el rendimiento de dichas parcelas.

Selección periódica

Este tipo de selección es una versión refinada de la masa y difiere de la siguiente manera:

Visualmente los individuos son seleccionados de poblaciones sometidas a pruebas de progenie. Las plantas seleccionadas sobre la base a los datos de la prueba de progenie se cruzan unos con otras en todas las formas posibles para producir semillas. Esto forma la base para la nueva población.

Generación de líneas por medio de semillas únicas (SSD)

Este método se introdujo como un medio para transformar toda la generación F2 tan rápido como sea posible en una generación de plantas homocigotas. Esto a pesar de que la población inicialmente sea muy heterogénea. Este método puede utilizarse para reducir el tiempo que se requiere para eliminar los genotipos que aún no son homocigotos. Debido a que con sólo se desea obtener una semilla por planta (no toda una mazorca), entonces las plantas pueden ser sembradas a muy alta densidad. También se puede intentar tener dos o tres ciclos por año. Esto puede hacer que el proceso para la liberación de un cultivar, se disminuya de 10 a 5 años. El mantenimiento de registros no es necesario a principios de cada generación.

La desventaja es que este método no elimina las plantas más débiles como en otros métodos, y no existe ninguna disposición para la selección de plantas superiores en la generación F2. Algunos investigadores han hecho modificaciones de este método para remediar esos problemas.

Una variedad desarrollada por este método será más uniforme que los elaborados por la selección masal, porque todas las plantas de tal variedad tendrán el mismo genotipo. Las semillas de plantas seleccionadas no se agregan. Si no que están separados y son utilizados para realizar pruebas de progenie. Esto se hace para estudiar el comportamiento reproductivo de las plantas seleccionadas.

Selección de hermanos completos

Las plantas de la población original se autofecundan para producir progenies S1, que son evaluados en el próximo ciclo. Esto se puede hacer con ensayos de múltiples repeticiones y en diferentes localidades para identificar las familias S1 prometedoras. El remanente de semillas S1 de estas familias seleccionadas se recombina en la tercera temporada. Como consecuencia, un ciclo se puede completar en tres ciclos. Las unidades de selección y recombinación son las progenies S1.

Selección de medios hermanos

La selección se hace en base a pruebas de progenie de rendimiento en lugar de la apariencia fenotípica de las plantas de sus padres. Las semillas seleccionadas de medios hermanos, que han sido polinizadas al azar por el polen de la población, se cultivan en filas con el propósito de selección. Una parte de la semilla se siembra para determinar la capacidad de rendimiento, de cualquier carácter de cada planta. La semilla de las familias más productivas se cosecha a granel para completar un ciclo de selección.

Método dobles haploides

Se pueden producir plantas haploides por la eliminación de cromosomas. Se pueden obtener a por medio del cultivo de microesporas. Sin embargo, las cruza con inductores de haploidía son mas utilizados hoy en día por su capacidad para producir plantas haploides con mucha mayor cantidad en comparación con los otros métodos.

Destaca que son necesarias para alterar el desarrollo de vías de microsporas de producir polen a la formación de plantas haploides. Actualmente, el llamado método inductor es muy utilizado en el mejoramiento del maíz (el maíz híbrido)

El número de cromosomas haploides de las plantas se duplica con el tratamiento de colchicina. Espontáneamente las plantas haploides se duplica, sin embargo, también puede ser producido directamente a partir de los tres métodos.

Los métodos de rescate de embriones pueden utilizarse para garantizar que las semillas de estos cruces no se aborten. Este método tiene el potencial de acortar los ciclos de mejoramiento genético en comparación con los métodos del pedigrí o de distribución masiva. Al igual que el único método de selección de semillas, a principios de generaciones no están sujetos a la selección, pero la mayoría de las líneas se eliminan durante la evaluación de campo los ensayos.

El objetivo es el mismo que para SSD: transformar el material genético de la F1 tan rápido como sea posible en un homocigoto en la mayor parte de las plantas. Este método en mano de obra es muy intenso y el más caro de los procedimientos porque aumenta la cantidad de generaciones por año.

Por este método para tener éxito, los vegetales deben ser genéticamente estables.

Retrocruzas:

Es un método de selección para el mejoramiento de genotipos. Este es un tipo de crusa repetida con el mismo padre recurrente. Se puede generar así una línea isogénica.

Un alelo puede ser incorporado en el fondo genético de una línea elite. El llamado padre recurrente (a la variedad) es altamente productiva y comercialmente de éxito, pero carece de este gen específico (por ejemplo, la resistencia a las enfermedades). Este rasgo es, por ejemplo, puede estar presente en una variedad silvestre o exótica (donante). Después de la primera crusa, se repite la crusa con el padre recurrente. Después de cada crusa las plantas híbridas con la combinación de genes deseados deben ser identificadas y seleccionadas. Después se cruzan de nuevo con el padre recurrente.

Esta técnica es fácil cuando los rasgos que se añaden son fácilmente heredados. Es decir, es adecuada para introgresar alelos dominantes y fácilmente identificables en las plantas híbridas. Si hay genes no deseados que están estrechamente emparentados con el gen deseado, los genes no deseados son transferido junto con el gen deseado y los descendientes pueden sufrir de un desperfecto. Una de las ventajas del método de la parte de atrás, es que las pruebas (para que no sea el rasgo deseado) no es necesaria. Este método se utiliza para, por ejemplo, transferencia de la esterilidad masculina en líneas puras, que es un requisito previo para muchos programas de mejoramiento híbrido. Mientras el Marcador de retrocruzas asistida es habitualmente aplicada en programas de mejoramiento de la introgresión de genes.

Métodos de selección tradicionales

Se ha generado una cantidad de variedades muy diversas con los métodos de selección tradicionales. Hay un cúmulo de oportunidades para la recombinación de genes, para piramizar alelos en cultivares superiores. Sin embargo, hay pocas oportunidades para la rotura de bloques de genes ligados y haplotipos. Los mejoradores están buscando superar estas deficiencias a través del aumento de la cantidad de cruas y recombinaciones realizadas.

También se esta tratado de iniciar las pruebas de rendimiento en las generaciones anteriores e introducir sistemas modificados para la mejora de poblaciones. Por ejemplo, la selección recurrente de poblaciones pertenecientes a grupos heteróticos recíprocos (*reciprocal recurrent selection* RRS).

Las líneas que básicamente son genéticamente idénticas, a excepción de un único gen, se llaman líneas isogénicas.

La cría híbrida es una categoría muy importante para fertilizar los cultivos y tiene las siguientes ventajas en comparación con la alternativa natural que es la población de cría:

- Más heterosis
- Máxima intensidad posible de selección (mejor genotipo posible de la población)
- Más fácil para crear variedades con múltiples genes de resistencia (cruza entre dos líneas complementarias)
- Método atractivo para las empresas privadas a causa de la depresión endogámica, que constituye un mecanismo natural para la protección de las variedades.

El incremento de semillas híbridas ha sido uno de los principales objetivos de la horticultura y prácticas agrícolas, porque son las plantas híbridas más productivas (debido al vigor híbrido) y más uniformes que las plantas procedentes de la polinización al azar. Para aumentar semillas híbridas, de polinización libre y polinización de medios hermanos (polinización por una planta de la misma línea) de las semillas madre debe ser rehuido.

Un método de castración de la línea femenina utilizada, que luego, naturalmente, es de polinización cruzada por el polen de la línea que actúa como padre. Esta debe plantarse en una fila adyacente. Sin embargo, este proceso es muy intensivo de mano de obra y caro.

Si las plantas se pueden hacer auto-incompatibles con la introducción de genes que controlan la auto-incompatibilidad, esto facilitaría la producción de semilla híbrida, ya que reduce los costos de mano de obra. Las semillas para los híbridos se generan por medio de la polinización cruzada entre dos diferentes líneas. Sin embargo, el paso necesario de los padres en la fertilización de formas homocigóticas por cuenta propia no es fácil con el material auto-incompatible. Lo más conveniente es utilizar un sistema de esterilidad de polen citoplásmico-génico.

Selección de polinización cruzada de cultivos

El estado natural para los cultivos de auto-fertilización es el homocigotico. Especies de plantas que el modo normal, las característica reproductiva y las características de la estructura de la población de las semillas es a través de establecer un alto grado de polinización cruzada. Cada planta recibe una mezcla de polen de un gran número de individuos que tienen diferentes genotipos. Estas poblaciones se caracterizan por un alto grado de heterocigosidad. Esto permite un libre flujo de genes entre los individuos

dentro de las poblaciones, generando así un enorme potencial de la variación genética, que se mantiene.

En el desarrollo de variedades híbridas, el objetivo es identificar a los heterocigotos más productivos de la población, que entonces se producen con la exclusión de los demás miembros de la población. En contraste, en el mejoramiento poblacional se prevé una mejora por medio de la eliminación gradual de los alelos nocivos y menos productivos. Esto se logra por ciclos repetidos de recombinación y selección. Los programas de mejoramiento de poblaciones son lentos, pero deben de ser constantes y con un compromiso a largo plazo. Por otro lado, la producción de híbridos tiene por objeto aprovechar al máximo los beneficios de la heterosis en mucho menos tiempo. Los dos enfoques son complementarios y no mutuamente excluyentes.

Selecciones con prueba cruzada de rendimiento

El propósito de este tipo de selección es una ligera desviación del concepto de población. La intra-mejora en el sentido de que la población se mejora, no sólo para el rendimiento, sino también con respecto a la capacidad de combinación con una población de referencia. Se trata del mejor aprovechamiento del vigor híbrido. Consiste de tres pasos:

1. Auto-polinización y la prueba de cruce de individuos
2. Evaluación de los cruces en los ensayos replicados
3. Recombinación cruzada remanente de las semillas de plantas seleccionadas

Métodos de selección para el desarrollo de cultivos de cría pura en la cruce entre variedades.

La introducción de germoplasma y líneas de cría se realizan para crear nuevas combinaciones de genes. En las generaciones siguientes, los genotipos óptimos (presumiblemente superior que tengan los genes y combinaciones de genes) son seleccionados y fijados en los estados homocigotos por medio de fertilización y selección. Son ampliamente probadas estas selecciones, con el objetivo de la liberación de algunos predominantes para el cultivo.

Cultivos de propagación asexual (selección de clones)

La reproducción asexual cubre aquellos modos de multiplicación de plantas, en los que la formación de gametos y fertilización no se lleva a cabo. Es decir, no hay meiosis ni fecundación. A falta de la recombinación genética que se da durante la meiosis, la composición genética de material vegetal que se multiplica sigue siendo esencialmente la misma. La

composición de alelos es exactamente la de la planta madre. A esto se le llama clon. Los clones superiores son seleccionados y propagados vegetativamente. La descendencia permite la liberación de variedades estables sin ningún tipo de deterioro debido a la segregación. Esta característica singular de la reproducción asexual ha ayudado a desarrollar una serie de cultivos de frutas y hortalizas, incluidas uvas, fresas, aguacates, platanos, mangos, manzanas, etc.

La uniformidad tan alta de los clones es algo destacado. En cuanto se encuentra una planta superior, esta se puede reproducir ilimitadamente de manera vegetal. Sin embargo existe el riesgo de patógenos altamente especializados. La variabilidad genética es necesaria para amortiguar los cultivos contra agentes patógenos, así como la estabilidad de la producción bajo condiciones ambientales variables.

Mutagénesis

El mejoramiento de plantas a través de mutaciones inducidas tiene claras ventajas y limitaciones. Cualquier clon vegetativo pueden ser tratados con mutágenos, e incluso con un único mutante deseable o una parte de un propágulo mutado puede ser multiplicado como un tipo mejorado de la variedad original.

Desarrollo de nuevos clones

El desarrollo y registro de nuevos clones debe tener lugar por medio de la selección local del clon en plantaciones antiguas, así como la importación de clones de alta calidad desde el extranjero, para la evaluación local.

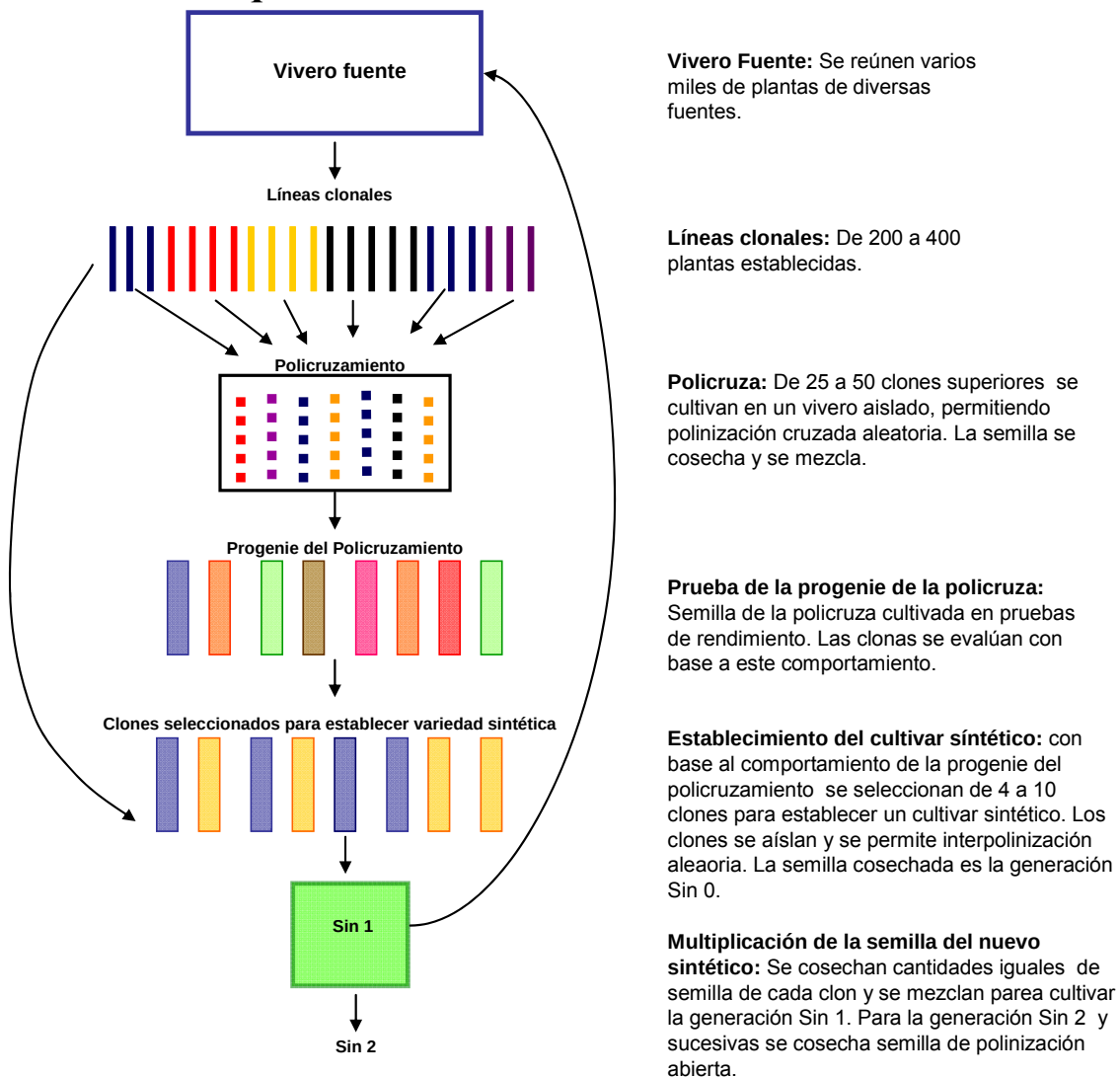
Un clon es la descendencia vegetativa de una planta madre específicas, pero no muestran ninguna desviación genética, morfológica o fisiológica de la planta madre. La evaluación se lleva a cabo con los diferentes clones seleccionados después de la selección. Los diferentes clones se comparan entre sí para determinar su calidad y capacidad de resistencia. La Cría no participa en la selección del clon, el clon no puede ser criado para la resistencia de ciertos tipos de virus, el énfasis debe ponerse en asegurarse de que el clon al salir de la guardería es libre de virus. Se han desarrollado técnicas para poner a prueba los clones para cualquier virus nocivo. Los virus nocivos a veces no muestran en las evaluaciones del anteproyecto de desarrollo fitosanitario (detección de virus y erradicación de virus) es, pues, realizado en los laboratorios e invernaderos, en paralelo con el terreno y la evaluación de la calidad en el campo del clon.

Los desarrolladores tuvieron que incorporar técnicas, como los cultivos de tejidos y propagación *in vitro* de material limitado para desarrollar clones libres de virus de la madre. El meristemo apical está libre de cualquier virus nocivo. Al utilizar el meristemo apical para el cultivo de tejidos pueden ser desarrollados un clon libre de virus.

Multiplicación de clones

Durante la primera fase de multiplicación, cada uno de los clones candidatos se mantiene en un almacén. A partir de ese tejido núcleo se harán todas las futuras multiplicaciones y evaluaciones. Durante la segunda fase de multiplicación, los portainjertos (el patrón) y el injerto se deberán mantener en lugares libres de insectos dentro de instalaciones de campo en zonas aisladas. El vástago y material patrón de la segunda fase de la fuente son injertadas y encalladas. Las plantas injertadas se plantaron en zonas aisladas para el establecimiento de bloques de la madre. Estos bloques se utilizan para fines de multiplicación. La tercera fase es la creación de bloques descendiente de la madre de las explotaciones de los productores contratados de la citada fuente en el pre-selección, se mantienen. Unos 4 km² de suelo virgen

11.1.6. Esquemas de selección



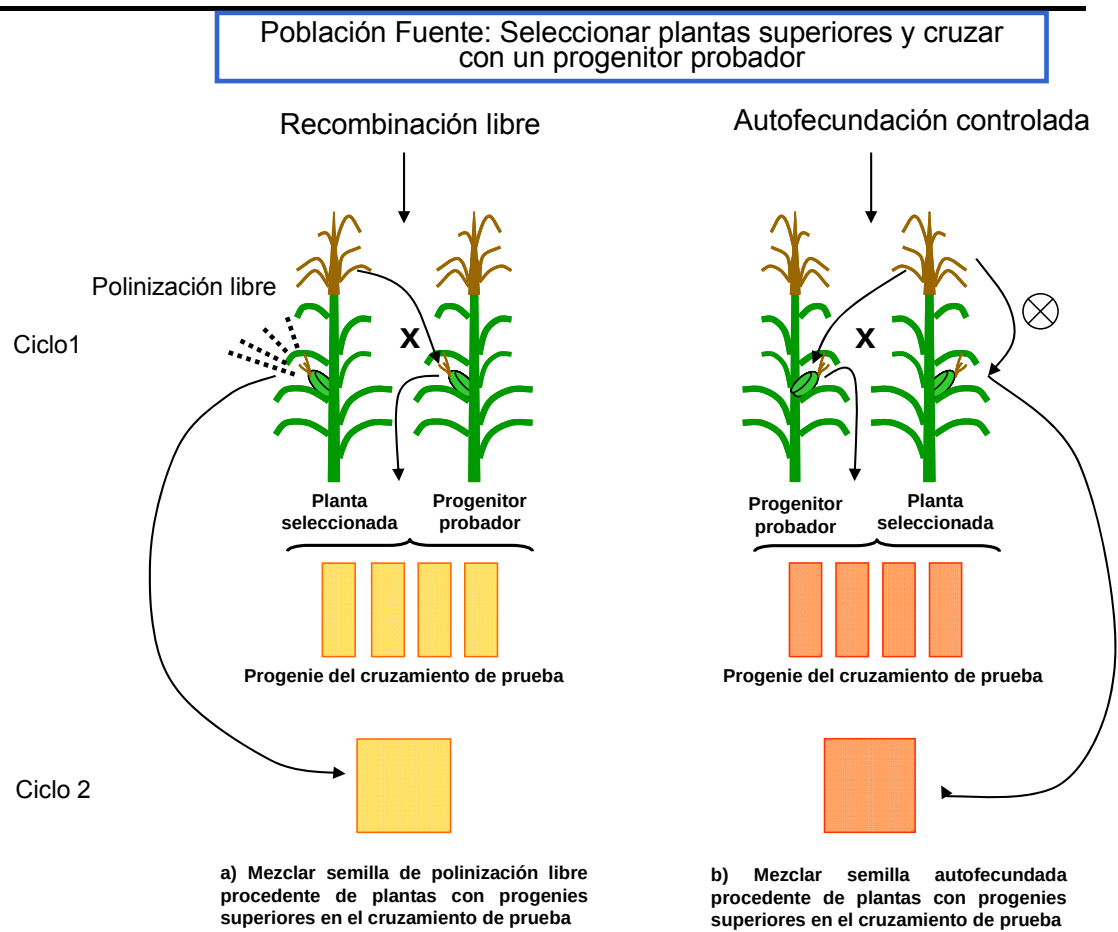


Figura 11-2. Selección de familias de medios hermanos basada en el rendimiento de la progenie del cruzamiento de prueba

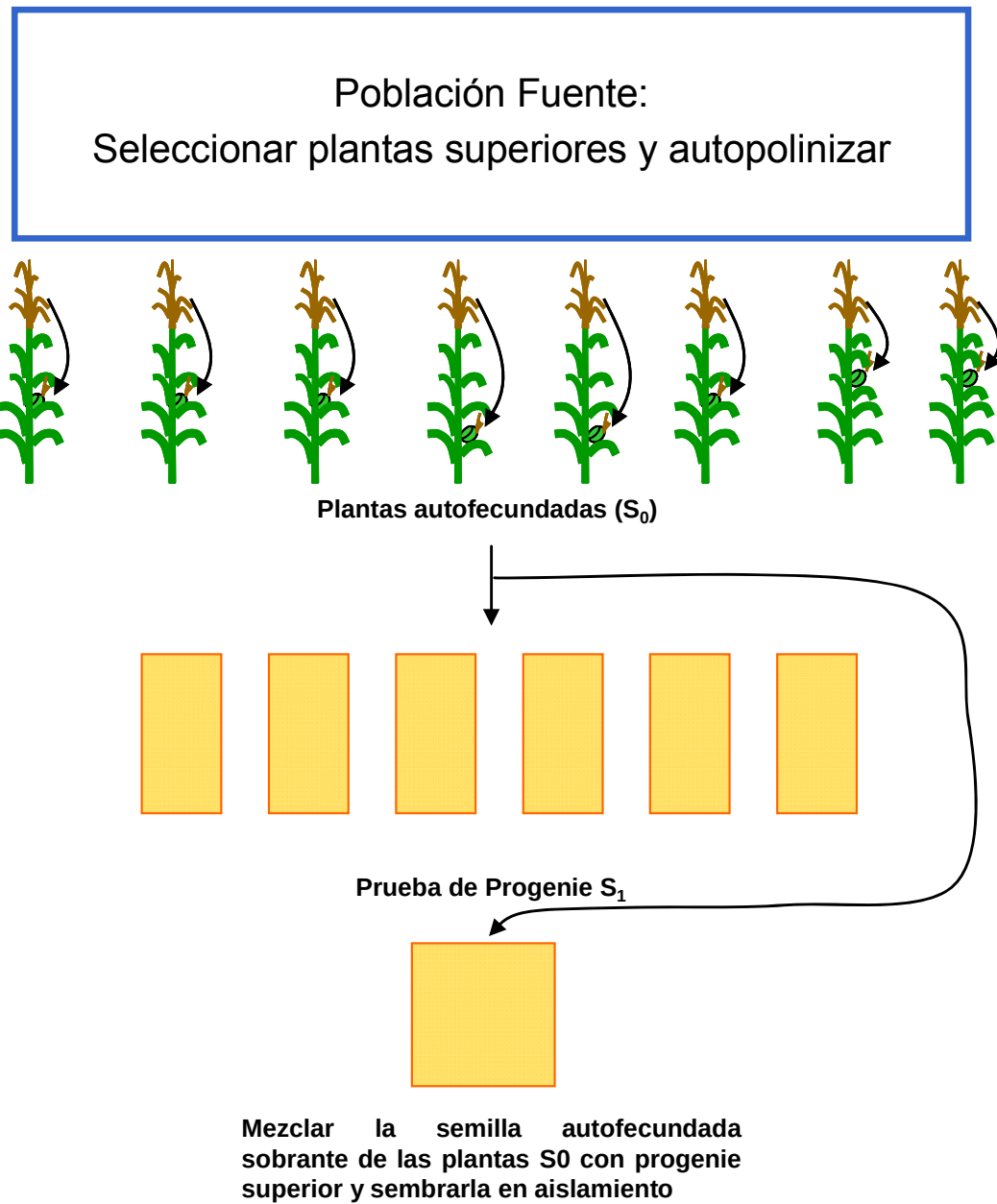


Figura 11-3. Selección basada en el rendimiento de la progenie S_1 .

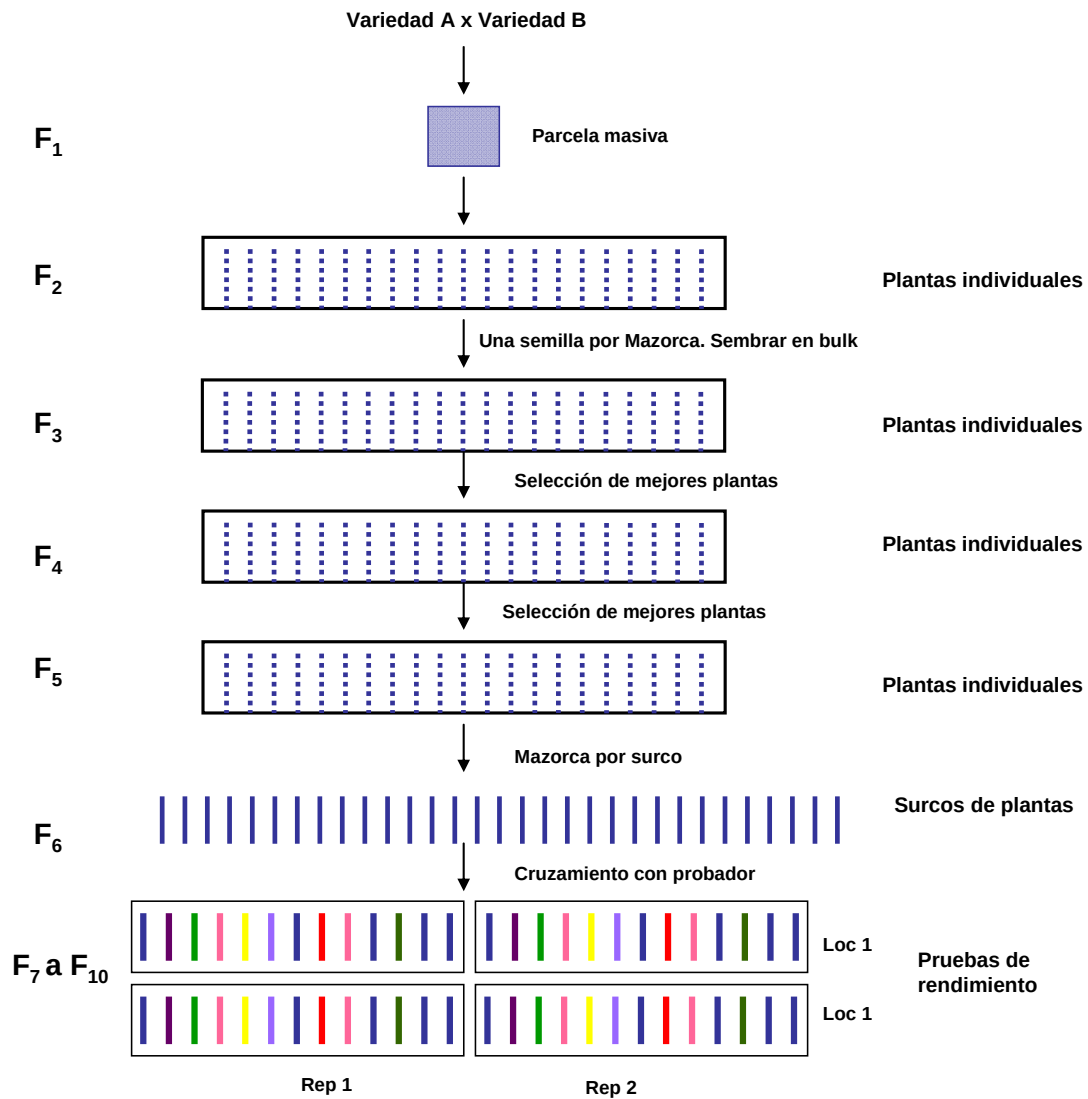


Figura 11-4. Método de selección de descendencia uniseminal. Las semillas cosechadas de la F_1 se siembran de manera espaciada en la F_2 . Una sola semilla cosechada de cada planta F_2 se utiliza para obtener la generación F_3 . Esto se repite hasta la generación F_5 . En esta generación las plantas se cosechan y se siembra en un surco de progenie en la F_6 . En la generación F_7 se establece un ensayo preliminar de rendimiento. Este ensayo se repite hasta la F_{10} . Algunos fitomejoradores combinan este método, con el método de selección por pedigree. Cultivan sólo las generaciones F_3 y F_4 utilizando el primero de estos métodos, acortando el tiempo que se requiere para hacer la prueba de rendimiento.

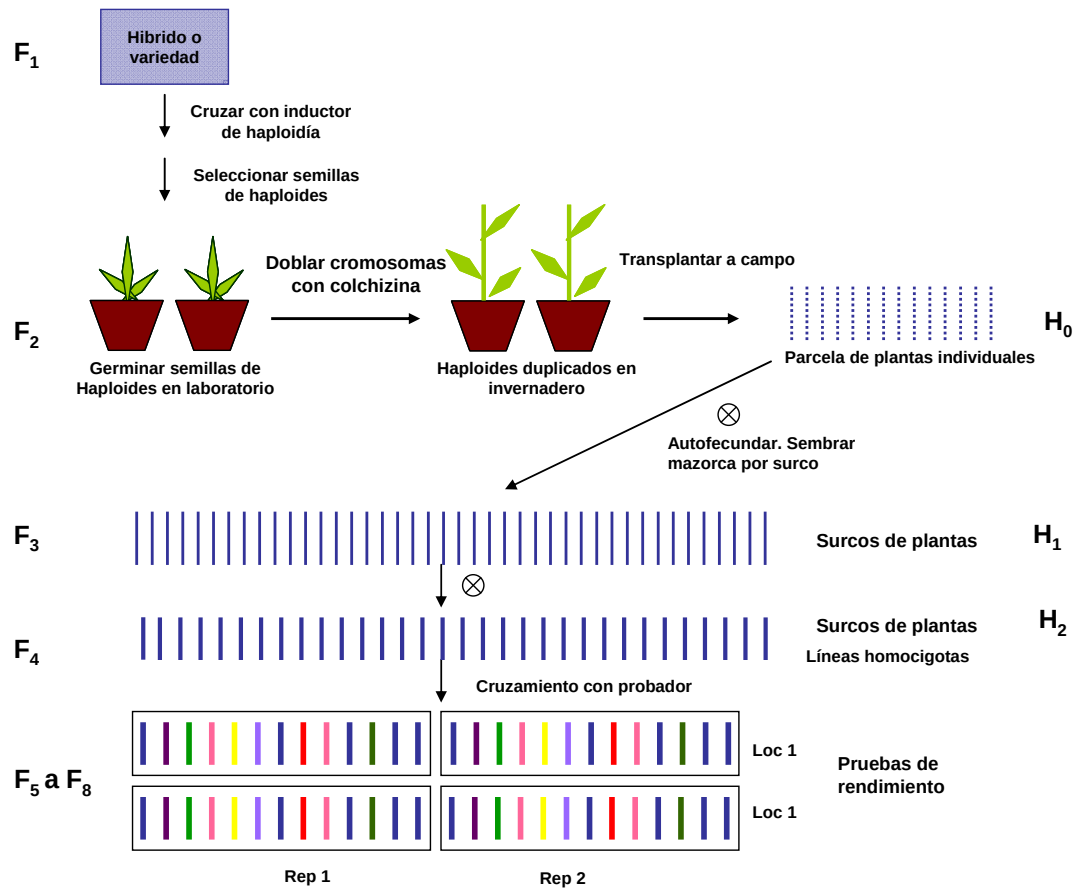


Figura 11-5. Método de haploides duplicados para la generación de líneas homocigotas.

Para el método de dobles haploides, se parte de una población elite o de un híbrido de alto rendimiento (F_1). Las plantas se cruzan con un inductor de haploidía. Las mazorcas se desgranán en bulk. Se seleccionan las semillas haploides (aleurona azul con embrión blanco). Se germinan las semillas haploides y se tratan con colchizina (inhibidor de mitosis) para generar plantas con cromosomas duplicados ($F_2=H_0$). Las plantas se transfieren a invernadero o directamente a campo en una parcela de plantas individuales. Ahí se autofecundan todas las plantas ($H_0 \rightarrow H_1$). Se cosechan y se siembran mazorca por surco. Se avanza una generación mas seleccionando las mejores líneas ($H_1 \rightarrow H_2$). Se hace cruzar con un probador y se evalúan los híbridos de las líneas haploides en ensayos de rendimiento en las generaciones F_5 a F_8 .

11.1.7. Mejoramiento por pedigrí

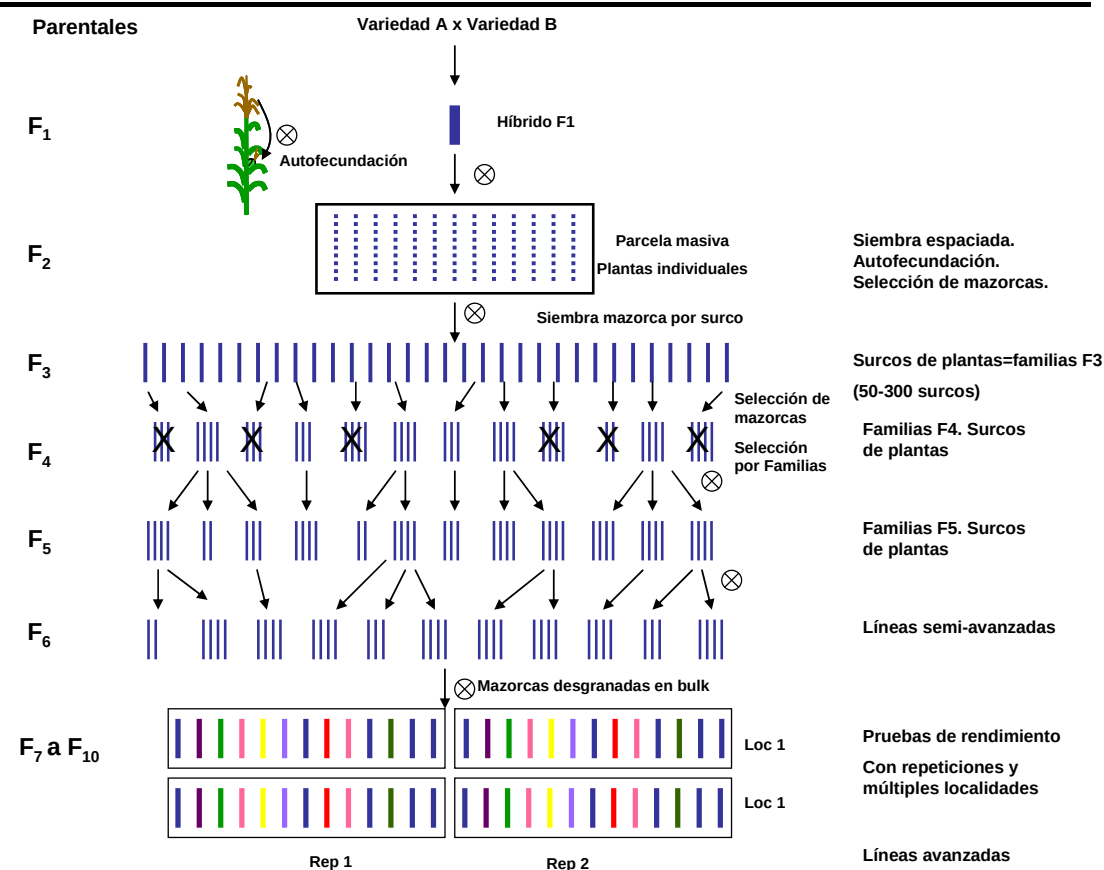


Figura 11-6. Diagrama que muestra la metodología de mejoramiento por pedigree.

Método de selección por pedigree: A partir de plantas F₂ seleccionadas, se hacen crecer en surcos progenies de 25-30 plantas en la F₃. De los mejores surcos se seleccionan las plantas superiores y las mazorcas mas bonitas se desgranar mazorca por sobre y se siembran en familias de surcos. La selección consiste en escoger las mejores plantas de las mejores familias. Al llegar a F₆, las familias empiezan a ser relativamente mas uniformes. En la F₇ se establecen pruebas preliminares de rendimiento mismas que se continúan hasta la F₁₀.

El método de pedigrí es generalmente rápido y fácil. Sin embargo, este método es muy intenso en mano de obra. Los padres se cruzan y se genera una F₁, que después se autofecunda para generar una población F₂, que después deriva en familias de líneas F₃, F₄, etc. La selección de plantas con nuevas combinaciones de genes se hace en base a la genealogía. La descendencia se somete repetidamente a la selección y siempre las plantas individuales se toman como fuente de nuevas familias descendientes (mazorca por sobre, mazorca por surco), hasta que se alcanza la uniformidad genética. Sólo entonces, las semillas de varias mazorcas se combinan y mezclan para tener más semillas para pruebas de extensas de combinaciones híbridas. Es importante llevar un registro detallado del origen de las plantas o las líneas seleccionadas. La intensidad de selección

puede ser variada, en la práctica, de acuerdo a la disponibilidad de infraestructura (espacio de campo, almacenaje de semilla, instalaciones, personal para desgrane, selección y preparación de ensayos, etc.).

Con una sola planta de selección se debe llevar a cabo en base a rasgos de alta heredabilidad. En la primera fase se deben concentrar los alelos mayoritarios y favorables. En una fase posterior, cuando las unidades de evaluación son parcelas y ya no plantas individuales, se puede hacer selección para algunos rasgos con menor heredabilidad. Una objeción en contra de este método es que la variación genética disponible para la selección de los rasgos cuantitativos, se puede reducir drásticamente en las primeras generaciones. Es por ello que es importante tener una población de familias lo mas grande posible desde el inicio de la generación F3. Por lo regular se deben empezar con 100 a 300 líneas S2. La purificación de semillas y la multiplicación suele ser incorporado en una de las últimas generaciones de selección de pedigrí.

La eficiencia de la cría es uno de los objetivos para la generación de pruebas. Esto se hace por la identificación temprana de familias superiores. La eliminación rápida de las poblaciones inferiores y la posterior concentración de los esfuerzos de selección en las poblaciones superiores se traducen en un aumento de la eficiencia. Una evaluación precisa de genotipos es esencial para el éxito de este método.

11.1.8. Selección recurrente

Tarea: Haga un diagrama de cómo funciona la selección recurrente

11.1.9. Metas de mejoramiento

11.1.9.1. Rendimiento

La agricultura se centra en el consumidor, para innovar y mejorar la calidad y cantidad de los suministros alimenticios en el mundo globalizado. Este es un punto de partida muy importante, si tenemos en cuenta que hasta ahora la biotecnología agrícola, a través de la genética, se había centrado en la lucha contra los insectos y la desratización

Tarea: Escriba un ensayo de 2 páginas sobre la siguiente cuestion: ¿Qué significa rendimiento para un agricultor?

11.1.9.2. Calidad nutricional

Entre las tendencias más importantes de la nueva agricultura, nos encontramos con la formación de nuevos productos y servicios, influidos por una demanda de alimentos de mejor calidad. Pero también, nos

encontramos con una expansión biotecnológica cada vez más rápida, con una mayor tecnología informática, con nuevas empresas relacionadas con el sector y con una expansión global en potencia.

Gracias a la biotecnología, la mejora de los alimentos, a través de la incorporación de nuevas propiedades nutricionales potencialmente beneficiosas, reducirá los riesgos vinculados con el padecimiento de algunas afecciones, como la osteoporosis o los procesos cardíacos.

Desde la mejora genética del maíz, se han desarrollado variedades especializadas tanto para calidad proteica (ricos en lisina y triptófano) como para alto contenido de aceite (de 6 a 9% de grasa cruda), las cuales están en proceso de adopción por los agricultores y usuarios; en México, los maíces de alta calidad proteica (conocidos como QPM, por su denominación en lengua inglesa: *quality protein maize*) se han venido promoviendo durante los últimos seis años con éxito limitado todavía; los maíces de alto contenido en aceite (conocidos como HOC: *high oil corn*) no ingresan al mercado nacional de semillas por su origen transgénico, así como las variantes en el manejo especializado para su producción.

La condición gemelar o poliembriónía en semillas de maíz es una característica natural que puede ser aprovechable como una vía alterna en el diseño de variedades de aplicación especial, buscando además de potencial de rendimiento, el valor nutritivo del grano, incrementando cantidad y calidad de aceites y proteína; esto, bajo la hipótesis de que dos o más embriones por semilla, permitirán incrementar la capacidad de almacenamiento de nutrientes de calidad (Castro, 1973; Rodríguez, 1981; Espinoza et al, 1998; Espinoza *et al.*, 1999). El planteamiento de que la condición gemelar o poliembriónica de las poblaciones PE (frecuencias poliendromicas) del IMM (instituto Mexicano del maíz) debe estar asociado a un mayor y mejor contenido de nutrientes en la semilla, principalmente en cuanto a aceites y proteína embrionaria, se ha venido documentando, inicialmente por la vía de análisis bromatológicos de las semillas (Espinoza *et al.*, 1999), y recientemente por la cuantificación de ciertos nutrientes específicos, como el caso de los ácidos grasos (Valdés *et al.*, 2004) y, planeado para realizarse en fechas del año 2004, la determinación de los aminoácidos lisina y triptofano en harina de granos completos. La hipótesis en esta línea de trabajo es que dos o más embriones por semilla, deben conferirle a ésta una mejor constitución nutritiva y cierta ventaja selectiva.

La selección aplicada para incrementar la poliembriónía en las poblaciones de maíz bajo estudio influye de manera positiva en los contenidos nutrimentales de las semillas, e.g. aumentando el nivel de grasa cruda por encima del promedio conocido en el maíz común y la proporción de ácidos grasos insaturados, mejorando la relación oleico:linoleico.

La cantidad de proteína cruda en maíces PE no presenta variaciones significativas a los testigos, sean maíz común o especializado del tipo

QPM; sin embargo, será importante corroborar, en estudios a realizar otros estudios, cual es el nivel de lisina y triptofano, estableciendo la hipótesis de que los maíces PE han acumulado mayores contenidos en estos aminoácidos esenciales por el hecho de presentar un mayor número de embriones por semilla.

Ingeniería Metabólica

La ingeniería metabólica ofrece un medio para mejorar la calidad y cantidad de flavonoides presentes en los alimentos que ingerimos. Siendo otro ejemplo de mejoras en el potencial nutricional de los alimentos

En tomate, que produce naturalmente sólo pequeñas cantidades de kaempferol y quercetina en la piel del fruto, se ha conseguido introducir y sobre-expresar los genes reguladores *Lc* y *C1* de maíz, lo que lleva a un incremento de la formación de kaempferol de más del 60%, principalmente en la pulpa de los frutos. Además, la introducción del gen CHI de *Petunia* aumenta la biosíntesis de quercetina en más del 70% en la piel. La expresión de *Lc* y *C1* en patata causa una marcada acumulación de kaempferol en los tubérculos. En el futuro, la disponibilidad del gen FLS, codificante de la enzima flavonol sintasa, obtenido de varias fuentes permitirá una ingeniería mucho más directa en la síntesis de los flavonoles.

11.1.9.3. Calidad industrial

En un futuro no muy lejano los científicos agrícolas podrán desarrollar polímeros industriales o productos fibrosos procedentes de fuentes renovables como el maíz o la soja.(Parker) tanto dentro como fuera del sector agrícola. Desde el siglo XX, el interés se ha incrementado. De un tiempo a esta parte, la cuestión más importante ha sido cómo perfeccionar el empleo de los excedentes por parte de los agricultores de los países industrializados. Una segunda cuestión importante ha sido la necesidad de encontrar sustitutos a los productos disponibles en ese momento, sustitutos que nos evitaban depender de importaciones foráneas.

Los biocombustibles son productos alternativos fabricados a partir de la "biomasa", que es materia orgánica, como madera, plantas o desechos orgánicos, que puede ser convertida directamente en líquidos combustibles para propósitos de transporte. Estos biocombustibles están siendo investigados y desarrollados como reemplazo de los combustibles derivados del petróleo y como sustitutos parciales de los aditivos, para de esta manera "estirar" las reservas de petróleo existentes y reducir la dependencia general de los combustibles fósiles.

Fabricación del etanol

El etanol o etil alcohol es un combustible de alto octanaje, producido a partir de la fermentación de las plantas de azúcar. En Estados Unidos, el E-10 (10% etanol / 90% gasolina) es la combinación alcohol/gasolina más ampliamente disponible para la compraventa minorista del transporte, pero también se producen mezclas de combustibles para automóviles superiores a E-85. (La operación con E-85 requiere de "vehículos de combustible flexible", fabricados especialmente (FFV), de los cuales actualmente hay en uso cinco millones). Aunque el etanol se está usando como oxigenante para el diesel tradicional, la mayoría de las investigaciones y del mercadeo se centran en su uso en mezclas de gasolina. El maíz es la principal reserva alimenticia para la fabricación de etanol en Estados Unidos. El etanol también se produce a partir de fuentes orgánicas tales como cebada, trigo, arroz, soya, girasol, papa, mandioca y melaza. Fuera de Norteamérica, la caña de azúcar y la remolacha de azúcar son las reservas alimenticias más comunes. También puede producirse a partir de hierbas silvestres, paja de trigo y otros elementos orgánicos, comúnmente considerados como basuras, tales como paja de arroz, desechos de maderamen y hojas y tallos de plantas (afrecho). El afrecho de maíz es el desecho agrícola más importante en toda América.

De acuerdo con la National Corn Growers Association, cerca del 13% del maíz cultivado en Estados Unidos (unos 1,43 mil millones de bultos) se transformaron en etanol en 2005. Un bulto de maíz produce cerca de 2,8 galones de etanol.

La molienda húmeda y la molienda seca son dos mecanismos de producción para generar etanol a partir del maíz. En la molienda seca, el grano entero primero es convertido en harina de maíz. El polvo con agua se transforma en una suspensión a la que se le agregan enzimas para convertir la fécula en dextrosa. Se agrega amoníaco para control del pH y como nutriente para la levadura. La masa se procesa en una cocción a alta temperatura para reducir los niveles de bacterias. Luego se enfría la masa y se agrega la levadura.

El proceso de fermentación toma por lo general entre 40 y 50 horas. Luego se separa el etanol y se concentra a 190 grados, usando destilación convencional; después se deshidrata a etanol anhidro de 200 grados. Enseguida se mezcla con un desnaturalizador, como la gasolina u otros destilados del petróleo.

En la molienda húmeda, el grano es embebido en agua y ácido sulfuroso diluido por más de 48 horas, para facilitar la separación del grano en sus partes componentes. Después de remojar, el compuesto acuoso de maíz es molido para separar el germen del maíz. Luego la fécula y cualquier

remanente acuoso de la masa se fermentan en un proceso similar al del método seco.

También hay mejoras en el potencial farmacéutico de los alimentos. Los isoflavonoides pueden actuar como fitoestrógenos, lo cual ha generado un gran interés en el uso de estos compuestos para tratar desórdenes hormonales en humanos. Los isoflavonoides sólo se presentan en las leguminosas, en las cuales el primer paso de la biosíntesis de las mismas está regulado por la isoflavona sintasa (IFS). Recientemente se ha clonado el gen que codifica para dicha enzima, lo cual ha abierto un nuevo campo de ingeniería para la formación de isoflavonoides en plantas de cultivo que normalmente no presentaban dichos compuestos. Se ha realizado un experimento inicial en *Arabidopsis thaliana* introduciendo el gen IFS de soja por el promotor 35S, resultando la conversión de la naringenina en el isoflavonoide genisteina, que pertenece al grupo de los fitoestrógenos de alto interés médico.

En maíz, las mutaciones recesivas de dos genes CHS: C2 y Whp dan como resultado polen estéril de color blanco, producto de la ausencia de flavonoides en él. En la *Petunia*, la supresión de sentido de CHS causa flores blancas y esterilidad a la vez. La expresión antisentido de CHS en las anteras causa también esterilidad masculina. Por su parte en las plantas de tabaco, la introducción y sobre-expresión del gen STS que codifica para la estibeleno sintasa, que compite con la CHS endógena por sustratos comunes, y por tanto también se produce esterilidad masculina y pigmentación alterada de las flores (Forkmann, G., S. Martens. 2001). Estudios adicionales muestran que los fenotipos estériles pueden ser complementados por la adición de flavonoles.

Supresión de la fertilidad del polen La esterilidad masculina es un requisito para el desarrollo de sistemas de semillas híbridos, por tanto, la generación de este hecho por ingeniería metabólica es una meta muy importante.

En consecuencia, la estrategia del antisentido o la supresión de sentido de FLS bajo el control de un promotor específico es una alternativa a la supresión completa de la síntesis de los flavonoides. Parece ser que esta es la estrategia más prometedora para la generación de esterilidad masculina. Biosíntesis de flavonoides por bacterias genéticamente modificadas. Recientemente mediante la ingeniería genética se ha logrado cultivar bacterias capaces de sintetizar flavonoides de tipo flavanonas. (Hwang E. I., M. Kaneko, Y. Ohnishi, S. Horinouchi. 2003)

11.1.9.4. Tolerancia a la sequía

Con el propósito de mejorar la tolerancia del maíz y el trigo a condiciones de escasez de agua, se adoptaron varias estrategias, entre ellas, generar por medio de ingeniería genética variedades que contienen diversas construcciones genéticas que podrían mejorar el desempeño de las variedades en condiciones desfavorables. Si bien existen ciertos asuntos que deben atenderse antes de que las variedades transgénicas lleguen a los agricultores (p. ej. propiedad intelectual, bioseguridad, conservación del ambiente y seguridad de los alimentos humanos y animales), si se demuestra que los sistemas genéticos basados en transgenes son efectivos, ofrecerán la posibilidad de mejorar el comportamiento de las plantas en condiciones desfavorables. De particular importancia es la naturaleza (unigénica y dominante) del transgene que hace más fácil la transferencia y el mantenimiento de los sistemas en cualquier variedad, en comparación con la de aquellos que se basan en genes múltiples y genética cuantitativa.

Los mecanismos moleculares de la respuesta a la escasez de agua se han estudiado principalmente en la planta modelo *Arabidopsis thaliana*. Los análisis de la expresión de genes inducidos por deshidratación han demostrado que por lo menos cuatro vías de señalización metabólica independientes funcionan en la inducción de genes inducidos por factores desfavorables, en respuesta a la deshidratación: dos son dependientes de ABA y dos son independientes. Varios de los genes inducidos por condiciones desfavorables, como rd29A en *A. thaliana*, son activados mediante una vía metabólica independiente de ABA. El gene 1 de Ligamiento al Elemento de Respuesta a la Deshidratación (DREB1 por sus siglas en inglés) y DREB2 son factores de transcripción que se unen al promotor de genes como rd29A, por lo que inducen la expresión en respuesta a la sequía, la sal y el frío.

DREB1A ha sido sobreexpresado en plantas transgénicas *Arabidopsis* y el fenotipo resultante ha mostrado una sólida inducción de la expresión de los genes objetivo en condiciones favorables, pero también se ha observado enanismo en los fenotipos de las plantas transgénicas. Las plantas también mostraron tolerancia al congelamiento y la deshidratación. Por el contrario, la sobreexpresión de DREB2A produjo una expresión débil de los genes objetivo en condiciones favorables y el crecimiento de las plantas transgénicas fue lento. La regulación de la expresión del gene DREB1A por mediación del promotor rd29A en condiciones desfavorables produjo plantas con mayor tolerancia al congelamiento, la sal y la sequía y sin cambios drásticos en el fenotipo normal de las plantas transformadas.³

³ Como parte de las actividades del CIMMYT para aumentar la tolerancia a la sequía en el trigo, JIRCAS, por conducto del Dr. Shinozaki, proporcionó

Avances: Se han generado varios cientos de eventos y se han probado en condiciones de escasez de agua en el invernadero de bioseguridad. Se ha confirmado que los eventos que han mostrado los más altos niveles de tolerancia a las condiciones de escasez de agua contienen el gene DREB y se están evaluando en condiciones de campo en el invernadero del CIMMYT en El Batán, México. Si se observa que el comportamiento de los transgénicos mejora en las condiciones que se están evaluando, se realizarán más ensayos de campo.

La sequía reduce la fijación de nitrógeno en leguminosas. En condiciones de sequía, se produce una limitación de carbono en los nódulos de las leguminosas que podría ser la causa del descenso de la fijación de nitrógeno en las mismas. Ésta es una de las conclusiones que ha presentado María Dolores Gálvez en su tesis defendida en la Universidad Pública de Navarra. El trabajo doctoral se titula: Metabolismo nodular en *Pisum sativum* L. en respuesta a estrés hídrico. Interacciones carbono/nitrógeno y posibles moléculas implicadas en la modulación de la respuesta.

Fijación de nitrógeno y sequía. La fijación biológica de nitrógeno es un proceso de gran interés agronómico y ecológico, puesto que el nitrógeno, después del agua y el carbono, es el nutriente que limita en mayor medida el crecimiento vegetal y la producción de los cultivos. Este proceso resulta particularmente sensible a condiciones ambientales adversas, como el estrés hídrico o sequía. Precisamente el objetivo de la tesis doctoral de María Dolores Gálvez ha sido investigar cómo se produce la regulación de la fijación biológica de nitrógeno en condiciones de sequía.

La reducción del nitrógeno atmosférico a amonio, o fijación de nitrógeno, la llevan a cabo únicamente determinados organismos procariotas. Entre ellos, los denominados genéricamente rizobios son capaces de establecer simbiosis con plantas leguminosas dando lugar a la formación de una nueva estructura radical: el nódulo.

Por un lado, la planta se beneficia del microorganismo, que se encarga de tomar el nitrógeno del aire y transformarlo en amonio de forma que se lo facilita a la planta. Este amonio se incorpora a esqueletos de carbono para formar aminoácidos y proteínas. Por su parte, el microorganismo obtiene de la planta los nutrientes necesarios para su crecimiento.

En condiciones de sequía, se detectó un descenso de la actividad sacarosa sintasa nodular. Este descenso se produjo simultáneamente al descenso de la fijación de nitrógeno, pudiéndose establecer una elevada correlación

el gene DREB1A de *A. thaliana* controlado por un promotor inducido por sequía, para probarlo primero en trigo y posteriormente en maíz

entre ambos procesos en condiciones adversas. Como consecuencia de la inhibición de la actividad sacarosa sintasa, se observó también un descenso de la concentración de azúcares fosfato y ácidos orgánicos, lo que indica un descenso en el flujo de carbono en los nódulos que limitaría, a su vez, el suministro de carbono al bacterioide, viéndose afectada la capacidad del bacterioide para fijar nitrógeno.

Con el fin de profundizar en el proceso de percepción y transducción del estrés hídrico que lleva a la inhibición de la fijación de nitrógeno, María Dolores Gálvez estudió también el ácido abscísico y las especies de oxígeno activado como posibles moléculas implicadas en la regulación de la fijación de nitrógeno.

11.1.9.5. Resistencia a enfermedades

El proceso por el cual las plantas desarrollan una enfermedad se puede dividir en cinco fases.

1. **Inoculación:** Es la introducción del agente patógeno dentro del tejido de la planta huésped. Por ejemplo las esporas son los principales agentes transmisibles de los hongos. El agua o el viento pueden transportar este agente, pero también los insectos, pájaros, mamíferos.
2. **Incubación:** Es un período de desarrollo durante el cual el agente patógeno experimenta una serie de cambios hasta llegar a una forma que le permite penetrar e infectar. Los hongos forman la estructura denominada anclaje de penetración.
3. **Penetración:** Es el proceso por el cual el agente se introduce dentro de la planta, puede ser un proceso activo como pasivo. Algunos agentes patógenos elaboran enzimas que disuelven las capas superficiales y la celulosa, o introducirse por los estomas, lenticelas o heridas. Algunas estructuras comunes son los apresorios.
4. **Infección:** Cuando el agente invade el tejido de una planta y se establece una relación parasitaria. También se generan nuevas esporas y la enfermedad se dispersa de forma exponencial.
5. **Enfermedad:** Cuando en la planta huésped se manifiesta una respuesta a la presencia del agente, como clorosis o necrosis. Son los síntomas del patógeno que muchas veces reducen considerablemente el rendimiento.

El mejoramiento para la resistencia a una enfermedad puede estar enfocado a cualquiera de esas etapas. Se pueden seleccionar genes que dificultan la penetración, genes que bloquean la infección sistémica, o bien genes que disminuyen la severidad de los síntomas.

11.2. Metodologías innovadoras

11.2.1. Retrospectiva histórica del mejoramiento molecular de plantas

Evolución del mejoramiento: de Darwin a Mendel hasta Hallauer, Watson y Crick

Es cierto que Darwin delineó los principios científicos de la hibridación y la selección. Mientras fue Mendel quien definió la asociación fundamental entre el genotipo y el fenotipo. Por medio de sus teorías fue posible hacer mejoramiento con un enfoque científico a inicios del siglo 20 (e.g. Shull, 1909). A pesar de que algunos mejoradores reconocieron inmediatamente la importancia de la genética mendeliana (genética cualitativa), la integración total se retrazo alrededor de 20 años, hasta que la genética cuantitativa reconcilió los principios de Mendel con la continua variación observada en la mayoría de los rasgos considerados importantes por los mejoradores (Paul and Kimmelman, 1988). Los principios de genética cuantitativa descritos en el famoso libro de Hallauer son primordiales para entender este proceso. Los subsecuentes avances en nuestro entendimiento de la biología de plantas, análisis e inducción de la variación genética, genética celular y cuantitativa, biología molecular, biotecnología y más recientemente de genómica, se han aplicado exitosamente para incrementar la base científica y su aplicación en el proceso de mejoramiento de plantas (e.g. Baenziger et al., 2006; Jauhar, 2006; Varshney et al., 2006).

La era de la biotecnología de plantas comienza a inicios de los años ochenta, con los conocidos reportes de producción de plantas transgénicas utilizando *Agrobacterium* (Bevan et al., 1983; Fraley et al., 1983; Herrera-Estrella et al., 1983). A partir de entonces se desarrollaron sistemas de marcadores moleculares para crear mapas genéticos de alta resolución y así explotar la relación entre marcadores y características importantes de los cultivos (Edwards et al., 1987; Paterson et al., 1988).

Para 1996 la comercialización de cultivos transgénicos demostró la exitosa integración de la biotecnología en el mejoramiento de plantas y en los programas de mejora de cultivos (Kozziel et al., 1993; Delannay et al., 1995). Como se observa en la Figura 11-8, la introgresión de uno, o algunos genes en un cultivo, comúnmente utilizado retrocruzas, es una práctica común de mejoramiento. Algunos métodos de retrocruza asistida por marcadores fueron desarrollados rápidamente para hacer introgresión de características transgénicas y reducir el arrastre por ligamiento (*linkage drag*), en donde los marcadores moleculares se utilizaron para explorar el

genoma y seleccionar aquellos individuos que contuvieran el transgen y la proporción más alta de alelos favorables del genoma parental recurrente (e.g. Ragot et al., 1995; Johnson and Mumm, 1996).

Durante los últimos 25 años, el desarrollo y aplicación continua de la biotecnología de plantas, marcadores moleculares y genómica, han establecido nuevas herramientas para la creación, análisis y manipulación de la variación genética, así como el desarrollo de cultivos mejorados. Ver reviews (Sharma et al., 2002; Varshney et al., 2006; Collard and Mackill, 2008). El mejoramiento molecular es actualmente una práctica estándar en varios cultivos, sobre todo en las empresas privadas de semillas. En las siguientes secciones revisaremos brevemente el impacto positivo tanto de la información molecular como de la ingeniería genética, en el paradigma del mejoramiento de plantas.

11.2.1.1. Principios y prácticas del mejoramiento

Esquemas de mejoramiento y concepto de ganancia genética

Conceptualmente, el mejoramiento de plantas es simple: cruzar los mejores padres e identificar y recupera la progenie que supera a los padres. En la práctica, el mejoramiento consta de tres pasos, en el cual, poblaciones o colecciones de germoplasma con variación genética útil son creadas o acopladas, se identifican los individuos con fenotipos superiores y se desarrollan cultivos mejorados a partir de los individuos seleccionados. Se ha desarrollado una amplia gama de acercamientos para hacer mejoramiento de plantas, dependiendo de la especie a cultivar y los objetivos del mejoramiento (Fehr, 1987; Stoskopf et al., 1993). Estos métodos de mejoramiento difieren en los tipos de población, procedimientos de selección y el producto final.

En la figura x se muestra un resumen de los métodos de mejoramiento que comúnmente se emplean en programas de desarrollo de cultivos. Como ya se mencionó, cuando la meta es mejorar un genotipo elite establecido, con características controladas por uno o pocos loci, se utiliza la retrocruza para lograr la introgrección de un solo gen (Fig. 1A), o una pirámide de pocos genes (Fig. 1B). Para características genéticas complejas, el desarrollo de germoplasma mejorado requiere la reorganización del genoma para producir nuevas combinaciones de genes favorables en la progenie. El método de mejoramiento por pedigrí genera estas nuevas combinaciones mediante la cruce y recombinación entre padres superiores y complementarios, así como mediante la selección entre la progenie segregante para rendimiento mejorado (Fig. 1C).

La selección periódica pretende incrementar de manera simultánea las frecuencias de alelos favorables en loci múltiples de la población mejorada, a través de la cruce entre individuos seleccionados (Fig. 1D). Para cultivos híbridos como el maíz, la selección periódica podría extenderse para mejorar el rendimiento de poblaciones complementarias distintas (e.g. grupos heterocigotos) que se utilizan como padres para generar combinaciones híbridas superiores. A esta práctica se le conoce como selección periódica recíproca.

Los principios de genética cuantitativa han sido particularmente poderosos como la base teórica tanto para el desarrollo de poblaciones, como para los métodos de selección y estabilización de genotipos deseables (Hallauer, 2007). Un concepto importante en la genética cuantitativa y el mejoramiento de plantas es la ganancia genética (DG), el cual se refiere al cambio predicho del valor promedio de una característica dentro de la población donde ocurre la selección. Sin tener en cuenta la especie, la característica de interés o el método de mejoramiento utilizado, el valor DG sirve como una expresión universal simple para la mejora genética esperada (Fehr, 1987; Falconer and Mackay, 1996). La figura 2 muestra la ecuación de DG y una expansión de sus términos para parámetros fundamentales de la genética cuantitativa. Aunque el mejoramiento de plantas utiliza claramente una sobresimplificación de los principios de la genética cuantitativa, la ecuación de ganancia genética relaciona los cuatro factores centrales que influyen el proceso de mejoramiento: el grado de variación fenotípica presente en la población (representado por su SD, s_P), la probabilidad de que una característica fenotípica se transmita de los padres a la progenie (heredabilidad, h^2) la proporción de la población seleccionada como padres de la siguiente generación (intensidad de selección, i , expresada como unidades de SD del promedio) y el tiempo necesario para completar un ciclo de selección (L). L , no sólo es función del número de generaciones que se requieren para completar un ciclo de selección, sino también, de que tan rápido se complete una generación y el número de generaciones por año.

Es claro que DG se puede aumentar incrementando s_P , h^2 o i , y disminuyendo L . Por lo tanto, la ecuación de ganancia genética provee un marco de comparación de la efectividad de estrategias particulares de mejoramiento, y comúnmente se utiliza como una guía para asignar juiciosamente los recursos necesarios para alcanzar los objetivos de mejoramiento. Cuando se considera dentro del contexto de ganancia genética, el mejoramiento molecular de plantas ofrece nuevos y poderosos acercamientos que permiten superar las limitaciones previas para maximizar DG. En las siguientes secciones se citan ejemplos en donde el mejoramiento molecular de plantas ejerce un impacto positivo en DG y

cada una de las variables que la integran. Por cuestión de brevedad, nos enfocaremos en ejemplos de maíz, en donde el mejoramiento molecular está más avanzado y se ha convertido en uno de los primeros cultivos en el que se han desarrollado híbridos comerciales mejorados.

El mejoramiento molecular de plantas expande la diversidad genética útil para la mejora de cultivos.

El máximo potencial para la ganancia genética es proporcional a la variación fenotípica (sP) presente en la población fuente original, y se mantiene en los ciclos subsiguientes de selección. La variación fenotípica está asociada positivamente a la diversidad genética, aunque también depende de factores ambientales y de las interacciones entre genotipo y ambiente. La diversidad genética se puede derivar de poblaciones cultivadas (naturales o sintéticas), progenie segregante de la cruce de líneas parentales seleccionadas, materiales exóticos que no están adaptados al medio ambiente, cruces interespecíficas entre especies silvestres, mutaciones naturales o inducidas, introducción de eventos transgénicos, o combinaciones de estos.

Sin embargo, no todas las variaciones fenotípicas son iguales. Por ejemplo, el uso de germoplasma exótico ha sido extremadamente exitoso para mejora varias especies de cultivo, pero puede haber dificultades si se introducen alelos no deseados, relacionados con pérdida de adaptación. La necesidad de diversidad genética debe ser balanceada mediante desempeño elite, puesto que es clave escoger los mejores padres, para maximizar la probabilidad de una mejora exitosa. En contraste, el incremento esperado en el desequilibrio por ligamento entre las poblaciones elite, derivado de intensas selecciones previas, también puede limitar la creación de nuevas combinaciones genéticas para ganancia futura. Con el entrecruzamiento de especies fuente, para lograr la recombinación genética se puede superar este problema, pero esto retrasa el desarrollo del cultivo.

Los marcadores moleculares y más recientemente, los esfuerzos de secuenciación del genoma con altos rendimientos, han incrementado dramáticamente el conocimiento y la habilidad de caracterizar la diversidad genética en el fondo común de germoplasma, para esencialmente cualquier especie de cultivo. Usando maíz como ejemplo, el reconocimiento de alelos de marcadores moleculares y la variación en la secuencia nucleotídica han proporcionado información básica sobre la diversidad genética antes y después de la domesticación de su antecesor silvestre, el teosinte, entre razas domesticadas distribuidas geográficamente y dentro de germoplasmas elite históricos (para una revisión, leer Cooper et al., 2004; Niebur et al., 2004; Buckler et al., 2006).

Esta información enriquecer las investigaciones sobre evolución de plantas y genómica comparativa, contribuye a entender la estructura de la población, proporciona medidas empíricas de las respuestas genéticas hacia la selección y sirve también para identificar y mantener reservas de variabilidad genética para una futura mina de alelos benéficos (McCouch, 2004; Slade et al., 2005). Además, el conocimiento de las relaciones genéticas entre recursos de germoplasma puede ser una guía al escoger padres para la producción de híbridos o poblaciones mejoradas (e.g. Dudley et al., 1992; Collard and Mackill, 2008).

Mientras que los marcadores moleculares y otras aplicaciones genómicas han sido altamente exitosas caracterizando la variación genética existente, la biotecnología de plantas genera nueva diversidad genética, que generalmente va más allá de especies relacionadas (Gepts, 2002; Johnson and McCuddin, 2008). La biotecnología permite acceso a genes que no estaban disponibles al utilizar cruces y crea un fondo de variación genética esencialmente infinito. Los genes pueden ser adquiridos de los genomas existentes en todos los reinos de vida, o pueden ser diseñados y ensamblados en el laboratorio. Ejemplos tanto sutiles como extremos, sobre el poder de los transgenes para introducir variación fenotípica, se pueden encontrar en tres diferentes transgenes que desarrollan resistencia en maíz y otros cultivos, contra los herbicidas de glifosfato. En el primer híbrido de maíz tolerante al glifosfato, se utilizó una versión modificada del gen endógeno de maíz que codifica para la 5-enol-piruvilshikimato-3-fosfato sintasa (Spencer et al., 2000). A este híbrido le siguieron una serie de eventos producidos por el gen de la 5-enol-piruvilshikimato-3-fosfato sintetasa aislado de *Agrobacterium* (Behr et al., 2004). Más recientemente un gen sintético con la actividad de glifosfato acetil transferasa aumentada, fue creado utilizando reordenamiento de genes (gene shuffling) y selección en sistemas microbianos (Castle et al., 2004). Cada uno de estos casos de maíz tolerante al glifosfato, ilustra también otro beneficio de la biotecnología, donde nuevas combinaciones de secuencias regulatorias (por ejemplo los promotores del virus del mosaico del la coliflor, y el de actina 1 de arroz) pueden ser usadas para lograr la expresión óptima de una característica, con respecto a la actividad general y la distribución relativa de tejido, lo que podría ser imposible con los genes endógenos (Heck et al., 2005).

El mejoramiento molecular de plantas incrementa la acción favorable de los genes

La genética cuantitativa utiliza el concepto teórico de heredabilidad para cuantificar la proporción de la variación fenotípica que es controlada por el genotipo. En la práctica, la heredabilidad está considerablemente influenciada por la arquitectura genética de la característica de interés, la cual esta descrita por el número de genes, la magnitud de sus efectos y el

tipo de acción del gen sobre el fenotipo. Un mejor conocimiento de la arquitectura genética y de la acción benéfica de la acción de los genes (la cual es más preferible para la selección) tiene un mayor impacto sobre la optimización de la ganancia genética. Para la fórmula de la ganancia genética, la heredabilidad (h^2) se utiliza en un sentido estricto, representando la proporción de la variación fenotípica debida a los efectos genéticos (aquellos que reflejan sustituciones de alelos o cambios en su proporción).

Los efectos genéticos aditivos también son considerados como valor de cultivo (valor de cría), puesto que su transferencia a la progenie es predecible. Las desviaciones de los efectos aditivos son significativas para muchas características, y están divididas en efectos dominantes que reflejan las interacciones entre los diferentes alelos en el mismo locus, y efectos epistáticos resultado de las interacciones entre los diferentes loci.

La acción de los genes y los valores de cultivo (valor de cría?) se caracterizan mediante pruebas de progenie, en las cuales los fenotipos de individuos de la población se comparan con los padres y hermanos producidos por autopolinización o cruce. Los esfuerzos previos para desarrollar un gran número de marcadores moleculares, mapas genéticos de alta densidad y mapeo de poblaciones apropiadamente estructurado, han hecho rutina la habilidad de definir simultáneamente la acción del gen y su valor de cultivo en cientos y a veces miles de loci distribuidos relativamente uniforme entre genomas completos, en una gran variedad de cultivos. Los resultados de estos estudios de mapeo proporcionan estimados altamente mejorados de el número de loci, efectos alelicos y acción del gen que controlan las características de interés. Y aún más importante, se pueden identificar fácilmente segmentos genómicos que muestren asociaciones estadísticamente significativas con características cuantitativas (loci de características cuantitativas, QTL por sus siglas en inglés). Además del mapeo genético en familias derivadas de cruces biparentales, nuevos avances en la genética de asociación con genes candidatos y acercamientos que combinan el análisis de desequilibrio por ligamiento en familias y poblaciones (Holland, 2007; Yu et al., 2008) en un futuro aumentarán el poder de los QTL.

La información sobre los QTL puede ser utilizada de diversas maneras para incrementar la heredabilidad y la acción benéfica de los genes.

Para los rasgos que tienen heredabilidad moderada o baja, como el rendimiento de grano, los QTL y otros marcadores moleculares asociados generalmente cuentan más para una mayor proporción de efectos genéticos aditivos, que el fenotipo solo. Además, el conocimiento de la arquitectura genética se puede explotar para añadir o eliminar alelos específicos que

contribuyan al valor de cultivo. Cuando la epistasis o el ligamiento genético entre loci con efectos antagónicos en una característica limitan la ganancia genética, la información de los QTL puede ser usada para romper estas indeseables relaciones. El éxito para incrementar la ganancia genética utilizando la información sobre QTL depende en gran medida de la magnitud de los efectos de los QTL, la estimación precisa de su posición, la estabilidad de sus efectos en múltiples ambientes y si los QTL son robustos en el germoplasma mejorado. La predicción de la posición de los QTL se puede aumentar con mapeos finos, los cuales facilitan las pruebas de los efectos de QTL y los valores en poblaciones adicionales.

Cuando la densidad de las recombinaciones observadas se acerca a la resolución de genes aislados, el cambio genético causal del QTL puede ser determinado (para revisión, Salvi y Tuberosa, 2005; Yu and Buckler, 2006; Belo' et al., 2008; Harjes et al., 2008). El aislamiento molecular de estos QTL permite el desarrollo de marcadores moleculares perfectos o funcionales con el potencial de resolución de la unidad fundamental de la herencia, el nucleótido, y aumenta dramáticamente la especificidad y precisión con las cuales los efectos genéticos se estiman y manipulan en los programas de mejoramiento.

El uso de transgenes podría simplificar la arquitectura genética de características deseables de formas superiores, o incluso que no son posibles cuando hay disponibles marcadores perfectos para QTL robustos de efecto amplio. Típicamente los transgenes condicionan efectos genéticos fuertes a particulares loci operacionales, los cuales también exhiben acción de genes dominantes en donde sólo se necesita una copia del evento para lograr la máxima expresión en un cultivar híbrido. Las características de los transgenes pueden reducir el complicado proceso de mejora cuantitativa a una solución directa y frecuentemente dramática. Un excelente ejemplo es la expresión de toxinas proteicas insecticidas de *Bacillus thuringiensis* (Bt) en híbridos transgénicos de maíz para reducir el daño por herbivoría provocado por *Ostrinia nubilalis* y *Diabrotica* spp.

La resistencia parcial de germoplasmas de maíz a estas plagas había sido caracterizada como una característica cuantitativamente heredable con poca heredabilidad (Papst et al., 2004; Tollefson, 2007), pero los eventos transgénicos con Bt ofrecen una alternativa heredable simple que es eficientemente manipulada en los programas de mejoramiento.

A través de la simplificación de la estructura génica, los transgenes también permiten la interrupción de las interacciones alélicas entre factores que controlan el rasgo de interés y otras características de rendimiento importantes. Por ejemplo, utilizando un recurso transgénico de resistencia a

insectos (por ejemplo un solo locus del transgen Bt) se puede facilitar la selección de alelos favorables para mejorar el rendimiento que están reprimen fuertemente unidos a genes endógenos de resistencia a estas plagas. Además, los eventos transgénicos pueden ser diseñados para desacoplar efectos pleiotrópicos negativos en fenotipos benéficos condicionados por mutaciones recesivas. Esta aplicación se puede ilustrar con el uso de ARN interferente para disminuir específicamente la expresión del gen de almacenamiento de zeína en semilla (Segal et al., 2003; Huang et al., 2004). Esta estrategia imita los efectos de la mutación *opaque2*, mejorando el perfil de aminoácidos en granos de maíz para alimentación animal, a la vez que evita la textura suave del endospermo y la susceptibilidad a hongos patógenos típicamente asociados con *opaque2*.

Los eventos transgénicos también pueden ser diseñados para intervenir en pasos regulatorios clave para una ruta metabólica completa o para rutas de desarrollo, tales como aquellos genes de acción para el rasgo correspondiente que son heredados en factores dominantes únicos que son menos sensible a los efectos ambientales.

Algunos ejemplos incluyen la expresión de un factor de transcripción que incrementa la tolerancia a la sequía (Nelson et al., 2007) y la alteración del balance entre los niveles del factor de transcripción GLOSS15 y su represor, el micro ARN 172, para retardar la floración en híbridos de maíz (Lauter et al., 2005).

La biotecnología también facilita el agrupamiento de transgenes que controlan un rasgo, o serie de rasgos, en un haplotipo de un solo locus definido por un evento transgénico.

Ejemplos de esta estrategia incluyen, el Arroz Dorado inicial (Ye et al., 2000), el reciente lanzamiento de híbridos de maíz triplemente transgénico Yield-Guard VT, en el cual, las características de tolerancia a herbicidas y resistencia a múltiples insectos se integran en un solo locus genómico (<http://www.yieldgardvt.com/VTScience/Default.aspx>), o la combinación de transgenes que simultáneamente incrementan la síntesis y disminuyen el catabolismo de lisina en semillas de maíz (Frizzi et al., 2008). Reportes recientes sobre mejoras en la tecnología de marcaje de genes (Ow, 2007) y la construcción de minicromosomas de plantas meióticamente transmisibles (Carlson et al., 2007; Yu et al., 2007) preparan el terreno para introducir más características con mayor complejidad. Con estos logros la biotecnología es ahora en posición de ensamblar diversidad genética útil, desde esencialmente cualquier recurso, en construcciones que concentren acción genética favorable y maximice la heredabilidad para un conjunto de características muy amplio.

Para cerrar esta sección sobre cómo el mejoramiento molecular de plantas incrementa la acción benéfica del gen, es importante enfatizar que los estudios de QTL, cuando se llevan a cabo con la escala y posición adecuadas para identificar genes causales, representan una estrategia genómica funcional muy poderosa.

La clonación de QTL ha producido nuevas perspectivas acerca de la biología de las características cuantitativas, que no hubieran sido descubiertas desde el análisis de genes mediante estrategias de pérdida de función, o sobreexpresión, en particulares impacto de la variación regulatoria en la variación fenotípica y en la evolución (e.g. Cong et al., 2002; Clark et al., 2006; Yan et al., 2004; Salvi et al., 2007).

Por otra parte, los marcadores moleculares, la genómica y la biotecnología se aplican de una red de trabajo iterativa, explotando así la diversidad genética para la mejora de cultivos. La información genética permite el descubrimiento de alelos benéficos mediante el mapeo y clonación de QTL, seguido del uso de la información aprendida sobre caracterización molecular de QTL, para diseñar estrategias transgénicas óptimas para el mejoramiento de plantas.

El mejoramiento molecular de plantas incrementa la eficiencia de la selección

El mejoramiento convencional de plantas, que se basa sólo en la selección de fenotipos, ha sido históricamente efectiva. Sin embargo, para algunas características, la selección fenotípica ha logrado poco progreso debido a los retos al medir fenotipos o identificar individuos con el mayor valor de mejoramiento. Los efectos de ambiente, interacción ambiente- genotipo y errores de medición, también contribuyen a las diferencias observadas. La evaluación de genotipos en ambientes múltiples con réplicas permite una mejor estimación del valor de mejoramiento, pero requiere gasto y tiempo adicionales. Para algunos rasgos, podría ser necesario sacrificar individuos para medir el fenotipo, o la expresión de un rasgo puede depender de condiciones variables del ambiente (por ejemplo, presión por enfermedades) y el estado de desarrollo (la calidad del grano solo se puede conocer después de la floración). Además los mejoradores típicamente den mejorar de manera simultánea una serie de características valoradas comercialmente, lo que podría limitar la cantidad de grano para seleccionar. Así como el mejoramiento molecular ayuda a expandir la diversidad genética, caracterizar la estructura genética y modificar la acción del gen, sus métodos también pueden ser aplicados para incrementar la eficiencia de la selección.

Una gran parte de la literatura ha considerado la utilidad de la selección asistida por marcadores moleculares y su ajuste con los diferentes métodos de mejoramiento (Fig. 1) tupiendo al lector a un gran número de revisiones excelentes del tema (Dekkers and Hospital, 2002; Holland, 2004; Johnson, 2004; Varshney et al., 2006; Collard and Mackill, 2008). Los genotipos de marcadores moleculares que se utilizan dentro de genes o muy cercanos a los QTL, que influyen rasgos bajo selección, pueden ser empleados como un suplemento a las observaciones fenotípicas en una guía de selección (Lande and Thompson, 1990). En los casos donde las correlaciones genéticas son altas, se pueden ganar eficiencias adicionales sustituyendo la selección genotípica por la fenotípica, durante algunos ciclos de selección, lo que puede reducir los esfuerzos de fenotipaje, así como el número de ciclos que permitan el uso de viveros fuera de temporada. Johnson (2004) resume un ejemplo anticipado de combinar los datos fenotípicos y las puntuaciones de marcadores moleculares para incrementar las ganancias en rendimiento de maíz y resistencia a plagas debida a la selección. Una estrategia efectiva para modificar simultáneamente características múltiples, es el uso de índices de selección que consideren múltiples factores para escoger el fenotipo mejorado final. Recientemente Eathington et al. (2007) reportaron su resultados sobre el uso de índices de múltiples características y la selección asistida por marcadores en cerca de 250 poblaciones únicas de maíz. El uso de marcadores moleculares incrementan la eficiencia de la mejora alrededor de 2 veces en relación a la selección fenotípica, también con ganancias similares en poblaciones de soya (*Glycine max*) y girasol (*Helianthus annuus*).

La selección asistida por marcadores también puede aumentar significativamente la ganancia genética para características en las que el fenotipo es difícil de evaluar ya sea por su costo, o por su dependencia e condiciones ambientales específicas. Los marcadores moleculares se pueden utilizar para fomentar la probabilidad de identificar realmente genotipos superiores, concentrándose en probar recursos de genotipos con el mayor potencial (i.e. con la eliminación temprana de genotipos inferiores), disminuyendo el número de la progenie que se necesita analizar para recuperar un nivel dado de ganancia, y permitiendo la mejora simultanea de características que están negativamente correlacionadas (Knapp, 1998). Algunos ejemplos exitosos incluyen resistencia de la soya a nematodos (Young, 1999), resistencia a enfermedades en cereales (revisado por Varshney et al., 2006) y tolerancia a sequía en maíz (Ribaut and Ragot, 2007; Tuberosa et al., 2007). La eficiencia de la selección fenotípica para lagunas características complejas se puede incrementar si se incluyen fenotipos bioquímicos o fisiológicos como características secundarias, cuando estos exhiben correlaciones genéticas fuertes con la característica

de interés y alta heredabilidad. Los recientes avances en genómica funcional permiten escalar a nivel de población el perfil de abundancia de RNA, niveles y actividades de proteínas y metabolitos que están asociados con características importantes. A parte de los marcadores moleculares que identifican variaciones en la secuencia de ADN, estas estrategias genómicas podrían proveer fenotipos secundarios adicionales como objetivos de selección (Jansen and Nap, 2001; Johnson, 2004), particularmente para aquellas características definidas por su respuesta a factores como medio, estado fisiológico o desarrollo.

La selección asistida por marcadores, también acelera el uso de transgenes en cultivos comerciales. Típicamente, esto se ha logrado mediante la retrocruza asistida por marcadores. Sin embargo, para mejoras futuras en biotecnología, como tolerancia a la sequía o a la limitación de nutrientes, el mejoramiento directo podría requerir de la cooptimización de expresión del transgen y el origen genético, puesto que los genes endógenos y los factores ambientales, podrían tener el potencial de influir en los fenotipos resultantes de las modificaciones transgénicas (Mumm, 2007).

Por supuesto, que el uso de marcadores moleculares, podría también contribuir con los esfuerzos de mejoramiento directo. Alternativamente los esfuerzos por descubrir genes adicionales o QTL, que son necesarios para el desempeño fiable de una característica, podrían proponer el diseño de nuevas construcciones de transgenes que agrupen transgenes primarios con modificadores genéticos, en una segunda generación de eventos transgénicos.

Incremento en la aprobación del mejoramiento molecular de plantas

La aceptación de las estrategias de mejoramiento molecular de plantas ha ocurrido a diferentes ritmos, dentro de las especies y las instituciones comprometidas con el mejoramiento de estos, debido a la influencia combinada de factores científicos, económicos y sociológicos. Algunas barreras científicas importantes, incluyendo la recalcitrancia de los cereales a la transformación mediada por *Agrobacterium* y la falta de conocimientos sobre el control genético, son definidas como objetivos importantes del mejoramiento. El desarrollo continuo de investigación y tecnología ha superado los obstáculos presentes en la transformación de plantas, en casi todos los cultivos y especies de horticultura de importancia económica (Wenzel, 2006). De manera similar la información obtenida de la investigación genómica en plantas y otros organismos ha generado bastante información sobre estructura y función génica, así como grandes cantidades de marcadores moleculares para ser usados en el mejoramiento de plantas.

A pesar de estos recursos, la genética específicamente de QTL robustos permanece ambigua, amén que los programas de mejoramiento y los sistemas de manejo de la información asociada, sean reestructurados para integrar por completo el conocimiento sobre pedigríes, fenotipos, y marcadores de genotipos que puedan ser influenciados para optimizar la respuesta a la selección (Cooper et al., 2004; Eathington et al., 2007). Aún con una integración de este tipo, modificar las funciones regulatorias sigue siendo un reto científico de mejoramiento molecular, puesto que es difícil determinar la secuencia de bases responsable de los cambios regulatorios y predecir los efectos fenotípicos (Morgante and Salamini, 2003).

Una vez que las tecnologías en biotecnología y genómica estén disponibles, los factores económicos generalmente dictan el grado en que estas innovaciones se integren a los programas existentes de mejora de plantas. El costo de ganar la aprobación regulatoria gubernamental, para la liberación comercial de variedades transgénicas (recientemente estimadas por Kalaitzandonakes et al., 2007, entre \$7-\$10 millones) es una barrera económica significativa. El costo asociado con el desarrollo, establecimiento y operación del mejoramiento molecular de plantases mayor que para las prácticas convencionales de mejoramiento (Koeberner and Summers, 2003; Morris et al., 2003), pues requiere inversiones importantes en nueva infraestructura de investigación y capacidad intelectual. Estos recursos sólo existen de forma inicial en compañías agrícolas privadas y en un puñado de instituciones públicas, promoviendo la aceleración con tendencia a incrementar la industrialización de los programas de mejoramiento en cultivos mayores como maíz, soya, algodón (*Gossypium hirsutum*) y trigo (Johnson, 2007). Cuando ha habido aceptación por las compañías, la balanza favorece a la biotecnología sobre los QTL, para mejorar características complejas, a pesar de aumentar los costos de desarrollo del producto, ya que los transgénicos pueden ser diseñados para producir efectos fenotípicos más fuertes, incluso dramáticos y generalmente se pueden desplegar más rápido en un amplio rango del germoplasma, resultando en nuevas soluciones aún más rápido.

A pesar de que actualmente el mejoramiento molecular es considerado por las grandes compañías, como un componente esencial en los esfuerzos de mejoramiento de cultivos mayores, la amplia aplicación de las estrategias moleculares modernas a las técnicas de mejoramiento convencionales, sigue siendo un tema de debate entre algunos mejoradores de plantas que participan en el sector público, particularmente, para cultivos menores (e.g. Gepts, 2002; Goodman, 2004). Además de los muy válidos factores científicos y económicos, que han retrasado o prevenido, la aceptación de las estrategias moleculares para alcanzar algunos objetivos del mejoramiento molecular, al menos hay tres razones adicionales que

contribuyen a esta visión. En primera, el mejoramiento molecular de plantas requiere entrenamiento y experiencia en biología molecular y mejoramiento de plantas. Los esfuerzos educativos que otorgan este tipo de entrenamiento interdisciplinario fueron establecidos a inicios de 1990, pero continúan limitados a un pequeño grupo de instituciones académicas con tradición en mejoramiento de plantas (Guner and Wehner, 2003; Gepts and Hancock, 2006; Guimaraes y Kueneman, 2006). Una segunda razón que reduce el entusiasmo para que la biotecnología sea abrazada por los mejoradores, son los problemas de aceptación de cultivos transgénicos entre algunos gobiernos y grupos de consumidores, como ha sucedido al posponer variedades de trigo con transgenes que brindan resistencia hacia los herbicidas con glifosfato (Sokstad, 2004). Finalmente, la exaltación sobre el potencial del mejoramiento molecular de plantas, también estimula cambios en el financiamiento de instituciones públicas para incrementar la capacidad intelectual y la infraestructura de las investigaciones en genómica y genética molecular, lo cual ocurre, ironicamente, a expensas de el mejoramiento convencional (Knight, 2003).

Este énfasis puede ser temporalmente necesario para establecer los fundamentos de la biología de plantas de siglo 21, pero actualmente, hay un reconocimiento creciente de que el incremento en la inversión en capacidad de mejoramiento de plantas y en investigación de transferencia, uniendo los métodos moleculares con los objetivos de mejoramiento, necesita del desarrollo del potencial de los avances recientes en biotecnología y genómica (Guimaraes and Kueneman, 2006; National Research Council, 2008).

11.2.1.2. CONCLUSIÓN

A pesar de los recientes avances y ejemplos exitosos del mejoramiento molecular de plantas, uno de los grandes retos actuales en la biología de plantas es identificar aquellas combinaciones de genes que llevan a una mejora significativa de cultivos. Este comentario se concluye sugiriendo que la estrategia más efectiva para acelerar estos esfuerzos, es integrar mejor las diferentes disciplinas de investigación y las actividades centrales que conforman el mejoramiento molecular de plantas. Como aquí se describe, y anteriormente por otros (e.g. Gepts y Hancock, 2006; Bliss, 2007), esta integración requiere el conocimiento de la organización genómica y de la función de los genes, fundamentos sólidos en estrategias estadísticas par estimar los efectos genéticos, bases fuertes sobre biología de plantas, experiencia con los métodos de laboratorio sobre biología molecular y genómica funcional, así como en práctica en campo sobre mejoramiento, y la habilidad de manejar grandes conjuntos de datos de diversos tipos. Aunque se reconoce, con conciencia y apreciación la importancia de cada una de estas disciplinas por todos los científicos de

plantas la formación de estudiantes, y los esfuerzos de entrenamiento, los fondos de financiamiento y los programas de investigación, todavía enfatizan en subgrupos específicos del paradigma de mejoramiento molecular de plantas. Lo anterior se justifica dada la amplitud de las disciplinas involucradas, sin embargo, se necesita hacer mayores esfuerzos para implementar programas educativos y de investigación que promuevan la participación activa en el mejoramiento molecular de plantas para la mejora de cultivos. Este tipo de programas debería ser desarrollado fomentando las colaboraciones entre grupos con experiencia complementaria. Absolutamente, las agencias de financiamiento deben expandir su portafolio de proyectos que da apoyo para la transferencia tecnológica mediante requerimiento de propuestas que integren esfuerzos en investigación básica y también en productos tangibles de mejoramiento de plantas. Algunos ejemplos dignos de imitarse incluyen el programa Harvest-Plus (<http://www.harvestplus.org/>), el MASwheat (Dubcovsky, 2004), y el U.S. Department of Agriculture Coordinated Agricultural Projects. El sector privado también debe seguir con inversiones que estimulen la integración y brinde el ambiente de entrenamiento apropiado para los futuros científicos que se integren a la fuerza de trabajo del mejoramiento molecular. Además de apoyo directo para entrenamiento de graduados y el patrocinio de investigación, las compañías seguidas pueden proporcionar apoyo "en especie" que ayude a reducir la brecha en tecnología en expansión entre el sector de investigación en mejoramiento molecular de plantas público y privado. Con los esfuerzos colectivos de la amplia comunidad de científicos comprometidos con la biología de plantas y la mejora de cultivos, el mejoramiento molecular de plantas podrá expandir su impacto y contribuciones que dan respuesta a las necesidades de incremento sustentable de la producción agrícola.

Tarea 11-1

Haga un resumen en español de los siguientes párrafos:

Plant breeding has a long history of integrating the latest innovations in biology and genetics to enhance crop improvement. Prehistoric selection for visible phenotypes that facilitated harvest and increased productivity led to the domestication of the first crop varieties (Harlan, 1992) and can be considered the earliest examples of biotechnology. Darwin outlined the scientific principles of hybridization and selection, and Mendel defined the fundamental association between genotype and phenotype, discoveries that enabled a scientific approach to plant breeding at the beginning of the 20th century (e.g. Shull, 1909). Despite the immediate recognition among some plant breeders of the importance of Mendelian genetics, full integration was delayed for nearly 20 years until quantitative genetics reconciled Mendelian principles with the continuous variation observed for most traits considered important by most plant breeders (Paul and Kimmelman, 1988). Subsequent advances in our understanding of plant biology, the analysis and induction of genetic variation, cytogenetics, quantitative genetics, molecular biology, biotechnology, and, most recently, genomics have been successively applied to further increase the

scientific base and its application to the plant breeding process (e.g. Baenziger et al., 2006; Jauhar, 2006; Varshney et al., 2006).

The plant biotechnology era began in the early 1980s with the landmark reports of producing transgenic plants using *Agrobacterium* (Bevan et al., 1983; Fraley et al., 1983; Herrera-Estrella et al., 1983). Molecular marker systems for crop plants were developed soon thereafter to create high-resolution genetic maps and exploit genetic linkage between markers and important crop traits (Edwards et al., 1987; Paterson et al., 1988).

By 1996, the commercialization of transgenic crops demonstrated the successful integration of biotechnology into plant breeding and crop improvement programs (Koziel et al., 1993; Delannay et al., 1995). As depicted in Figure 1, introgression of one or a few genes into a current elite cultivar via backcrossing is a common plant breeding practice. Methods for marker assisted backcrossing were developed rapidly for the introgression of transgenic traits and reduction of linkage drag, where molecular markers were used in genome scans to select those individuals that contained both the transgene and the greatest proportion of favorable alleles from the recurrent parent genome (e.g. Ragot et al., 1995; Johnson and Mumm, 1996).

During the past 25 years, the continued development and application of plant biotechnology, molecular markers, and genomics has established new tools for the creation, analysis, and manipulation of genetic variation and the development of improved cultivars (for review, see Sharma et al., 2002; Varshney et al., 2006; Collard and Mackill, 2008). Molecular breeding is currently standard practice in many crops, with the following sections briefly reviewing how molecular information and genetic engineering positively impacts the plant breeding paradigm.

11.2.2. Modernización de viejas herramientas

Metodologías de mejoramiento:

Metodologías empíricas	Metodologías clásicas	Metodologías modernas	Metodologías innovadoras
Selección y serendipia.	Hibridización, recombinación, selección por pedigrí.	Marcadores moleculares, fisiología, bioquímica.	Estrategias de fenotipo, dobles haploides, MARS, transformación, etc.

Unidad de selección:

Metodologías
clásicas



Metodologías
innovadoras

Fenotipo	Genotipo	Haplotipo
Poblaciones.	Plantas.	Combinación
Selección en	Líneas	de alelos
base a	individuales.	individuales.
grupos de	Selección en	Selección en
individuos.	base al	base a
Pools de	genotipo de	haplotipos.
genes.	un solo	
	individuo.	

11.2.2.1. Parámetros primarios y secundarios

En maíz, el parámetro agronómico más importante es el rendimiento de grano por hectárea. Este por su parte puede ser expresado en dos componentes: el número total de granos (KN de *kernel number*) multiplicado por el peso promedio de ellos (KW de *kernel weight*). Para seleccionar líneas de maíz, por lo regular se les compara en términos de los rendimientos de sus combinaciones híbridas. Es decir, no importa tanto el rendimiento de la línea *per se*, sino mas bien el rendimiento de los híbridos que puede formar esa línea con algún probador del grupo heterotico opuesto.

Adicionalmente a los parámetros primarios de rendimiento, existen otros parámetros que también son característicos de cada genotipo. Estos pueden ser por ejemplo, el número de mazorcas por planta (EPP de *Ear number per plant*), los días de floración femenina (SD de *Silking days*), días de floración masculina (AD de *anthesis days*), intervalo de floración femenina-masculina (ASI de *anthesis-silking interval*), la senescencia (SEN), el enrollamiento de las hojas bajo sequía (ROL), el contenido de clorofila en determinada etapa de desarrollo (SPAD), etc. A estos y otros parámetros que no son primariamente de rendimiento de grano se les llama caracteres secundarios (*secondary traits*). Por lo regular son más fáciles o más rápidos de medir que el rendimiento final de grano. Algunos de los parámetros secundarios mas importantes, son el número promedio de mazorcas por planta (EPP) y el intervalo de floración (ASI), sobre todo bajo condiciones de sequía.

11.2.2.2. Índices de Selección

Muchas veces, al agricultor no solo le interesa un solo carácter sino varios. Los índices de selección se usan precisamente cuando uno quiere evaluar varios parámetros al mismo tiempo.

Ejemplo de un análisis multicaracter en animales: Aunque las hijas del toro 1 produzcan más leche que las hijas del toro 2 estas pueden tener mejor calidad de las proteínas para la producción de queso. Por ello se utilizan herramientas matemáticas para combinar todos los parámetros de selección en una sola ecuación. El resultado se expresa como un índice de selección. De esta forma se puede seleccionar no solo para un único carácter (producción de leche) si no también para una combinación de varios caracteres a la vez (producción y calidad). Por ejemplo, nos puede interesar la producción de queso en términos de kilos de queso de buena calidad por vaca, más que la producción de litros de leche por vaca. Los factores que se le asignan a cada variable que se usa para calcular el índice de selección permiten darle un peso diferente a cada uno de los parámetros de interés.

Ejemplo de un análisis multicaracter en plantas:

El índice de selección también permite incluir parámetros secundarios para así mejorar la selección de un **carácter primario**. La inclusión de **caracteres secundarios** ayuda a reducir el error de las mediciones y de los experimentos. De esta forma se puede discriminar mejor los genotipos que son genéticamente superiores a pesar de una variabilidad ambiental muy grande.

Veamos un ejemplo que se refiere a un experimento de campo bajo condiciones de sequía:

		Genotipo	Rendimiento de grano	Floración Masculina	Intervalo Floración	Mazorcas por planta	Índice de Selección
REP	PLOT	ENTRY	GYF (ton/ha)	AD (days)	ASI (days)	EPP	Selection Index
1	1	1	3.00	60	6	1.05	-0.22
2	4	1	1.60	63	13	0.66	-4.46
3	7	1	3.46	60	12	1	-1.22
1	2	2	2.99	63	13	0.88	-2.32
2	5	2	3.23	63	10	1.04	-0.87
3	8	2	4.87	62	8	1.1	1.34
1	3	3	2.71	63	3	1.46	1.62
2	6	3	4.95	63	1	1.52	4.41
3	9	3	3.19	64	2	1.3	1.71

Tomando los datos de la tabla arriba debemos encontrar el mejor genotipo (ENTRY). En primer lugar nos interesa el rendimiento de grano en campo (GYF). Para ello calculamos los promedios de las 3 repeticiones por entrada y obtenemos la siguiente tabla resumen:

PROMEDIOS	ENTRY	GYF (ton/ha)	AD (days)	ASI (days)	EPP	Selection Index
	1	2.69	61.0	10.3	0.9	-2.0
	2	3.70	62.7	10.3	1.0	-0.6
	3	3.62	63.3	2.0	1.4	2.6

Parece que la entrada 2 con 3.7 toneladas es la que mas rinde, sin embargo, ¿es una diferencia significativa? Para contestar esa pregunta debemos calcular los errores estándares y hacer las pruebas T.

Errores Estándar	ENTRY	GYF (ton/ha)	AD (days)	ASI (days)	EPP	Selection Index
	1	0.56	1.00	2.19	0.12	1.28
	2	0.59	0.33	1.45	0.07	1.06
	3	0.68	0.33	0.58	0.07	0.92

Pruebas T (valor de P)	ENTRY	GYF (p value)	AD (p value)	ASI (p value)	EPP (p value)	Selection Index (p value)
1 vs 2		0.14	0.12	0.50	0.26	0.23
1 vs 3		0.18	0.07	0.03	0.02	0.03
2 vs 3		0.47	0.12	0.01	0.01	0.04

Podemos ver que los errores estándares del rendimiento son de 0.56 a 0.68, que son mas grandes que las diferencias de rendimiento entre genotipos. La prueba T de la comparación entre la entrada 2 y la 3 nos da un valor de p no significativo ($p=0.47$). De hecho, si consideramos solo el rendimiento, ningún genotipo es significativamente mejor o peor que el otro ya que los valores p son mayores a 0.05. ¿Que se hace en esos casos? Una opción es volver a sembrar esos mismos genotipos ya sea con muchas más repeticiones o en más localidades. Solo así se obtendrán suficientes datos de rendimiento para tener mayor certeza de escoger el genotipo superior. Sin embargo, repetir ensayos o hacerlos más grandes es bastante laborioso y caro. Afortunadamente los mejoradores han desarrollado algunos procedimientos para ahorrarse estos costos.

El uso de parámetros fisiológicos secundarios y el cálculo de índices de selección es un ejemplo. De muchos experimentos anteriores bajo condiciones de sequía se ha visto que el rendimiento de grano (GYF) esta correlacionado con el intervalo de floración (ASI) y también con el número de mazorcas por planta (EPP). La correlación es positiva para EPP y negativa para ASI. Si el rendimiento es el parámetro primario, al EPP y ASI se les llama parámetros secundarios. Si queremos ver cual genotipo es mejor, adicionalmente a buscar la entrada que tenga el GYF mayor, también se puede considerar el genotipo con el EPP mayor o el ASI menor.

Si revisamos la tabla de resultados vemos que el genotipo 3 tiene un EPP mayor y un ASI menor a los demás. Los tres genotipos son muy parecidos en términos de rendimiento. Sin embargo, si adicionalmente consideramos el EPP y el ASI podemos ver que la entrada 3 es mucho mejor que todas las demás. Esto lo concluimos de los datos que ya tenemos del ensayo sin necesidad de repetir el experimento. Es decir, el uso correcto de las matemáticas y la estadística a veces nos ayuda a ahorrarnos trabajo experimental.

Para combinar los diferentes parámetros en un solo índice, primero hay que estandarizar los valores numéricos de cada parámetro. No podemos sumar toneladas con días de floración o número de mazorcas. Las unidades de cada medición son muy diferentes. Si no normalizamos los datos no podremos calcular un índice de selección.

Una forma de estandarizar los datos es restarle el valor promedio a cada dato, y el resultado dividirlo entre la desviación estándar. De esta forma obtenemos un valor sin unidades, que se refiere a que tantas desviaciones estándares arriba (valor positivo) o abajo (valor negativo) del promedio se encuentra cada genotipo con relación al parámetro secundario. El selection index se calcula entonces sumando o restando los diferentes valores estandarizados dependiendo si están positivamente o negativamente correlacionados con el rendimiento. El índice de selección de las tablas anteriores fue calculado tomando en cuenta los datos estandarizados de GFY, EPP y ASI de la siguiente manera:

$$\text{Selection Index} = \text{GFY}_{\text{est}} - \text{ASI}_{\text{est}} + \text{EPP}_{\text{est}}$$

$$\text{GFY}_{\text{est}} = \frac{\text{GFY} - \text{Promedio}(\text{GFY})}{\text{DesviaciónEstandar}(\text{GFY})}$$

$$\text{ASI}_{\text{est}} = \frac{\text{ASI} - \text{Promedio}(\text{ASI})}{\text{DesviaciónEstandar}(\text{ASI})}$$

$$\text{EPP}_{\text{est}} = \frac{\text{EPP} - \text{Promedio}(\text{EPP})}{\text{DesviaciónEstandar}(\text{EPP})}$$

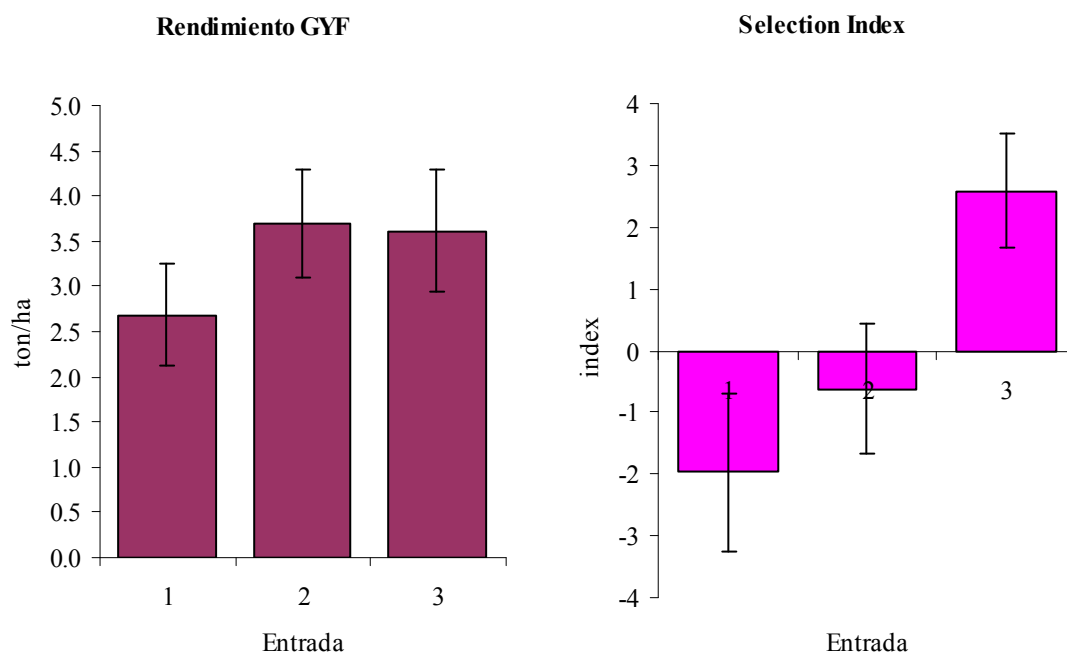


Figura 11-7. Grafica de rendimiento de grano (*grain yield field GYF*). Las diferencias de GYF no son significativas. A la derecha se muestra la grafica con los valores del índice de selección. Las diferencias si son significativas.

11.2.3. Selección asistida por marcadores

Tradicionalmente, la única forma de saber si la descendencia de una cruce en particular había heredado caracteres útiles (p. ej. tolerancia a la sequía, resistencia a enfermedades o calidad de grano) era sembrando los materiales en parcelas y evaluando las plantas adultas. La selección asistida por mejoradores acelera el proceso de mejoramiento, ya que hace posible rastrear la presencia de genes útiles en cada generación. Esto no excluye la necesidad de hacer evaluaciones en el campo, pero hace muchísimo más eficiente el proceso. “Para la selección en el campo necesitamos tiempo, espacio y recursos, y nuestra capacidad es limitada”, explica Gary Atlin, mejorador de maíz del CIMMYT, “pero si aplicamos MAS podremos emplear los recursos de manera más eficaz, concentrándonos en las mejores líneas y filtrando aquellas que no hayan heredado los caracteres que nos interesan.”

Cuando los investigadores quieren saber si una línea de trigo o de maíz en particular tiene la versión útil de un gene (por ejemplo, resistencia en lugar de susceptibilidad a las enfermedades), emplean secciones adyacentes e identificables de DNA conocidas como marcadores, a las cuales se aplica tintura fluorescente para distinguirlas. Las distintas versiones de marcadores y genes se llaman alelos. El DNA que se encuentra junto al cromosoma suele mantenerse unido durante generaciones, de manera que

cierto alelo de un marcador, por lo general, es heredado junto con el alelo útil de un gene adyacente. Con el uso de nuevas máquinas de electroforesis capilar, la muestra es forzada a través de un tubo capilar estrecho bajo la influencia de una corriente eléctrica. Un láser al final del tubo detecta los diferentes alelos de los marcadores fluorescentes y esto indica al científico si la muestra contiene los alelos que está buscando.

Además de ser rápida y menos costosa en lo que respecta a mano de obra, la electroforesis capilar permite hacer pruebas con más de un marcador y colocar más de una muestra al mismo tiempo en cada tubo. Al utilizar colores de tinte fluorescente ligeramente distintos para cada muestra, es posible distinguir los marcadores en cada una de ellas, igual que los grupos de corredores que utilizan camisetas de distintos colores. Para lograr una máxima eficiencia, los científicos pueden también formar grupos de muestras que pueden examinar con pequeños intervalos de tiempo, igual que cuando los corredores se preparan para el momento de salir. El CIMMYT podrá incluso generar un marcadores tipo SNP (single nucleotide polymorphism), que son automatizables y permiten hacer ensayos con numerosos caracteres simultáneamente y aporta muchos más datos por muestra.

Todo esto quiere decir que las nuevas máquinas tienen mayor capacidad de producción y que se pueden analizar más muestras con la misma cantidad de mano de obra, reduciendo de manera impresionante el costo individual de las muestras. Actualmente la mayor limitante del uso de los marcadores es el costo. Si la selección asistida por marcadores fuera significativamente más económica, sin duda la emplearía en el mejoramiento de maíz dentro de instituciones de mejoramiento públicas.

En efecto, se puede hacer la transferencia de genes en variedades mejoradas en poco tiempo. Si uno hace una retrocruza entre un donador que contiene una característica genética y un progenitor mejorado con buen comportamiento agronómico, está tratando de seleccionar una característica del donador, aunque distinta de todos sus otros genes. Con varios marcadores, la selección asistida por marcadores indica exactamente cuál progeñie posee la combinación de los genes deseados de un donador y los genes buenos del otro progenitor. Se pueden obtener resultados en dos generaciones, en comparación con las cuatro o cinco que normalmente había que evaluar.

El problema de la selección asistida por marcadores es encontrar genes con efectos substanciales, en especial de caracteres complejos, como la tolerancia a la sequía en el maíz. Se que aún habrá que encontrar esos genes. “En el pasado, era difícil aplicar en el mejoramiento donadores con genes o caracteres únicos útiles pero que por lo demás poseen cualidades agronómicas pobres, en virtud de que esto implicaba introducir mucho

material que no era de utilidad. Por medio de la selección asistida por marcadores nos podemos deshacer del material que no es útil, y esto nos abre la posibilidad de encontrar el gen útil en una amplia variedad de progenitores. La tecnología de los genotipos está volviéndose menos costosa y más eficiente para identificar genes todo el tiempo." Fuente: <http://www.cimmyt.org/spanish/wps/news/2007/aug/genetic.htm>

El mejoramiento genético actual se base en el efecto infinitesimal, basándonos en fenotipos y genealogía. Se aplica una metodología estadística y se predicen cuáles son los mejores alelos (mide la parte aditiva del fenotipo de un individuo). La idea es poder usar marcadores moleculares o los propios genes para tener la inferencia genotípica, lo que no nos ahorra hacer lo que hemos mencionado. Si desconocemos el gen, entonces podemos estudiar un microsatélite que esté asociado al gen. En estos caso se habla de *Marker Assisted Selection* (**MAS**). Si en cambio sí usamos el gen porque lo conocemos entonces se llama *Gene Assisted Selection* (**GAS**).

Puedo tener marcadores directos (en el mismo gen mayor) o indirectos (ligado al gen mayor). Lo ideal seria tener un marcador directo, pero también presenta problemas.

MAS es útil para caracteres:

- Unigenicos o oligogénicos.
- limitados a un solo ambiente no disponible en la estación experimental.
- difíciles o caros de medir, como la resistencia a enfermedades.
- mesurables solo cuando la planta madura, como el rendimiento.

El MAS en especialmente útil en animales para caracteres:

- de baja heredabilidad, como los reproductivos.
- limitados a un sexo, como la producción de leche.
- difíciles o caros de medir, como la suceptibilidad a enfermedades.
- mesurables después del sacrificio del animal, como por Ej., la calidad de la carne...

11.2.4. Nuevos esquemas de mejoramiento

Cambio de enfoque para la unidad de selección Individuo → Haplotipo

Unidad de selección:

Metodologías
clásicas



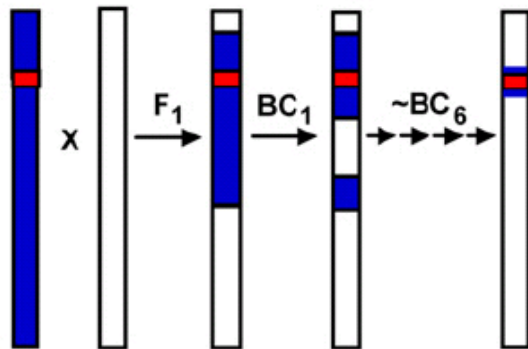
Metodologías
innovadoras

Fenotipo
Poblaciones.
Selección en
base a
grupos de
individuos.
Pools de
genes.

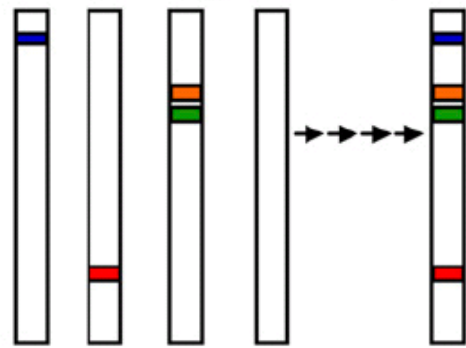
Genotipo
Plantas.
Líneas
individuales.
Selección en
base al
genotipo de
un solo
individuo.

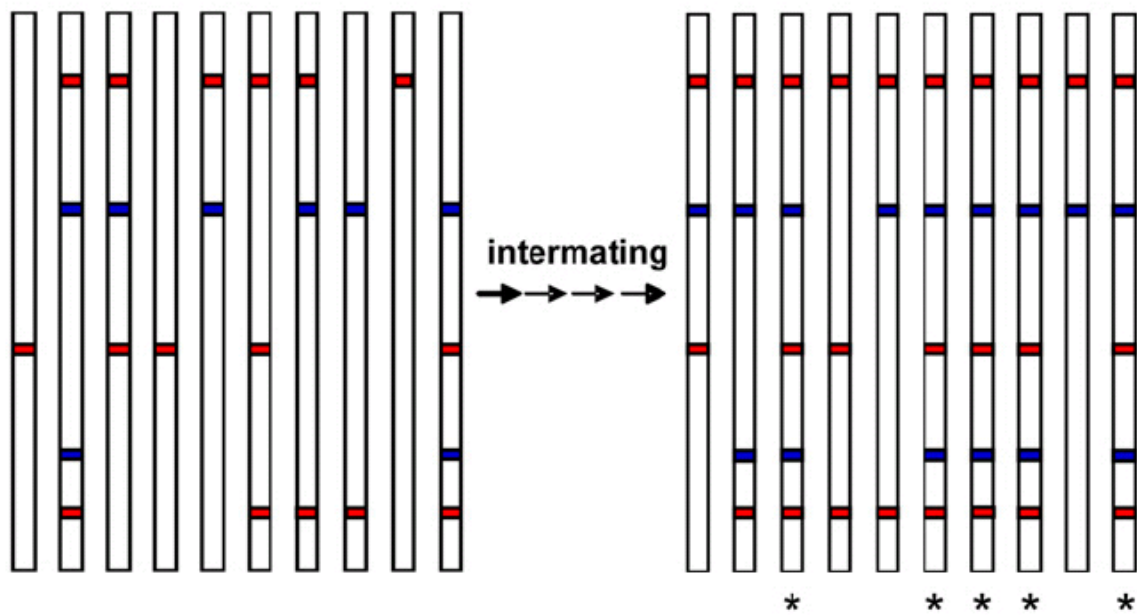
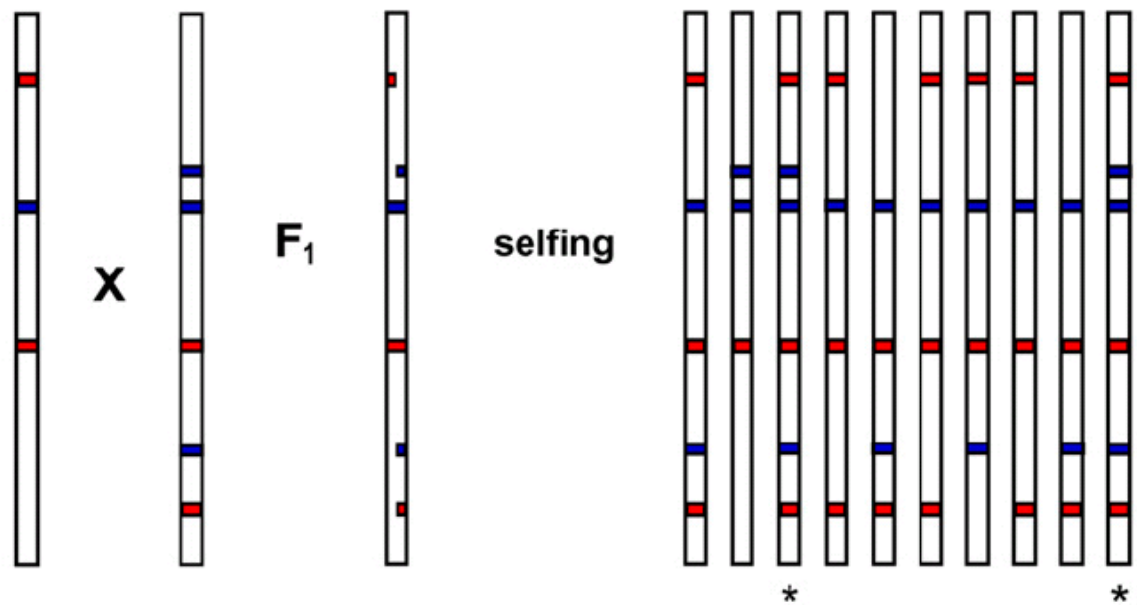
Haplotipo
Combinación
de alelos
individuales.
Selección en
base a
haplotipos.

A backcrossing



B gene pyramiding



D recurrent selection**C pedigree breeding****Figura 11-8.**

Common breeding and selection schemes. Each vertical bar is a graphical representation of the genome for an individual within a breeding population, with colored segments indicating genes and/or QTLs that influence traits under selection. Genes associated with different traits are shown in different colors (e.g. red, blue). “X” indicates a cross between parents, and arrows depict successive crosses of the same type. Asterisk below an individual signifies a desirable genotype. A, Backcrossing. A donor line (blue bar) featuring a specific gene of interest (red) is crossed to an elite line targeted for improvement (white bar), with progeny repeatedly backcrossed to the elite line. Each backcross cycle involves selection for the gene of interest and recovery of increased proportion of elite line

genome. B, Gene pyramiding. Genes/QTLs associated with different beneficial traits (blue, red, orange, green) are combined into the same genotype via crossing and selection. C, Pedigree breeding. Two individuals with desirable and complementary phenotypes are crossed; F1 progeny are self-pollinated to fix new, improved genotype combinations. D, Recurrent selection. A population of individuals (10 in this example) segregate for two traits (red, blue), each of which is influenced by two major favorable QTLs. Intermating among individuals and selection for desirable phenotypes/genotypes increases the frequencies of favorable alleles at each locus. For this example, no individual in the initial population had all of the favorable alleles, but after recurrent selection half of the population possesses the desired genotype. For hybridized crops, recurrent selection can be performed in parallel within two complementary populations to derive lines that are then crossed to form hybrids; this method is called reciprocal recurrent selection.

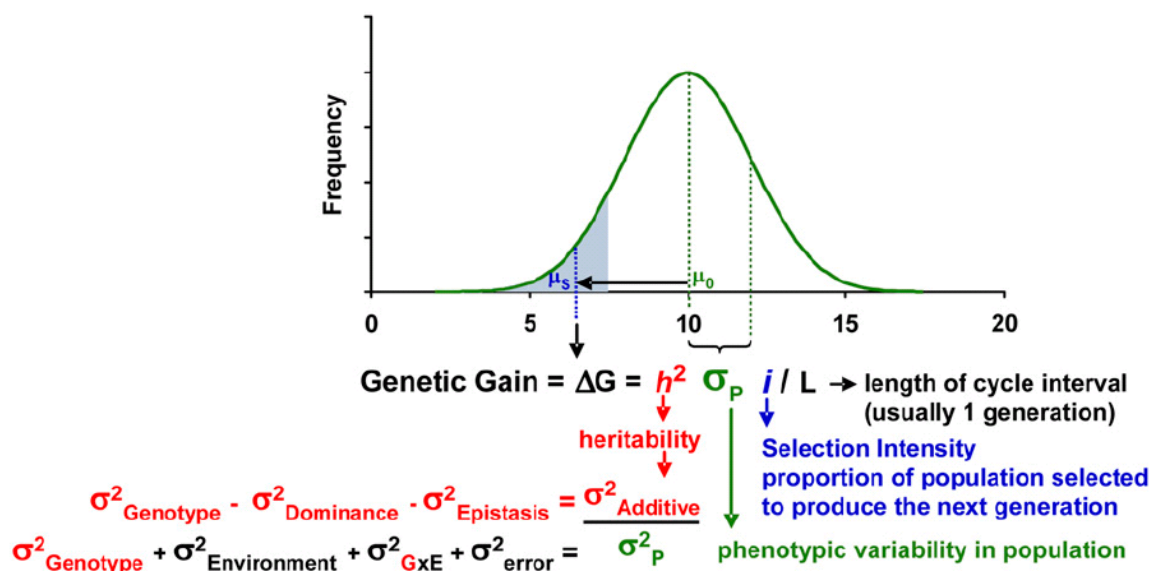


Figura 11-9.

The genetic gain equation and its component variables. The top portion illustrates an idealized distribution showing the frequency of individuals within a breeding population (y axis) that exhibit various classes of phenotypic values (x axis). Mean phenotypic value (μ_0) of the original population (shown as entire area under the normal curve) and mean (μ_s) for the group of selected individuals (shaded in blue) are indicated. In this generalized example, trait improvement is achieved by selecting for a lower phenotypic value, e.g. grain moisture at harvest in maize. Components of variation (σ^2) that contribute to the SD of the phenotypic distribution (σ_P) are indicated below the histogram.

11.3. Retos para el futuro

Uso de NIR como marcadores bioquímicos

Uso de DIESI-MS como marcadores bioquímicos

Uso de pirosecuenciador para genotipaje (marcadores mas baratos)

Crió conservación de polen (facilidad para hacer cruzas de prueba para mejoramiento)

Agente que incremente la tasa de recombinación genética (tamaño de haplotipos mas pequeños)

Agente que afecte la metilación del DNA y de esta forma se reduzca la Heterosis. Tener plantas con vigor hibrido sin que sean cruzas. Compañías semilleras interesadas en mejor producción de semilla hibrida.

Mutación inducida en ciertos genes (oligos) sin que sea transgénico

Dobles haploides (ya se esta haciendo con maíz para generar líneas homocigóticas mucho mas rápido)

12. Mejoramiento molecular

12.1. Desarrollo histórico del mejoramiento molecular de plantas

El mejoramiento de plantas es un término utilizado para describir la creación, selección y establecimiento de fenotipos superiores de plantas, dentro del desarrollo de cultivos mejorados que respondan a las necesidades tanto de productores como de consumidores. Las principales metas del mejoramiento de plantas con cultivos agrícolas y hortícolas, se han enfocado a mejorar el rendimiento, la calidad nutricional y otras características de valor comercial. Dentro de la escala global, el paradigma del mejoramiento de plantas ha sido extraordinariamente exitoso, encontrando ejemplos importantes como el desarrollo de híbridos de maíz (*Zea mays*; Duvick, 2001), la introducción de variedades de trigo (*Triticum aestivum*) y arroz (*Oryza sativa*) que generaron la revolución verde (Everson y Golin, 2003) y la reciente comercialización de cultivos transgénicos (James 2007). Estos y muchos otros productos resultado del mejoramiento de plantas, han contribuido a los múltiples beneficios que la sociedad en general ha recibido de los mayores suministros sustentables de carbono, los cuales son cosechados como alimento, bosques, fibras y combustible.

El mejoramiento de plantas tiene una larga historia integrando las últimas innovaciones en biología y genética para optimizar el desarrollo de cultivos mejorados. La selección prehistórica de fenotipos visibles que facilitaban la cosecha e incrementaban la productividad, llevaron a la domesticación de las primeras variedades de cultivos (Harlan, 1992). Esta domesticación puede ser considerada como uno de los primeros ejemplos de biotecnología.

Con Darwin, que delinea los principios científicos para la hibridación y selección, y Mendel, quien define la asociación fundamental entre genotipo y fenotipo, fue posible hacer mejoramiento de plantas desde un acercamiento científico a inicios del siglo 20 (e.g. Shull, 1909).

A pesar de que algunos mejoradores reconocieron inmediatamente la importancia de la genética mendeliana, la integración total se retrazo alrededor de 20 años, hasta que la genética cuantitativa reconcilió los principios de Mendel con la continua variación observada en la mayoría de los rasgos considerados importantes por los mejoradores (Paul and Kimmelman, 1988). Los subsecuentes avances en nuestro entendimiento de

la biología de plantas, análisis e inducción de la variación genética, genética celular y cuantitativa, biología molecular, biotecnología y más recientemente de genómica, se han aplicado exitosamente para incrementar la base científica y su aplicación en el proceso de mejoramiento de plantas (e.g. Baenziger et al., 2006; Jauhar, 2006; Varshney et al., 2006).

La era de la biotecnología de plantas comienza a inicios de los años ochenta, con los conocidos reportes de producción de plantas transgénicas utilizando *Agrobacterium* (Bevan et al., 1983; Fraley et al., 1983; Herrera-Estrella et al., 1983). A partir de entonces se desarrollaron sistemas de marcadores moleculares para diversos, para crear mapas genéticos de alta resolución y así explotar la relación entre marcadores y características importantes de los cultivos (Edwards et al., 1987; Paterson et al., 1988).

Para 1996 la comercialización de cultivos transgénicos demostró la exitosa integración de la biotecnología en el mejoramiento de plantas y en los programas de mejora de cultivos (Koziel et al., 1993; Delannay et al., 1995). Como se observa en la figura 1, la introgresión de uno, o algunos genes en un cultivo, comúnmente utilizado retrocruzas, es una práctica común de mejoramiento. Algunos métodos de retrocruza asistida por marcadores fueron desarrollados rápidamente para hacer introgresión de características transgénicas y reducir el arrastre por ligamiento (linkage drag), en donde los marcadores moleculares se utilizaron para explorar el genoma y seleccionar aquellos individuos que contuvieran el transgen y la proporción más alta de alelos favorables del genoma parental recurrente (e.g. Ragot et al., 1995; Johnson and Mumm, 1996).

Durante los últimos 25 años, el desarrollo y aplicación continua de la biotecnología de plantas, marcadores moleculares y genómica, han establecido nuevas herramientas para la creación, análisis y manipulación de la variación genética, así como el desarrollo de cultivos mejorados (para revisar, ver Sharma et al., 2002; Varshney et al., 2006; Collard and Mackill, 2008). Puesto que el mejoramiento molecular es una práctica estándar actualmente en varios cultivos, en las siguientes secciones revisaremos brevemente el impacto positivo tanto de la información molecular como de la ingeniería genética, en el paradigma del mejoramiento de plantas.

12.2. Principios y prácticas del mejoramiento molecular de plantas

Esquemas de mejoramiento y concepto de ganancia genética

Conceptualmente, el mejoramiento de plantas es simple: cruzar los mejores padres e identificar y recupera la progenie que supera a los padres. En la práctica, el mejoramiento consta de tres pasos, en el cual, poblaciones o colecciones de germoplasma con variación genética útil son creadas o acopladas, se identifican los individuos con fenotipos superiores y se desarrollan cultivos mejorados a partir de los individuos seleccionados. Se ha desarrollado una amplia gama de acercamientos para hacer mejoramiento de plantas, dependiendo de la especie a cultivar y los objetivos del mejoramiento (Fehr, 1987; Stoskopf et al., 1993). Estos métodos de mejoramiento difieren en los tipos de población, procedimientos de selección y el producto final.

En la figura x se muestra un resumen de los métodos de mejoramiento que comúnmente se emplean en programas de desarrollo de cultivos. Como ya se mencionó, cuando la meta es mejorar un genotipo elite establecido, con características controladas por uno o pocos loci, se utiliza la retrocruza para lograr la introgregación de un solo gen (Fig. 1A), o una pirámide de pocos genes (Fig. 1B). Para características genéticas complejas, el desarrollo de germoplasma mejorado requiere la reorganización del genoma para producir nuevas combinaciones de genes favorables en la progenie. El método de mejoramiento por pedigrí genera estas nuevas combinaciones mediante la cruce y recombinación entre padres superiores y complementarios, así como mediante la selección entre la progenie segregante para rendimiento mejorado (Fig. 1C).

La selección periódica pretende incrementar de manera simultanea las frecuencias de alelos favorables en loci múltiples de la población mejorada, a través de la cruce entre individuos seleccionados (Fig. 1D). Para cultivos híbridos como el maíz, la selección periódica podría extenderse para mejorar el rendimiento de poblaciones complementarias distintas (grupos heteróticos) que se utilizan como padres para generar combinaciones híbridas superiores. A esta práctica se le conoce como selección periódica recíproca.

Los principios de genética cuantitativa han sido particularmente poderosos como la base teórica tanto para el desarrollo de poblaciones, como para los métodos de selección y estabilización de genotipos deseables (Hallauer, 2007). Un concepto importante en la genética cuantitativa y el mejoramiento de plantas es la ganancia genética (DG), el cual se refiere al cambio predicho del valor promedio de una característica dentro de la población donde ocurre la selección. Sin tener en cuenta la especie, la característica de interés o el método de mejoramiento utilizado, el valor DG sirve como una expresión universal simple para la mejora genética esperada (Fehr, 1987; Falconer and Mackay, 1996). La figura 2 muestra la

ecuación de DG y una expansión de sus términos para parámetros fundamentales de la genética cuantitativa. Aunque el mejoramiento de plantas utiliza claramente una sobresimplificación de los principios de la genética cuantitativa, la ecuación de ganancia genética relaciona los cuatro factores centrales que influyen en el proceso de mejoramiento: el grado de variación fenotípica presente en la población (representado por su SD, sP), la probabilidad de que una característica fenotípica se transmita de los padres a la progenie (heredabilidad, h^2) la proporción de la población seleccionada como padres de la siguiente generación (intensidad de selección, i , expresada como unidades de SD del promedio) y el tiempo necesario para completar un ciclo de selección (L). L , no sólo es función del número de generaciones que se requieren para completar un ciclo de selección, sino también, de que tan rápido se complete una generación y el número de generaciones por año.

Es claro que DG se puede aumentar incrementando sP , h^2 o i , y disminuyendo L . Por lo tanto, la ecuación de ganancia genética provee un marco de comparación de la efectividad de estrategias particulares de mejoramiento, y comúnmente se utiliza como una guía para asignar juiciosamente los recursos necesarios para alcanzar los objetivos de mejoramiento. Cuando se considera dentro del contexto de ganancia genética, el mejoramiento molecular de plantas ofrece nuevos y poderosos acercamientos que permiten superar las limitaciones previas para maximizar DG. En las siguientes secciones se citan ejemplos en donde el mejoramiento molecular de plantas ejerce un impacto positivo en DG y cada una de las variables que la integran. Por cuestión de brevedad, nos enfocaremos en ejemplos de maíz, en donde el mejoramiento molecular está más avanzado y se ha convertido en uno de los primeros cultivos en el que se han desarrollado híbridos comerciales mejorados.

El mejoramiento molecular de plantas expande la diversidad genética útil para la mejora de cultivos.

El máximo potencial para la ganancia genética es proporcional a la variación fenotípica (sP) presente en la población fuente original, y se mantiene en los ciclos subsiguientes de selección. La variación fenotípica está asociada positivamente a la diversidad genética, aunque también depende de factores ambientales y de las interacciones entre genotipo y ambiente. La diversidad genética se puede derivar de poblaciones cultivadas (naturales o sintéticas), progenie segregante de la cruce de líneas parentales seleccionadas, materiales exóticos que no están adaptados al medio ambiente, cruces interespecíficas entre especies silvestres, mutaciones naturales o inducidas, introducción de eventos transgénicos, o combinaciones de estos.

Sin embargo, no todas las variaciones fenotípicas son iguales. Por ejemplo, el uso de germoplasma exótico ha sido extremadamente exitoso para mejorar varias especies de cultivo, pero puede haber dificultades si se introducen alelos no deseados, relacionados con pérdida de adaptación. La necesidad de diversidad genética debe ser balanceada mediante desempeño elite, puesto que es clave escoger los mejores padres, para maximizar la probabilidad de una mejora exitosa. En contraste, el incremento esperado en el desequilibrio por ligamento entre las poblaciones elite, derivado de intensas selecciones previas, también puede limitar la creación de nuevas combinaciones genéticas para ganancia futura. Con el entrecruzamiento de especies fuente, para lograr la recombinación genética se puede superar este problema, pero esto retrasa el desarrollo del cultivo.

Los marcadores moleculares y más recientemente, los esfuerzos de secuenciación del genoma con altos rendimientos, han incrementado dramáticamente el conocimiento y la habilidad de caracterizar la diversidad genética en el fondo común de germoplasma, para esencialmente cualquier especie de cultivo. Usando maíz como ejemplo, el reconocimiento de alelos de marcadores moleculares y la variación en la secuencia nucleotídica han proporcionado información básica sobre la diversidad genética antes y después de la domesticación de su antecesor silvestre, el teosinte, entre razas domesticadas distribuidas geográficamente y dentro de germoplasmas elite históricos (para una revisión, leer Cooper et al., 2004; Niebur et al., 2004; Buckler et al., 2006).

Esta información enriquece las investigaciones sobre evolución de plantas y genómica comparativa, contribuye a entender la estructura de la población, proporciona medidas empíricas de las respuestas genéticas hacia la selección y sirve también para identificar y mantener reservas de variabilidad genética para una futura minería de alelos benéficos (McCouch, 2004; Slade et al., 2005). Además, el conocimiento de las relaciones genéticas entre recursos de germoplasma puede ser una guía al escoger padres para la producción de híbridos o poblaciones mejoradas (e.g. Dudley et al., 1992; Collard and Mackill, 2008).

Mientras que los marcadores moleculares y otras aplicaciones genómicas han sido altamente exitosas para caracterizar la variación genética existente, la biotecnología de plantas genera nueva diversidad genética, que va más allá de las barreras sexuales entre especie (Gepts, 2002; Johnson and McCuddin, 2008). La biotecnología permite acceso a genes que no estaban disponibles por medio de cruza y crea un fondo de variación genética esencialmente ilimitado. Los genes pueden ser adquiridos de los genomas existentes en todos los reinos de vida, o pueden ser diseñados y

ensamblados en el laboratorio. Ejemplos tanto sutiles como extremos, sobre el poder de los transgenes para introducir variación fenotípica, se pueden encontrar en tres diferentes transgenes que desarrollan resistencia en maíz y otros cultivos, contra los herbicidas de glifosfato. En el primer híbrido de maíz tolerante al glifosfato, se utilizó una versión modificada del gen endógeno de maíz que codifica para la 5-enol-piruvilshikimato-3-fosfato sintasa (Spencer et al., 2000). A este híbrido le siguieron una serie de eventos producidos por el gen de la 5-enol-piruvilshikimato-3-fosfato sintetasa aislado de *Agrobacterium* (Behr et al., 2004). Más recientemente un gen sintético con la actividad de glifosfato acetil transferasa aumentada, fue creado utilizando reordenamiento de genes (gene shuffling) y selección en sistemas microbianos (Castle et al., 2004). Cada uno de estos casos de maíz tolerante al glifosfato, ilustra también otro beneficio de la biotecnología, donde nuevas combinaciones de secuencias regulatorias (por ejemplo los promotores del virus del mosaico del la coliflor, y el de actina 1 de arroz) pueden ser usadas para lograr la expresión óptima de una característica, con respecto a la actividad general y la distribución relativa de tejido, lo que podría ser imposible con los genes endógenos (Heck et al., 2005).

El mejoramiento molecular de plantas incrementa la acción favorable de los genes

La genética cuantitativa utiliza el concepto teórico de heredabilidad para cuantificar la proporción de la variación fenotípica que es controlada por el genotipo. En la práctica, la heredabilidad está considerablemente influenciada por la arquitectura genética de la característica de interés, la cual esta descrita por el número de genes, la magnitud de sus efectos y el tipo de acción del gen sobre el fenotipo. Un mejor conocimiento de la arquitectura genética y de la acción benéfica de la acción de los genes (la cual es más preferible para la selección) tiene un mayor impacto sobre la optimización de la ganancia genética. Para la fórmula de la ganancia genética, la heredabilidad (h^2) se utiliza en un sentido estricto, representando la proporción de la variación fenotípica debida a los efectos genéticos (aquellos que reflejan sustituciones de alelos o cambios en su proporción).

Los efectos genéticos aditivos también son considerados como valor de cultivo (valor de cría), puesto que su transferencia a la progenie es predecible. Las desviaciones de los efectos aditivos son significativas para muchas características, y están divididas en efectos dominantes que reflejan las interacciones entre los diferentes alelos en el mismo locus, y efectos epistáticos resultado de las interacciones entre los diferentes loci.

La acción de los genes y los valores de cultivo (valor de cría) se caracterizan mediante pruebas de progenie, en las cuales los fenotipos de

individuos de la población se compran con los padres y hermanos producidos por autopolinización o cruza. Los esfuerzos previos para desarrollar un gran número de marcadores moleculares, mapas genéticos de alta densidad y mapeo de poblaciones apropiadamente estructurado, han hecho rutina la habilidad de definir simultáneamente la acción del gen y su valor de cultivo en cientos y a veces miles de loci distribuidos relativamente uniforme entre genomas completos, en una gran variedad de cultivos. Los resultados de estos estudios de mapeo proporcionan estimados altamente mejorados del número de loci, efectos alélicos y acción del gen que controlan las características de interés. Y aún más importante, se pueden identificar fácilmente segmentos genómicos que muestren asociaciones estadísticamente significativas con características cuantitativas (loci de características cuantitativas, QTL por sus siglas en inglés). Además del mapeo genético en familias derivadas de cruza biparentales, nuevos avances en la genética de asociación con genes candidatos y acercamientos que combinan el análisis de desequilibrio por ligamiento en familias y poblaciones (Holland, 2007; Yu et al., 2008) en un futuro aumentaran el poder de los QTL.

La información sobre los QTL puede ser utilizada de diversas maneras para incrementar la heredabilidad y la acción benéfica de los genes.

Para los rasgos que tienen heredabilidad moderada o baja, como el rendimiento de grano, los QTL y otros marcadores moleculares asociados generalmente cuentan más para una mayor proporción de efectos genéticos aditivos, que el fenotipo solo. Además, el conocimiento de la arquitectura genética se puede explotar para añadir o eliminar alelos específicos que contribuyan al valor de cultivo. Cuando la epistasis o el ligamiento genético entre loci con efectos antagónicos en una característica limitan la ganancia genética, la información de los QTL puede ser usada para romper estas relaciones indeseables. El éxito para incrementar la ganancia genética utilizando la información sobre QTL depende en gran medida de la magnitud de los efectos de los QTL, la estimación precisa de su posición, la estabilidad de sus efectos en múltiples ambientes y si los QTL son robustos en el germoplasma mejorado. La predicción de la posición de los QTL se puede aumentar con mapeos finos, los cuales facilitan las pruebas de los efectos de QTL y los valores en poblaciones adicionales.

Cuando la densidad de las recombinaciones observadas se acerca a la resolución de genes aislados, el cambio genético causal del QTL puede ser determinado (para revisión, Salvi y Tuberosa, 2005; Yu and Buckler, 2006; Belo' et al., 2008; Harjes et al., 2008). El aislamiento molecular de estos QTL permite el desarrollo de marcadores moleculares perfectos o funcionales con el potencial de resolución de la unidad fundamental de la

herencia, el nucleótido, y aumenta dramáticamente la especificidad y precisión con las cuales los efectos genéticos se estiman y manipulan en los programas de mejoramiento.

El uso de transgenes podría simplificar la arquitectura genética de características deseables de formas superiores, o incluso que no son posibles cuando hay disponibles marcadores perfectos para QTL robustos de efecto amplio. Típicamente los transgenes condicionan efectos genéticos fuertes a particulares loci operacionales, los cuales también exhiben acción de genes dominantes en donde sólo se necesita una copia del evento para lograr la máxima expresión en un cultivar híbrido. Las características de los transgenes pueden reducir el complicado proceso de mejora cuantitativa a una solución directa y frecuentemente dramática. Un excelente ejemplo es la expresión de toxinas proteicas insecticidas de *Bacillus thuringiensis* (Bt) en híbridos transgénicos de maíz para reducir el daño por herbivoría provocado por *Ostrinia nubilalis* y *Diabrotica spp.*

La resistencia parcial de germoplasmas de maíz a estas plagas había sido caracterizada como una característica cuantitativamente heredable con poca heredabilidad (Papst et al., 2004; Tollefson, 2007), pero los eventos transgénicos con Bt ofrecen una alternativa heredable simple que es eficientemente manipulada en los programas de mejoramiento.

A través de la simplificación de la estructura génica, los transgenes también permiten la interrupción de las interacciones alélicas entre factores que controlan el rasgo de interés y otras características de rendimiento importantes. Por ejemplo, utilizando un recurso transgénico de resistencia a insectos (por ejemplo un solo locus del transgen Bt) se puede facilitar la selección de alelos favorables para mejorar el rendimiento que están reprimen fuertemente unidos a genes endógenos de resistencia a estas plagas. Además, los eventos transgénicos pueden ser diseñados para desacoplar efectos pleiotrópicos negativos en fenotipos benéficos condicionados por mutaciones recesivas. Esta aplicación se puede ilustrar con el uso de RNA interferente para disminuir específicamente la expresión del gen de almacenamiento de zeína en semilla (Segal et al., 2003; Huang et al., 2004). Esta estrategia imita los efectos de la mutación opaque2, mejorando el perfil de aminoácidos en granos de maíz para alimentación animal, a la vez que evita la textura suave del endospermo y la susceptibilidad a hongos patógenos típicamente asociados con opaque2.

Los eventos transgénicos también pueden ser diseñados para intervenir en pasos regulatorios clave para una ruta metabólica completa o para rutas de desarrollo, tales como aquellos genes de acción para el rasgo

correspondiente que son heredados en factores dominantes únicos que son menos sensible a los efectos ambientales.

Algunos ejemplos incluyen la expresión de un factor de transcripción que incrementa la tolerancia a la sequía (Nelson et al., 2007) y la alteración del balance entre los niveles del factor de transcripción GLOSS15 y su represor, el micro RNA 172, para retardar la floración en híbridos de maíz (Lauter et al., 2005).

La biotecnología también facilita el agrupamiento de transgenes que controlan un rasgo, o serie de rasgos, en un haplotipo de un solo locus definido por un evento transgénico.

Ejemplos de esta estrategia incluyen, el Arroz Dorado inicial (Ye et al., 2000), el reciente lanzamiento de híbridos de maíz triplemente trasgénico Yield-Guard VT, en el cual, las características de tolerancia a herbicidas y resistencia a múltiples insectos se integran en un solo locus genómico (<http://www.yieldgardvt.com/VTScience/Default.aspx>), o la combinación de transgenes que simultáneamente incrementan la síntesis y disminuyen el catabolismo de lisina en semillas de maíz (Frizzi et al., 2008). Reportes recientes sobre mejoras en la tecnología de marcaje de genes (Ow, 2007) y la construcción de minicromosomas de plantas meióticamente transmisibles (Carlson et al., 2007; Yu et al., 2007) preparan el terreno para introducir más características con mayor complejidad. Con estos logros la biotecnología es ahora en posición de ensamblar diversidad genética útil, desde esencialmente cualquier recurso, en construcciones que concentren acción genética favorable y maximice la heredabilidad para un conjunto de características muy amplio.

Para cerrar esta sección sobre cómo el mejoramiento molecular de plantas incrementa la acción benéfica del gen, es importante enfatizar que los estudios de QTL, cuando se llevan a cabo con la escala y posición adecuadas para identificar genes causales, representan una estrategia genómica funcional muy poderosa.

La clonación de QTL ha producido nuevas perspectivas acerca de la biología de las características cuantitativas, que no hubieran sido descubiertas desde el análisis de genes mediante estrategias de pérdida de función, o sobre-expresión, en particulares impacto de la variación regulatoria en la variación fenotípica y en la evolución (e.g. Cong et al., 2002; Clark et al., 2006; Yan et al., 2004; Salvi et al., 2007).

Por otra parte, los marcadores moleculares, la genómica y la biotecnología se aplican de una red de trabajo iterativa, explotando así la diversidad

genética para la mejora de cultivos. La información genética permite el descubrimiento de alelos benéficos mediante el mapeo y clonación de QTL, seguido del uso de la información aprendida sobre caracterización molecular de QTL, para diseñar estrategias transgénicas óptimas para el mejoramiento de plantas.

El mejoramiento molecular de plantas incrementa la eficiencia de la selección

El mejoramiento convencional de plantas, que se basa sólo en la selección de fenotipos, ha sido históricamente efectiva. Sin embargo, para algunas características, la selección fenotípica ha logrado poco progreso debido a los retos al medir fenotipos o identificar individuos con el mayor valor de mejoramiento. Los efectos de ambiente, interacción ambiente- genotipo y errores de medición, también contribuyen a las diferencias observadas. La evaluación de genotipos en ambientes múltiples con réplicas permite una mejor estimación del valor de mejoramiento, pero requiere gasto y tiempo adicionales. Para algunos rasgos, podría ser necesario sacrificar individuos para medir el fenotipo, o la expresión de un rasgo puede depender de condiciones variables del ambiente (por ejemplo, presión por enfermedades) y el estado de desarrollo (la calidad del grano solo se puede conocer después de la floración). Además los mejoradores típicamente den mejorar de manera simultánea una serie de características valoradas comercialmente, lo que podría limitar la cantidad de grano para seleccionar. Así como el mejoramiento molecular ayuda a expandir la diversidad genética, caracterizar la estructura genética y modificar la acción del gen, sus métodos también pueden ser aplicados para incrementar la eficiencia de la selección.

Una gran parte de la literatura ha considerado la utilidad de la selección asistida por marcadores moleculares y su ajuste con los diferentes métodos de mejoramiento (Fig. 1) tupiendo al lector a un gran número de revisiones excelentes del tema (Dekkers and Hospital, 2002; Holland, 2004; Johnson, 2004; Varshney et al., 2006; Collard and Mackill, 2008). Los genotipos de marcadores moleculares que se utilizan dentro de genes o muy cercanos a los QTL, que influyen rasgos bajo selección, pueden ser empleados como un suplemento a las observaciones fenotípicas en una guía de selección (Lande and Thompson, 1990). En los casos donde las correlaciones genéticas son altas, se pueden ganar eficiencias adicionales sustituyendo la selección genotípica por la fenotípica, durante algunos ciclos de selección, lo que puede reducir los esfuerzos de fenotipaje, así como el número de ciclos que permitan el uso de viveros fuera de temporada. Johnson (2004) resume un ejemplo anticipado de combinar los datos fenotípicos y las puntuaciones de marcadores moleculares para incrementar las ganancias en rendimiento de maíz y resistencia a plagas debida a la selección. Una estrategia efectiva para modificar

simultáneamente características múltiples, es el uso de índices de selección que consideren múltiples factores para escoger el fenotipo mejorado final. Recientemente Eathington et al. (2007) reportaron su resultados sobre el uso de índices de múltiples características y la selección asistida por marcadores en cerca de 250 poblaciones únicas de maíz. El uso de marcadores moleculares incrementan la eficiencia de la mejora alrededor de 2 veces en relación a la selección fenotípica, también con ganancias similares en poblaciones de soya (*Glycine max*) y girasol (*Helianthus annuus*).

La selección asistida por marcadores también puede aumentar significativamente la ganancia genética para características en las que el fenotipo es difícil de evaluar ya sea por su costo, o por su dependencia e condiciones ambientales específicas. Los marcadores moleculares se pueden utilizar para fomentar la probabilidad de identificar realmente genotipos superiores, concentrándose en probar recursos de genotipos con el mayor potencial (i.e. con la eliminación temprana de genotipos inferiores), disminuyendo el número de la progenie que se necesita analizar para recuperar un nivel dado de ganancia, y permitiendo la mejora simultánea de características que están negativamente correlacionadas (Knapp, 1998). Algunos ejemplos exitosos incluyen resistencia de la soya a nematodos (Young, 1999), resistencia a enfermedades en cereales (revisado por Varshney et al., 2006) y tolerancia a sequía en maíz (Ribaut and Ragot, 2007; Tuberosa et al., 2007). La eficiencia de la selección fenotípica para lagunas características complejas se puede incrementar si se incluyen fenotipos bioquímicos o fisiológicos como características secundarias, cuando estos exhiben correlaciones genéticas fuertes con la característica de interés y alta heredabilidad. Los recientes avances en genómica funcional permiten escalar a nivel de población el perfil de abundancia de RNA, niveles y actividades de proteínas y metabolitos que están asociados con características importantes. A parte de los marcadores moleculares que identifican variaciones en la secuencia de ADN, estas estrategias genómicas podrían proveer fenotipos secundarios adicionales como objetivos de selección (Jansen and Nap, 2001; Johnson, 2004), particularmente para aquellas características definidas por su respuesta a factores como medio, estado fisiológico o desarrollo.

La selección asistida por marcadores, también acelera el uso de transgenes en cultivares comerciales. Típicamente, esto se ha logrado mediante la retrocruza asistida por marcadores. Sin embargo, para mejoras futuras en biotecnología, como tolerancia a la sequía o a la limitación de nutrientes, el mejoramiento directo podría requerir de la co-optimización de expresión del transgen y el origen genético, puesto que los genes endógenos y los

factores ambientales, podrían tener el potencial de influir en los fenotipos resultantes de las modificaciones transgénicas (Mumm, 2007).

Por supuesto, que el uso de marcadores moleculares, podría también contribuir con los esfuerzos de mejoramiento directo. Alternativamente los esfuerzos por descubrir genes adicionales o QTL, que son necesarios para el desempeño fiable de una característica, podrían proponer el diseño de nuevas construcciones de transgenes que agrupen transgenes primarios con modificadores genéticos, en una segunda generación de eventos transgénicos.

Incremento en la aprobación del mejoramiento molecular de plantas

La aceptación de las estrategias de mejoramiento molecular de plantas ha ocurrido a diferentes ritmos, dentro de las especies y las instituciones comprometidas con el mejoramiento de estos, debido a la influencia combinada de factores científicos, económicos y sociológicos. Algunas barreras científicas importantes, incluyendo la recalcitrancia de los cereales a la transformación mediada por *Agrobacterium* y la falta de conocimientos sobre el control genético, son definidas como objetivos importantes del mejoramiento. El desarrollo continuo de investigación y tecnología ha superado los obstáculos presentes en la transformación de plantas, en casi todos los cultivos y especies de horticultura de importancia económica (Wenzel, 2006). De manera similar la información obtenida de la investigación genómica en plantas y otros organismos ha generado bastante información sobre estructura y función génica, así como grandes cantidades de marcadores moleculares para ser usados en el mejoramiento de plantas.

A pesar de estos recursos, la genética específicamente de QTL robustos permanece ambigua, amenos que los programas de mejoramiento y los sistemas de manejo de la información asociada, sean reestructurados para integrar por completo el conocimiento sobre pedigríes, fenotipos, y marcadores de genotipos que puedan ser influenciados para optimizar la respuesta a la selección (Cooper et al., 2004; Eathington et al., 2007). Aún con una integración de este tipo, modificar las funciones regulatorias sigue siendo un reto científico de mejoramiento molecular, puesto que es difícil determinar la secuencia de bases responsable de los cambios regulatorios y predecir los efectos fenotípicos (Morgante and Salamini, 2003).

Una vez que las tecnologías en biotecnología y genómica estén disponibles, los factores económicos generalmente dictan el grado en que estas innovaciones se integren a los programas existentes de mejora de plantas. El costo de ganar la aprobación regulatoria gubernamental, para la liberación comercial de variedades transgénicas (recientemente estimadas por Kalaitzandonakes et al., 2007, entre \$7-\$10 millones) es una barrera

económica significativa. El costo asociado con el desarrollo, establecimiento y operación del mejoramiento molecular de plantases mayor que para las prácticas convencionales de mejoramiento (Koeberner and Summers, 2003; Morris et al., 2003), pues requiere inversiones importantes en nueva infraestructura de investigación y capacidad intelectual. Estos recursos sólo existen de forma inicial en compañías agrícolas privadas y en un puñado de instituciones públicas, promoviendo la aceleración con tendencia a incrementar la industrialización de los programas de mejoramiento en cultivos mayores como maíz, soya, algodón (*Gossypium hirsutum*) y trigo (Johnson, 2007). Cuando ha habido aceptación por las compañías, la balanza favorece a la biotecnología sobre los QTL, para mejorar características complejas, a pesar de aumentar los costos de desarrollo del producto, ya que los transgénicos pueden ser diseñados para producir efectos fenotípicos más fuertes, incluso dramáticos y generalmente se pueden desplegar más rápido en un amplio rango del germoplasma, resultando en nuevas soluciones aún más rápido.

A pesar de que actualmente el mejoramiento molecular es considerado por las grandes compañías, como un componente esencial en los esfuerzos de mejoramiento de cultivos mayores, la amplia aplicación de las estrategias moleculares modernas a las técnicas de mejoramiento convencionales, sigue siendo un tema de debate entre algunos mejoradores de plantas que participan en el sector público, particularmente, para cultivos menores (e.g. Gepts, 2002; Goodman, 2004). Además de los muy válidos factores científicos y económicos, que han retrasado o prevenido, la aceptación de las estrategias moleculares para alcanzar algunos objetivos del mejoramiento molecular, al menos hay tres razones adicionales que contribuyen a esta visión. En primera, el mejoramiento molecular de plantas requiere entrenamiento y experiencia en biología molecular y mejoramiento de plantas. Los esfuerzos educativos que otorgan este tipo de entrenamiento interdisciplinario fueron establecidos a inicios de 1990, pero continúan limitados a un pequeño grupo de instituciones académicas con tradición en mejoramiento de plantas (Guner and Wehner, 2003; Gepts and Hancock, 2006; Guimarañes y Kueneman, 2006). Una segunda razón que reduce el entusiasmo para que la biotecnología sea abrazada por los mejoradores, son los problemas de aceptación de cultivos transgénicos entre algunos gobiernos y grupos de consumidores, como ha sucedido al posponer variedades de trigo con transgenes que brindan resistencia hacia los herbicidas con glifosfato (Sokstad, 2004). Finalmente, la exaltación sobre el potencial del mejoramiento molecular de plantas, también estimula cambios en el financiamiento de instituciones públicas para incrementar la capacidad intelectual y la infraestructura de las investigaciones en genómica y genética molecular, lo cual ocurre, irónicamente, a expensas de el mejoramiento convencional (Knight, 2003).

Este énfasis puede ser temporalmente necesario para establecer los fundamentos de la biología de plantas de siglo 21, pero actualmente, hay un reconocimiento creciente de que el incremento en la inversión en capacidad de mejoramiento de plantas y en investigación de transferencia, uniando los métodos moleculares con los objetivos de mejoramiento, necesita del desarrollo del potencial de los avances recientes en biotecnología y genómica (Guimaraes and Kueneman, 2006; National Research Council, 2008).

Perspectivas

A pesar de los recientes avances y ejemplos exitosos del mejoramiento molecular de plantas, uno de los grandes retos actuales en la biología de plantas es identificar aquellas combinaciones de genes que llevan a una mejora de cultivos significativa. La estrategia más efectiva para acelerar estos esfuerzos, es integrar mejor las diferentes disciplinas de investigación y las actividades centrales que conforman el mejoramiento molecular de plantas. Como aquí se describe, y anteriormente por otros (Gepts y Hancock, 2006; Bliss, 2007), esta integración requiere el conocimiento de la organización genómica y de la función de los genes, fundamentos sólidos en estrategias estadísticas par estimar los efectos genéticos, bases fuertes sobre biología de plantas, experiencia con los métodos de laboratorio sobre biología molecular y genómica funcional, así como en práctica en campo sobre mejoramiento, y la habilidad de manejar grandes conjuntos de datos de diversos tipos. Aunque se reconoce, con conciencia y apreciación la importancia de cada una de estas disciplinas por todos los científicos de plantas la formación de estudiantes, y los esfuerzos de entrenamiento, los fondos de financiamiento y los programas de investigación, todavía enfatizan en subgrupos específicos del paradigma de mejoramiento molecular de plantas. Lo anterior se justifica dada la amplitud de las disciplinas involucradas, sin embargo, se necesita hacer mayores esfuerzos para implementar programas educativos y de investigación que promuevan la participación activa en el mejoramiento molecular de plantas para la mejora de cultivos. Este tipo de programas debería ser desarrollado fomentando las colaboraciones entre grupos con experiencia complementaria. Absolutamente, las agencias de financiamiento deben expandir su portafolio de proyectos que da apoyo para la transferencia tecnológica mediante requerimiento de propuestas que integren esfuerzos en investigación básica y también en productos tangibles de mejoramiento de plantas. Algunos ejemplos dignos de imitarse incluyen el programa Harvest-Plus (<http://www.harvestplus.org/>), el MASwheat (Dubcovsky, 2004), y el U.S. Department of Agriculture Coordinated Agricultural Projects. El sector privado también debe seguir con inversiones que estimulen la integración y brinde el ambiente de entrenamiento apropiado para los futuros científicos que se integren a la fuerza de trabajo del mejoramiento molecular. Además de apoyo directo para entrenamiento de graduados y el patrocinio de investigación, las compañías pueden proporcionar apoyo "en especie" que ayude a reducir la brecha en tecnología en expansión entre el sector de investigación en mejoramiento molecular de plantas público y privado. Con los esfuerzos colectivos de la amplia comunidad de científicos comprometidos con la biología de plantas y la mejora de cultivos, el mejoramiento molecular de plantas podrá expandir su impacto y contribuciones que dan respuesta a las necesidades de incremento sustentable de la producción agrícola.

12.3. Figuras de marcadores y QTLs

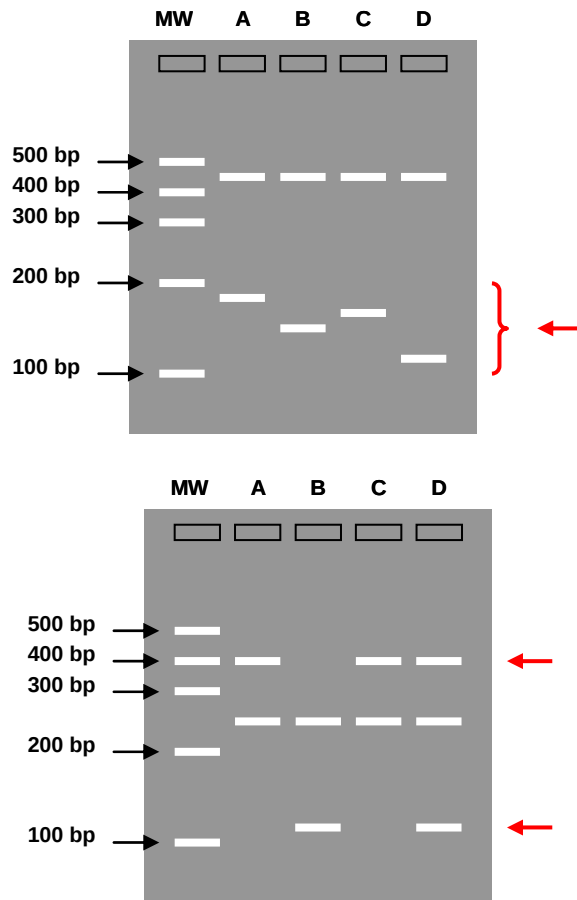


Figura 12-1. Diagram representing hypothetical DNA markers between genotypes A, B, C and D. Polymorphic markers are indicated by arrows. Markers that do not discriminate between genotypes are called monomorphic markers. (a) Example of SSR markers. The polymorphic marker reveals size differences for the marker alleles of the four genotypes, and represent a single genetic locus. (b) Examples of markers generated by the RAPD technique. Note that these markers are either present or absent. Often, the sizes of these markers in nucleotide base pairs (bp) are also provided; these sizes are estimated from a molecular weight (MW) DNA ladder. For both polymorphic markers, there are only two different marker alleles.

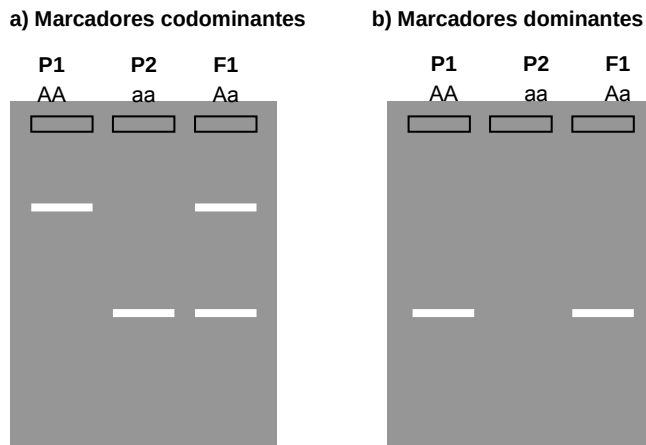


Figura 12-2. Comparison between (a) codominant and (b) dominant markers. Codominant markers can clearly discriminate between homozygotes and heterozygotes whereas dominant markers do not. Genotypes at two marker loci (A and B) are indicated below the gel diagrams.

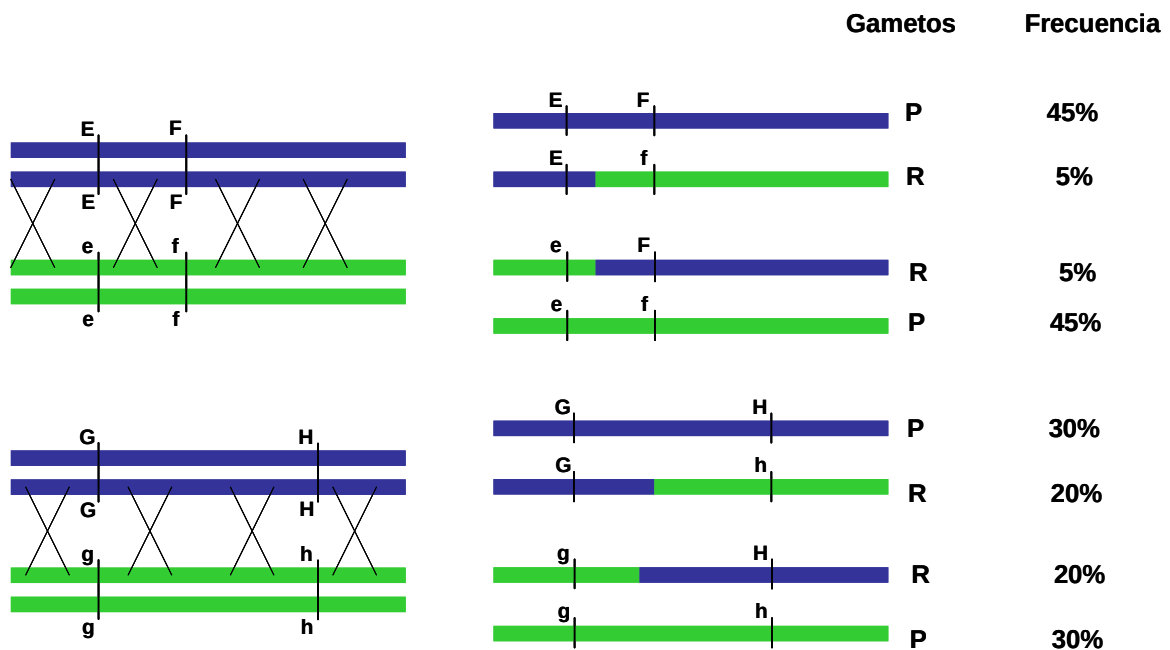


Figura 12-3. Diagram indicating cross-over or recombination events between homologous chromosomes that occur during meiosis. Gametes that are produced after meiosis are either parental (P) or recombinant (R). The smaller the distance between two markers, the smaller the chance of recombination occurring between the two markers. Therefore, recombination between markers G and H should occur more frequently than recombination between markers E and F. This can be observed in a segregating mapping population. By analysing the number of recombinants in a population, it could be determined that markers E and F are closer together compared to G and H.

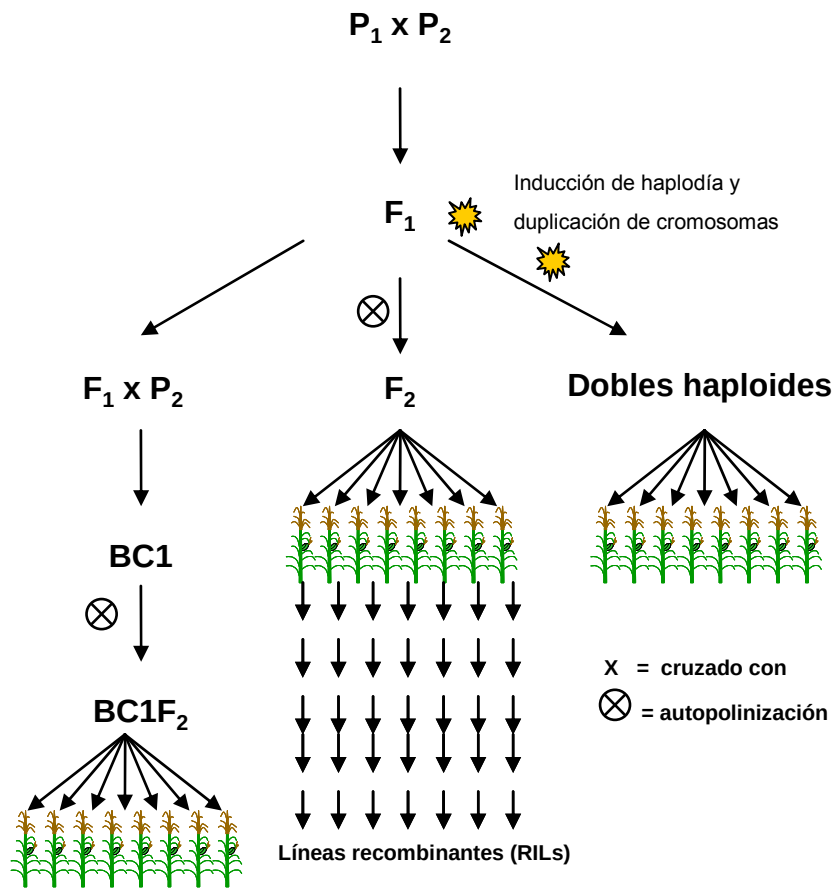


Figura 12-4. Diagram of main types of mapping populations for self-pollinating species.

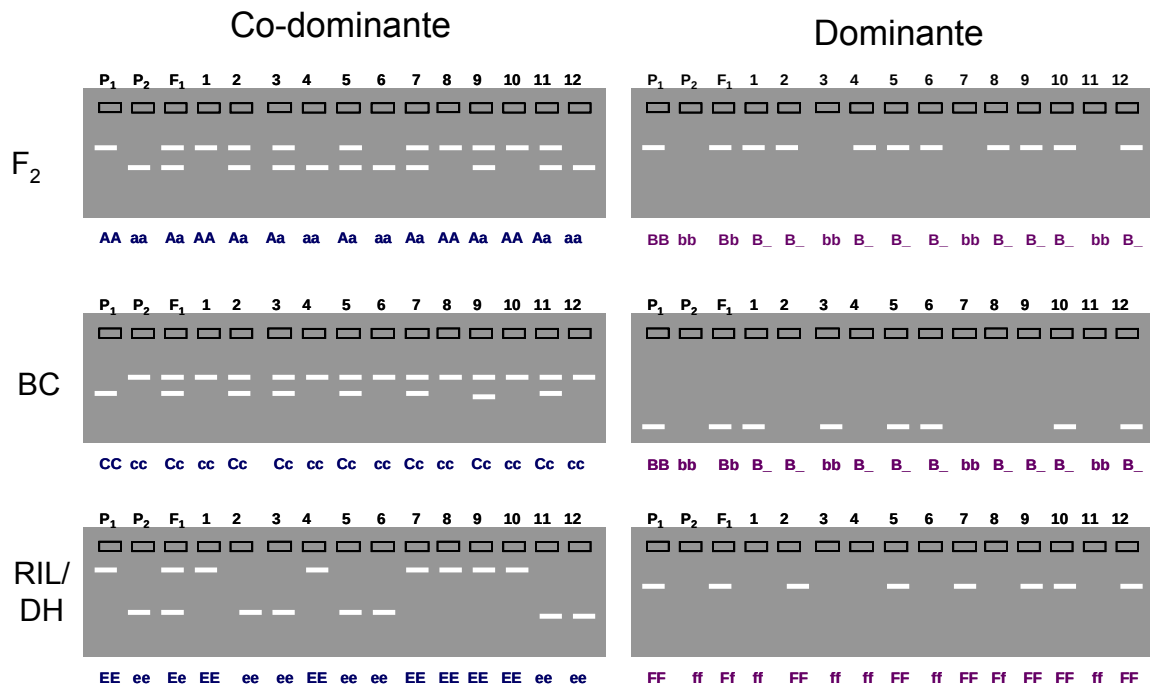


Figura 12-5. Hypothetical gel photos representing segregating codominant markers (left-hand side) and dominant markers (right-hand side) for typical mapping populations. Codominant markers indicate the complete genotype of a plant. Note that dominant markers cannot discriminate between heterozygotes and one homozygote genotype in F₂ populations. The segregation ratios of markers can be easily understood by using Punnett squares to derive population genotypes.

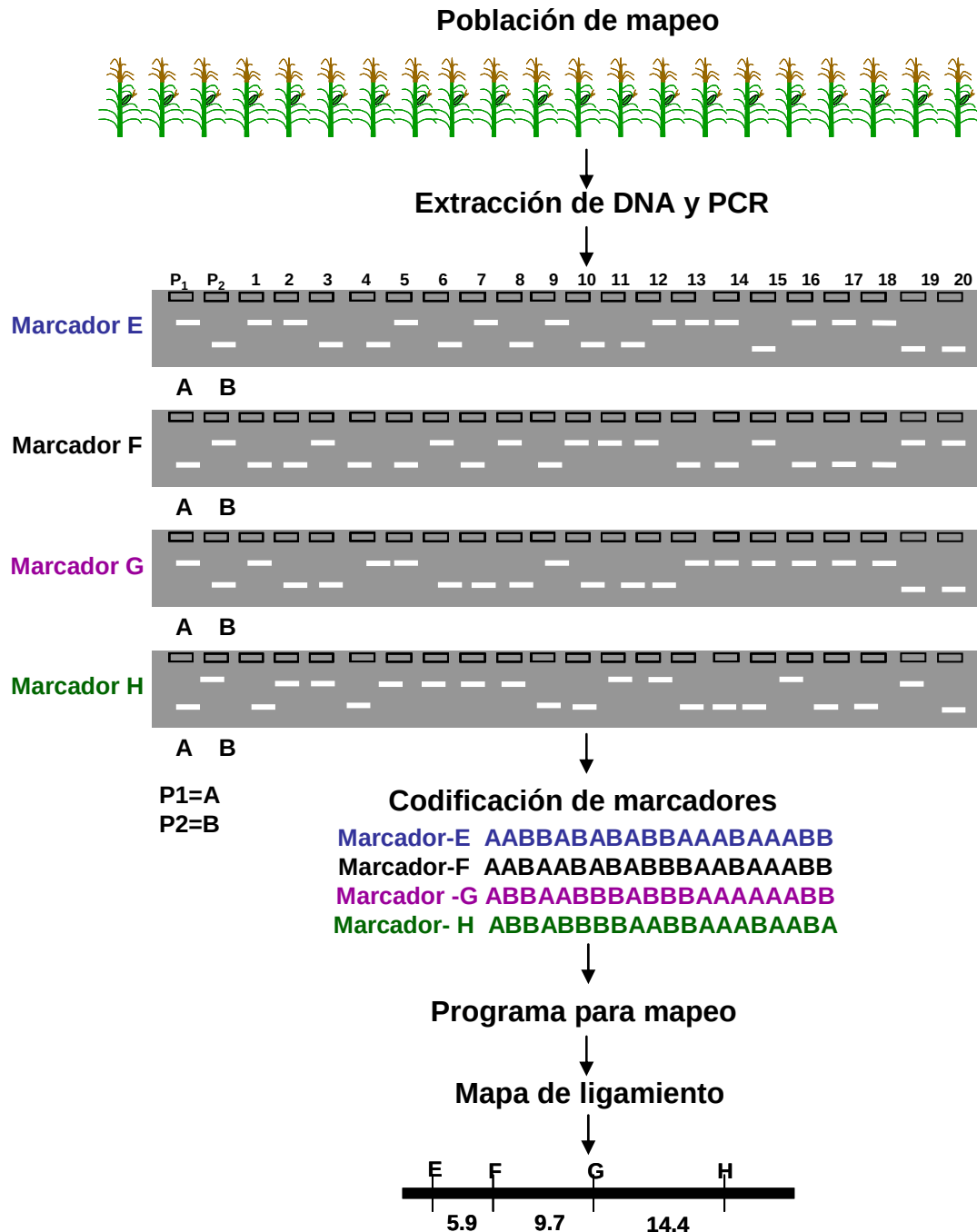


Figura 12-6. Construction of a linkage map based on a small recombinant inbred population (20 individuals). The first parent (P1) is scored as an ‘A’ whereas the second parent (P2) is scored as a ‘B’. Coding of marker data varies depending on the type of population used. This linkage map was constructed using Map Manager QTX (Manly et al., 2001) using the Haldane mapping function.

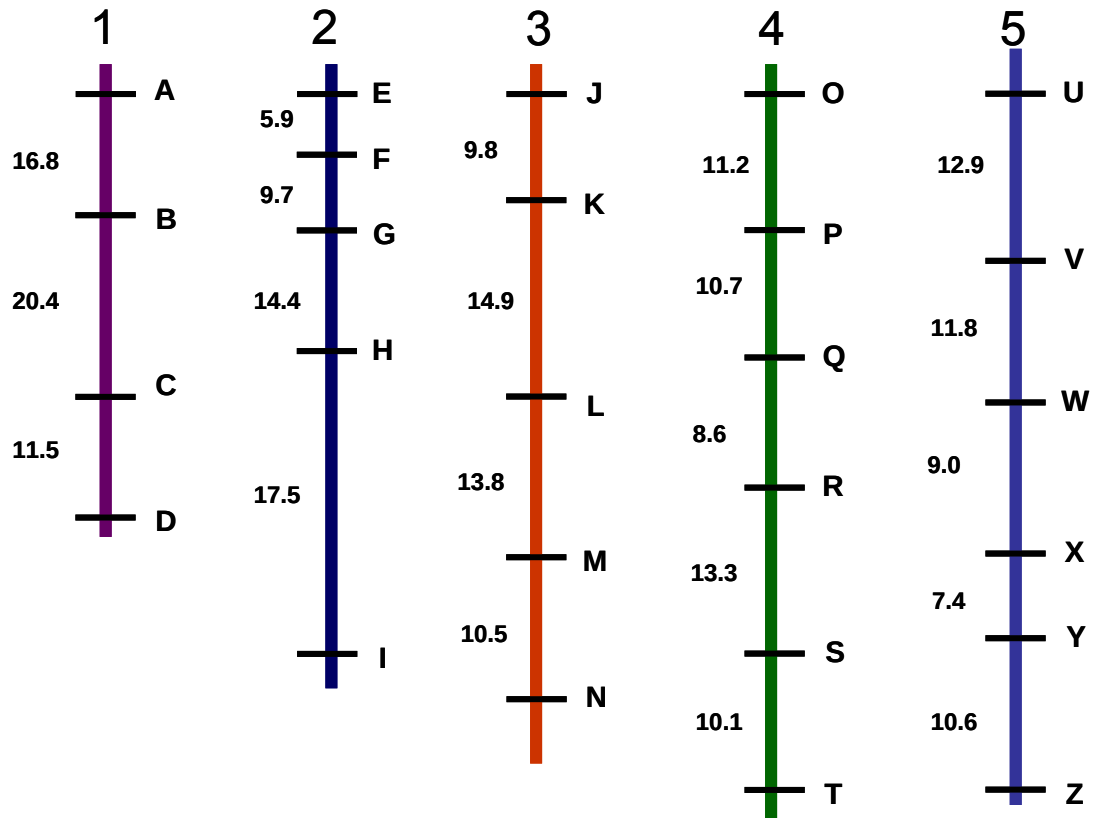


Figure 12-7. Hypothetical ‘framework’ linkage map of five chromosomes (represented by linkage groups) and 26 markers. Ideally, a framework map should consist of evenly spaced markers for subsequent QTL analysis. If possible, the framework map should also consist of anchor markers that are present in several maps, so that they can be used to compare regions between maps.

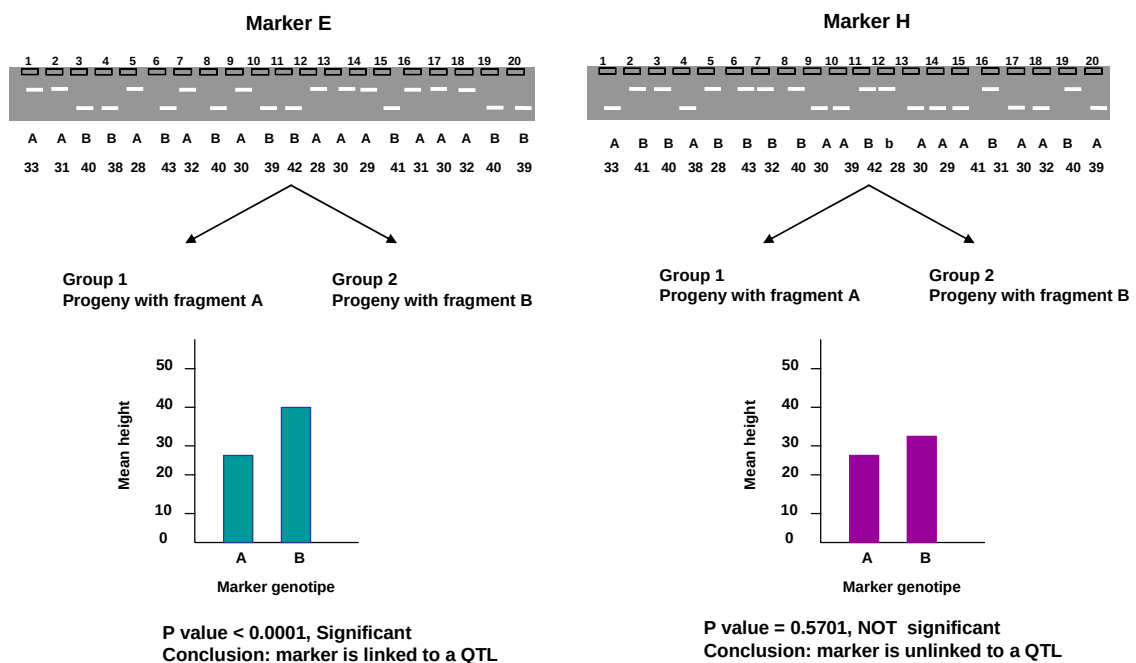


Figure 12-8

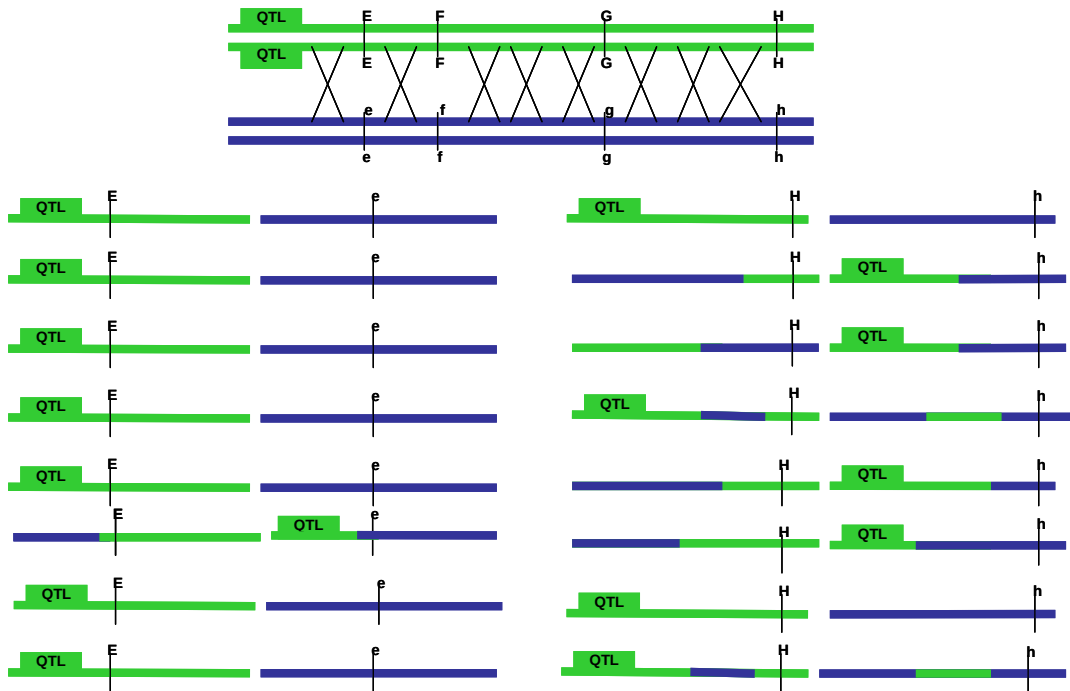


Figura 12-9

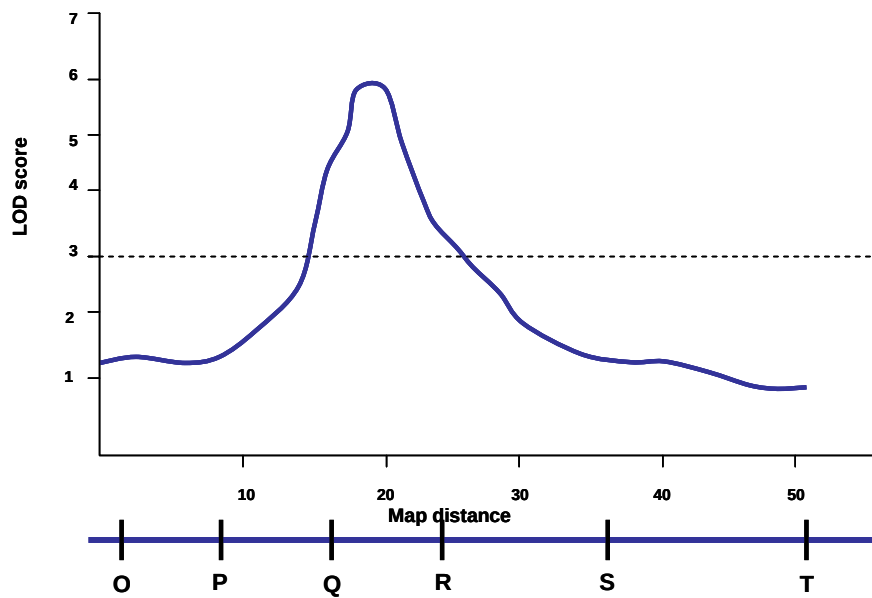


Figura 12-10

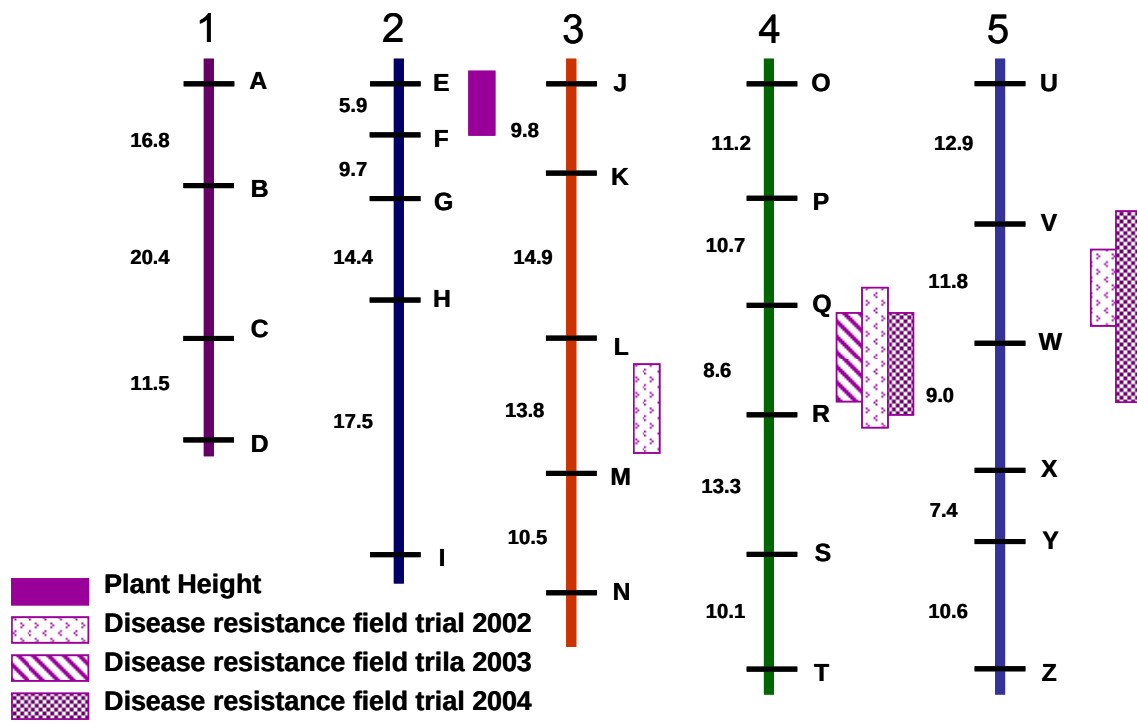


Figura 12-11

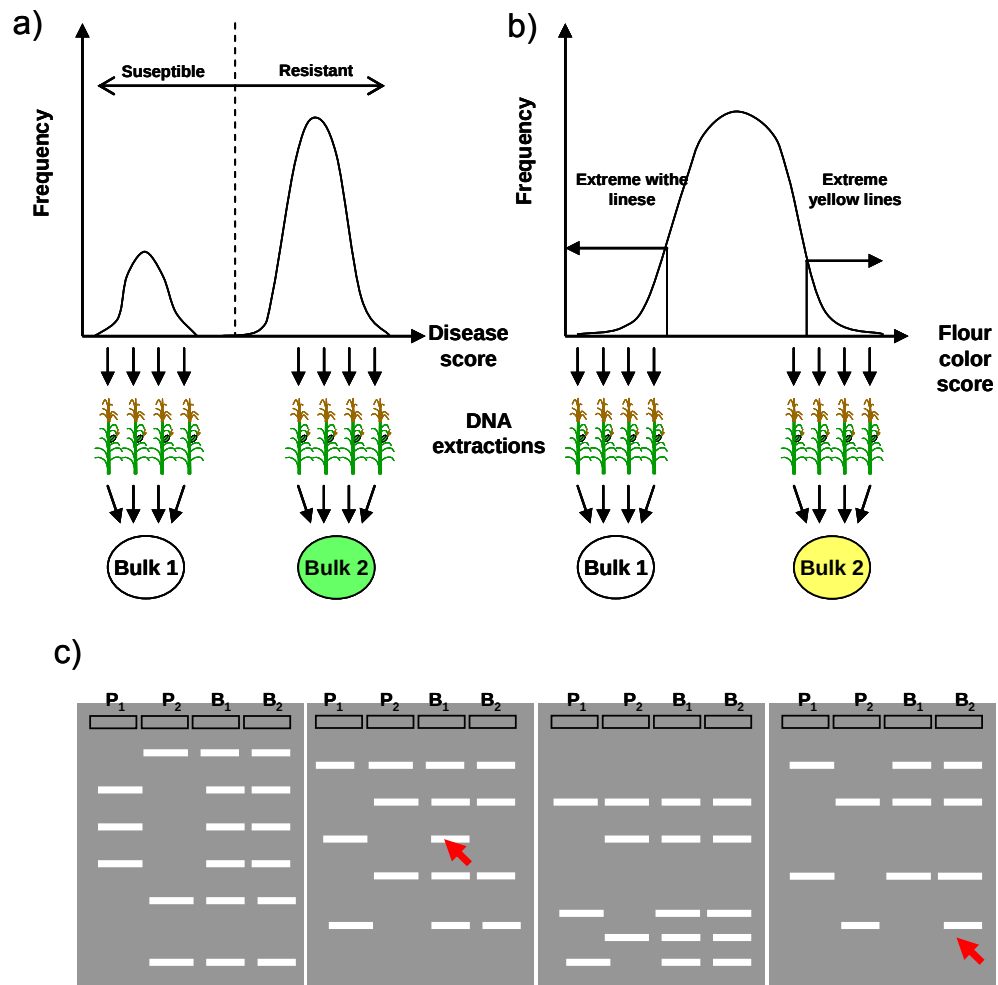


Figura 12-12

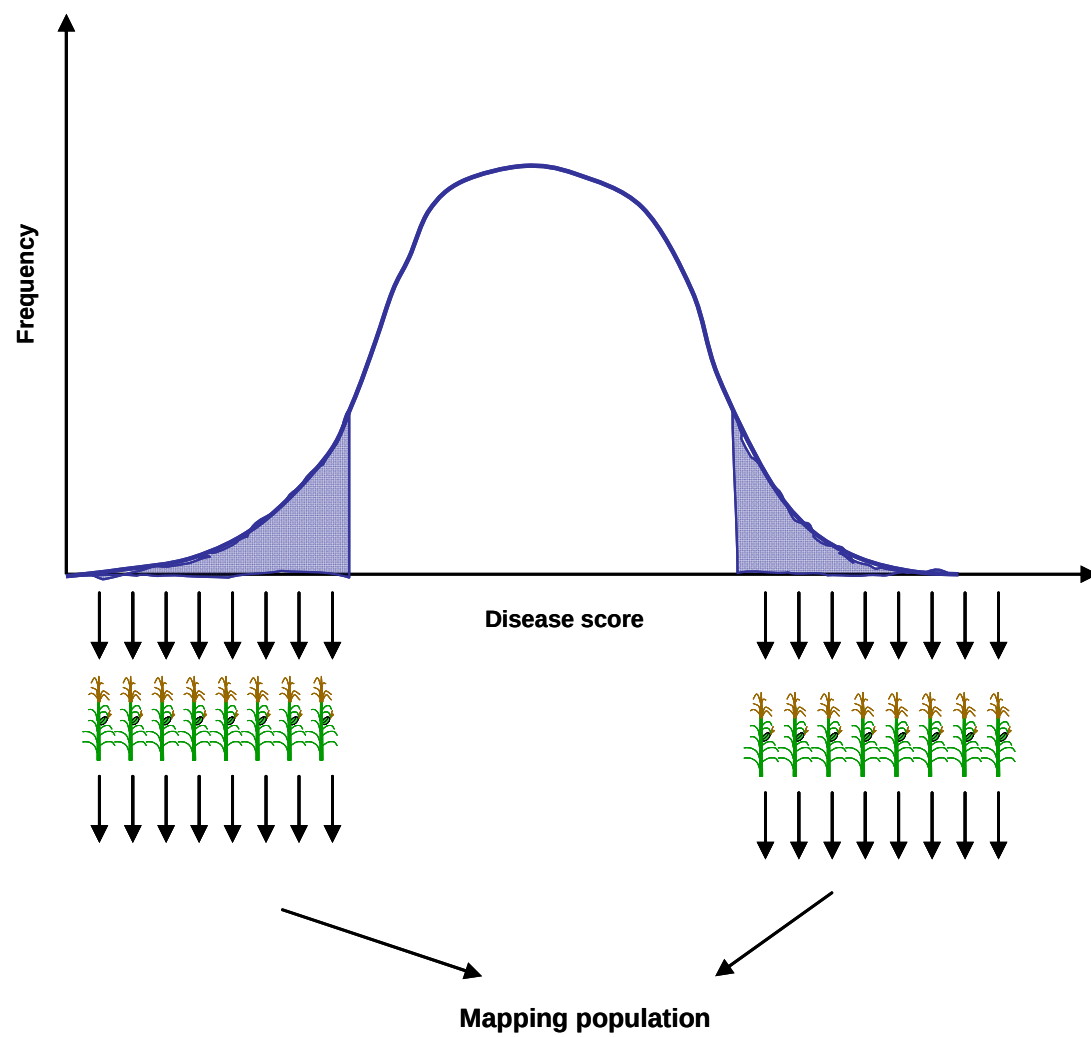


Figura 12-13

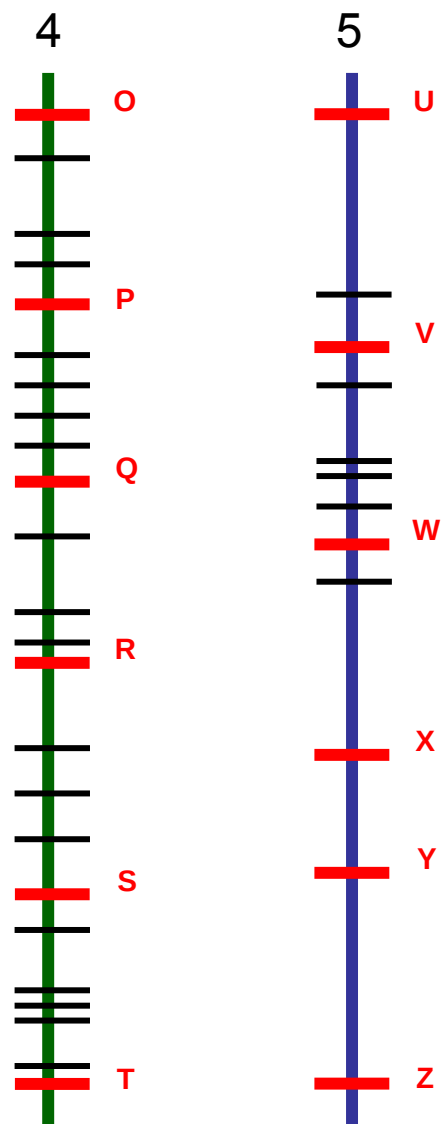
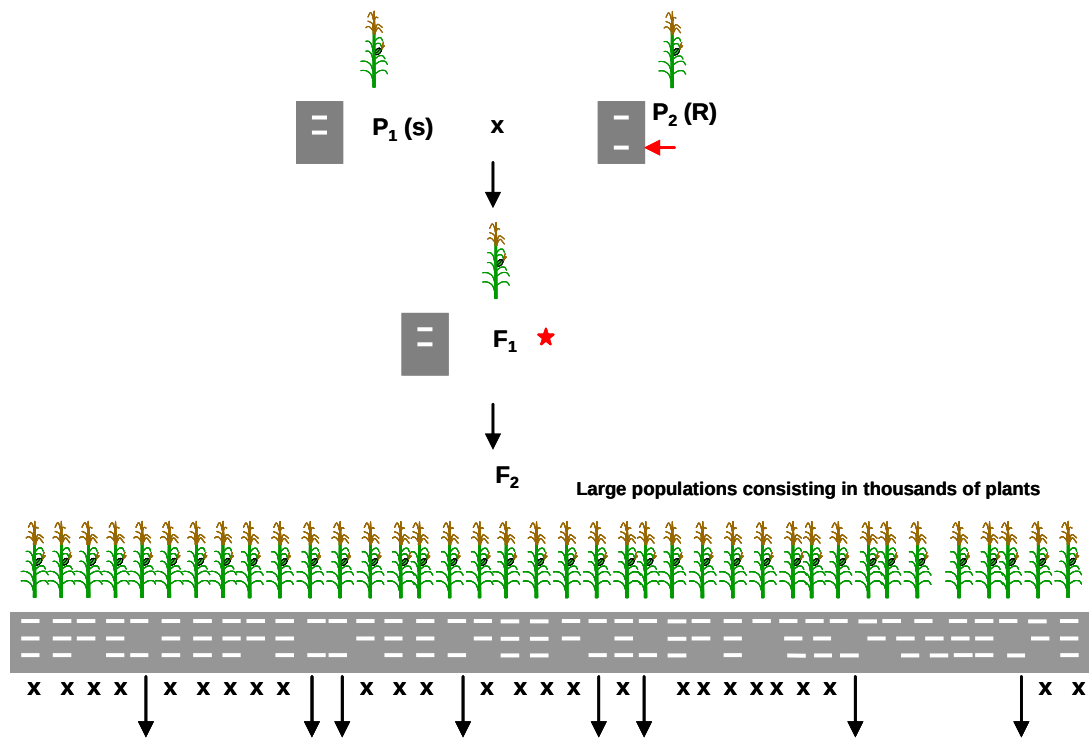


Figura 12-14



Early –generation marker assisted selection

Figura 12-15

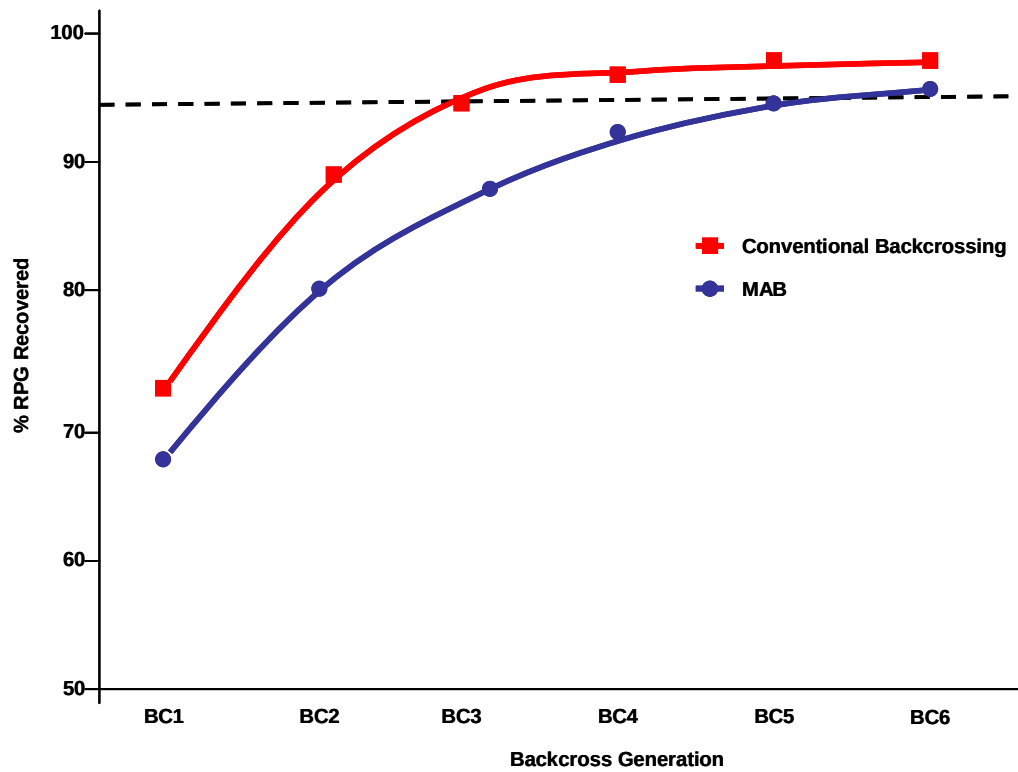


Figura 12-16

12.3.1. Referencias

Tarea 12-1

Escoja 5 artículos de la siguiente lista, y haga una tabla de las metodologías que utilizaron en cada uno de ellos. Haga también un resumen de los resultados principales que se presentan en las figuras y tablas de esos artículos.

- Ablett, G.A., A. Karakousis, L. Banbury, M. Cakir, T.A. Holton, P. Langridge & R.J. Henry, 2003. Application of SSR markers in the construction of Australian barley genetic maps. *Aust J Agric Res* 54: 1187–1195.
- Anderson, J., G. Churchill, J. Autrique, S. Tanksley & M. Sorrells, 1993. Optimizing parental selection for genetic linkage maps. *Genome* 36: 181–186.
- Asins, M., 2002. Present and future of quantitative trait locus analysis in plant breeding. *Plant Breed* 121: 281–291.
- Baird, V., A. Abbott, R. Ballard, B. Sosinski & S. Rajapakse, 1997. DNA Diagnostics in Horticulture, p. 111–130 In: P. Gresshoff (Ed.), *Current Topics in Plant Molecular Biology: Technology Transfer of Plant Biotechnology*. CRC Press, Boca Raton.
- Baird, W.V., R.E. Ballard, S. Rajapakse & A.G. Abbott, 1996. Progress in Prunus mapping and application of molecular markers to germplasm improvement. *HortScience* 31: 1099–1106.
- Barone, A., 2004. Molecular marker-assisted selection for potato breeding. *Am J Potato Res* 81: 111–117.
- Barrett, B., A. Griffiths, M. Schreiber, N. Ellison, C. Mercer, J. Bouton, B. Ong, J. Forster, T. Sawbridge, G. Spangenburg, G. Bryan & D. Woodfield, 2004. A microsatellite map of white clover. *Theor Appl Genet* 109: 596–608.
- Basten, C.J., B.S. Weir & Z.-B. Zeng, 1994. Zmap-a QTL cartographer. In: J.S.G.C. Smith, B.J. Benkel, W.F. Chesnais, J.P. Gibson, B.W. Kennedy & E.B. Burnside (Eds.), *Proceedings of the 5th World Congress on Genetics Applied to Livestock Production: Computing Strategies and Software*, Guelph, Ontario, Canada. Published by the Organizing Committee, 5th World Congress on Genetics Applied to Livestock Production.
- Basten, C., B. Weir, Z.-B. Zeng, 2001. QTL Cartographer. Department of Statistics, North Carolina State University, Raleigh, NC.
- Beattie, A., J. Larsen, T. Michaels & K. Pauls, 2003. Mapping quantitative trait loci for a common bean (*Phaseolus vulgaris* L.) ideotype. *Genome* 46: 411–422.
- Beavis, W., 1998. QTL Analyses: Power, Precision and Accuracy. In: A.H. Paterson (Ed.), *Molecular Dissection of Complex Traits*. CRC Press, Boca Raton.
- Beckmann, J. & M. Soller, 1986. Restriction fragment length polymorphisms in plant genetic improvement. *Oxford Surveys of Plant Mol Biol Cell Biol* 3: 197–250.
- Bernacchi, D., T. Beck-Bunn, Y. Eshed, S. Inai, J. Lopez, V. Petiard, H. Sayama, J. Uhlig, D. Zamir & S. Tanksley, 1998. Advanced backcross QTL analysis of tomato. II. Evaluation of near-isogenic lines carrying single-donor introgressions for desirable wild QTL-alleles derived from *Lycopersicon hirsutum* and *L. pimpinellifolium*. *Theor Appl Genet* 97: 170–180.
- Blair, M., A. Garriss, A. Iyer, B. Chapman, S. Kresovich & S. Mc-Couch, 2003. High resolution genetic mapping and candidate gene identification at the xa5 locus for bacterial blight resistance in rice (*Oryza sativa* L.). *Theor Appl Genet* 107: 62–73.
- Bohn, M., S. Groh, M.M. Khairallah, D.A. Hoisington, H.F. Utz & A.E. Melchinger, 2001. Re-evaluation of the prospects of marker-assisted selection for improving insect resistance against *Diatraea* spp. in tropical maize by cross validation and independent validation. *Theor Appl Genet* 103: 1059–1067.
- Brondani, C., P. Hideo, N. Rangel, T. Cristina, O. Borba, R. Pereira & V. Brondani, 2003. Transferability of microsatellite and sequence tagged site markers in *Oryza* species. *Hereditas* 138: 187–192.
- Cakir, M., S. Gupta, G.J. Platz, G.A. Ablett, R. Loughman, L.C. Emebiri, D. Poulsen, C.D. Li, R.C.M. Lance, N.W. Galwey, M.G.K. Jones & R. Appels, 2003. Mapping and validation of the genes for resistance to *Pyrenophora teres* f. *teres* in barley (*Hordeum vulgare* L.). *Aust J Agric Res* 54: 1369–1377.
- Campbell, A.W., G. Daggard, F. Bekes, A. Pedler, M.W. Sutherland & R. Appels, 2001. Targeting AFLP-DNA markers to specific traits and chromosome regions. *Aust J Agric Res* 52: 1153–1160.
- Cato, S., R. Gardner, J. Kent & T. Richardson, 2001. A rapid PCR-based method for genetically mapping ESTs. *Theor Appl Genet* 102: 296–306.
- Chalmers, K.J., A.W. Campbell, J. Kretschmer, A. Karakousis, P.H. Henschke, S. Pierens, N. Harker, M. Pallotta, G.B. Cornish, M.R. Shariflou, L.R. Rampling, A. McLauchlan, G. Daggard, P.J. Sharp, T.A. Holton, M.W. Sutherland, R. Appels & P. Langridge, 2001. Construction of three linkage maps in bread wheat, *Triticum aestivum*. *Aust J Ag Res* 52: 1089–1119.
- Chunwongse, J., S. Doganlar, C. Crossman, J. Jiang & S.D. Tanksley, 1997. High-resolution genetic map the *Lv* resistance locus in tomato. *Theor Appl Genet* 95: 220–223.
- Churchill, G.A. & R.W. Doerge, 1994. Empirical threshold values for quantitative trait mapping. *Genetics* 138: 963–971.
- Collard, B.C.Y., E.C.K. Pang & P.W.J. Taylor, 2003. Selection of wild *Cicer* accessions for the generation of mapping populations segregating for resistance to ascochyta blight. *Euphytica* 130: 1–9.
- Collins, H.M., J.F. Panozzo, S.J. Logue, S.P. Jefferies & A.R. Barr, 2003. Mapping and validation of chromosome regions associated with high malt extract in barley (*Hordeum vulgare* L.). *Aust J Agric Res* 54: 1223–1240.
- Danesh, D., S. Aarons, G. McGill & N. Young, 1994. Genetic dissection of oligogenic resistance to bacterial wilt in tomato. *Mol Plant-Microbe Interact* 7: 464–471.
- Darvasi, A., A. Weinreb, V. Minke, J.I. Weller & M. Soller, 1993. Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. *Genetics* 134: 943–951.
- Daviewala, A., K. Chowdari, S. Kumar, A. Reddy, P. Ranjekar & V. Gupta, 2000. Use of three different marker systems to estimate genetic diversity of Indian elite rice varieties. *Genetica* 108: 269–284.

- Doerge, R.W., 2002. Mapping and analysis of quantitative trait loci in experimental populations. *Nat Rev Genet* 3: 43–52.
- Donini, P., P. Stephenson, G. Bryan & R. Koebner, 1998. The potential of microsatellites for high throughput genetic diversity assessment in wheat and barley. *Genet Resour Crop Evol* 45: 415–451.
- Dreher, K., M. Khairallah, J. Ribaut & M. Morris, 2003. Money matters (I): Costs of field and laboratory procedures associated with conventional and marker-assisted maize breeding at CIMMYT. *Mol Breed* 11: 221–234.
- Eagles, H., H. Bariana, F. Ogonnaya, G. Rebetzke, G. Hollamby, R. Henry, P. Henschke & M. Carter, 2001. Implementation of markers in Australian wheat breeding. *Aust J Agric Res* 52: 1349–1356.
- Ellis, M.H., W. Spielmeier, K.R. Gale, G.J. Rebetzke & R.A. Richards, 2002. “Perfect” markers for the Rht-B1b and Rht-D1b dwarfing genes in wheat. *Theor Appl Genet* 105: 1038–1042.
- Faris, J.D., K.M. Haen & B.S. Gill, 2000. Saturation mapping of a gene-rich recombination hot spot region in wheat. *Genetics* 154: 823–835.
- Fasoula, V.A., D.K. Harris, M.A. Bailey, D.V. Phillips & H.R. Boerma, 2003. Identification, mapping, and confirmation of a soybean gene for bud blight resistance. *Crop Sci* 43: 1754–1759.
- Flandez-Galvez, H., P.K. Ades, R. Ford, E.C.K. Pang & P.W.J. Taylor, 2003a. QTL analysis for ascochyta blight resistance in an intraspecific population of chickpea (*Cicer arietinum* L.). *Theor Appl Genet* 107: 1257–1265.
- Flandez-Galvez, H., R. Ford, E.C.K. Pang & P.W.J. Taylor, 2003b. An intraspecific linkage map of the chickpea (*Cicer arietinum* L.) genome based on sequence tagged microsatellite site and resistance gene analog markers. *Theor Appl Genet* 106: 1447–1456.
- Foolad, M. & R. Jones, 1993. Mapping salt-tolerance genes in tomato (*Lycopersicon esculentum*) using trait-based marker analysis. *Theor Appl Genet* 87: 184–192.
- Ford, R., E.C.K. Pang & P.W.J. Taylor, 1999. Genetics of resistance to ascochyta blight (*Ascochyta lentis*) of lentil and the identification of closely linked RAPD markers. *Theor Appl Genet* 98: 93–98.
- Forster, J.W., E.S. Jones, R. Kolliker, M.C. Drayton, J.L. Dumsday, M.P. Dupal, K.M. Guthridge, N.L. Mahoney, E. van Zijl de Jong & K.F. Smith, 2000. Development and implementation of molecular markers for crop improvement. In: ‘Molecular Breeding of Forage Crops.’ In: G. Spangenberg (ed.), *Proceedings of the 2nd International Symposium, Lorne and Hamilton, Victoria Australia, November 19–24*, pp. 101–133, Kluwer Academic Publishers, London.
- Fregene, M., E. Okogbenin, C. Mba, F. Angel, M.C. Suarez, G. Janneth, P. Chavarriaga, W. Roca, M. Bonierbale & J. Tohme, 2001. Genome mapping in cassava improvement: Challenges, achievements and opportunities. *Euphytica* 120: 159–165.
- Frisch, M., M. Bohn & A.E. Melchinger, 1999. Comparison of selection strategies for marker-assisted backcrossing of a gene. *Crop Sci* 39: 1295–1301.
- Frisch, M., M. Bohn & A.E. Melchinger, 2000. PLABSIM: Software for simulation of marker-assisted backcrossing. *J Hered* 91: 86–87.
- Gale, M.D. & K.M. Devos, 1998. Plant comparative genetics after 10 years. *Science* 282: 656–659.
- Gardiner, J., E. Coe, S. Melia-Hancock, D. Hoisington & S. Chao, 1993. Development of a core RFLP map in maize using an immortalised F2 population. *Genetics* 134: 917–930.
- Gebhardt, C. & J.P.T. Valkonen, 2001. Organization of genes controlling disease resistance in the potato genome. *Annu Rev Phytopathol* 39: 79–102.
- George, M.L.C., B.M. Prasanna, R.S. Rathore, T.A.S. Setty, F. Kasim, M. Azrai, S. Vasal, O. Balla, D. Hautea, A. Canama, E. Regalado, M. Vargas, M. Khairallah, D. Jeffers & D. Hoisington, 2003. Identification of QTLs conferring resistance to downy mildews of maize in Asia. *Theor Appl Genet* 107: 544–551.
- Giovannoni, J., R. Wing, M. Ganai & S. Tanksley, 1991. Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations. *Nucleic Acid Res* 19: 6553–6558.
- Glover, K.D., D. Wang, P.R. Arelli, S.R. Carlson, S.R. Cianzio & B.W. Diers, 2004. Near isogenic lines confirm a soybean cyst nematode resistance gene from PI 88788 on linkage group J. *Crop Sci* 44: 936–941.
- Gu, W., N. Weeden, J. Yu & D. Wallace, 1995. Large-scale, cost-effective screening of PCR products in marker-assisted selection applications. *Theor Appl Genet* 91: 465–470.
- Gupta, P., R. Varshney, P. Sharma & B. Ramesh, 1999. Molecular markers and their applications in wheat breeding. *Plant Breed* 118: 369–390.
- Gupta, P.K., J.K. Roy & M. Prasad, 2001. Single nucleotide polymorphisms: A new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. *Curr Sci* 80: 524–535.
- Hackett, C., 2002. Statistical methods for QTL mapping in cereals. *Plant Mol Biol* 48: 585–599.
- Haley, C. & L. Andersson, 1997. Linkage mapping of quantitative trait loci in plants and animals, pp. 49–71. In: P. Dear (Ed.), *Genome mapping—A practical approach*, Oxford University Press, New York.
- Han, F., A. Kleinhofs, S. Ullrich, A. Kilian, M. Yano & T. Sasaki, 1998. Synteny with rice: Analysis of barley malting quality QTLs and rpg4 chromosome regions. *Genome* 41: 373–380.
- Harker, N., L.R. Rampling, M.R. Shariflou, M.J. Hayden, T.A. Holton, M.K. Morell, P.J. Sharp, R.J. Henry & K.J. Edwards, 2001. Microsatellites as markers for Australian wheat improvement. *Aust J Agric Res* 52: 1121–1130.
- Hartl, D. & E. Jones, 2001. *Genetics: Analysis of Genes and Genomes*, Jones and Bartlett Publishers, Sudbury, MA.
- Hayashi, K., N. Hashimoto, M. Daigen & I. Ashikawa, 2004. Development of PCR-based SNP markers for rice blast resistance genes at the Piz locus. *Theor Appl Genet* 108: 1212–1220.
- Henry, R., 1997. Molecular markers in plant improvement. In: *Practical Applications of Plant Molecular Biology*, pp. 99–132, Chapman and Hall, London.
- Hittalmani, S., H.E. Shashidhar, P.G. Bagali, N. Huang, J.S. Sidhu, V.P. Singh & G.S. Khush, 2002. Molecular mapping of quantitative trait loci for plant growth, yield and yield related traits across three diverse locations in a doubled haploid rice population. *Euphytica* 125: 207–214.
- Hori, K., T. Kobayashi, A. Shimizu, K. Sato, K. Takeda & S. Kawasaki, 2003. Efficient construction of high-density linkage map and its application to QTL analysis in barley. *Theor Appl Genet* 107: 806–813.
- Huettel, B., P. Winter, K. Weising, W. Choumane, F. Weigand & G. Kahl, 1999. Sequence-tagged microsatellite site markers for chickpea (*Cicer arietinum* L.). *Genome* 42: 210–217.
- Ishimaru, K., M. Yano, N. Aoki, K. Ono, T. Hirose, S.Y. Lin, L. Monna, T. Sasaki & R. Ohsugi, 2001. Toward the mapping of physiological and agronomic characters on a rice function map: QTL analysis and comparison between QTLs and expressed sequence tags. *Theor Appl Genet* 102: 793–800.
- Jahufer, M., M. Cooper, J. Ayres & R. Bray, 2002. Identification of research to improve the efficiency of breeding strategies for white clover in Australia: A review. *Aust J Agric Res* 53: 239–257.

- Jahufer, M., B. Barret, A. Griffiths & D. Woodfield, 2003. DNA fingerprinting and genetic relationships among white clover cultivars. In: J. Morton (Ed.), *Proceedings of the New Zealand Grassland Association*, Vol. 65, pp. 163–169, Taieri Print Limited, Dunedin.
- Jampatong, C., M. McMullen, B. Barry, L. Darrah, P. Byrne & H. Kross, 2002. Quantitative trait loci for first- and second-generation European corn borer resistance derived from maize inbred Mo47. *Crop Sci* 42: 584–593.
- Jansen, R., 1993. Interval mapping of multiple quantitative trait loci. *Genetics* 135: 205–211.
- Jansen, R. & P. Stam, 1994. High resolution of quantitative traits into multiple loci via interval mapping. *Genetics* 136: 1447–1455.
- Jones, E.S., W.A. Breese, C.J. Liu, S.D. Singh, D.S. Shaw & J.R. Witcombe, 2002. Mapping quantitative trait loci for resistance to downy mildew in pearl millet: Field and glasshouse screens detect the same QTL. *Crop Sci* 42: 1316–1323.
- Jones, N., H. Ougham & H. Thomas, 1997. Markers and mapping: We are all geneticists now. *New Phytol* 137: 165–177.
- Joshi, C. & H. Nguyen, 1993. RAPD (random amplified polymorphic DNA) analysis based intervarietal genetic relationships among hexaploid wheats. *Plant Sci* 93: 95–103.
- Joshi, S., P. Ranjekar & V. Gupta, 1999. Molecular markers in plant genome analysis. *Curr Sci* 77: 230–240.
- Jung, G., H.M. Ariyaratne, D.P. Coyne & J. Nienhuis, 2003. Mapping QTL for bacterial brown spot resistance under natural infection in field and seedling stem inoculation in growth chamber in common bean. *Crop Sci* 43: 350–357.
- Jung, G., P.W. Skroch, J. Nienhuis, D.P. Coyne, E. Arnaud-Santana, H.M. Ariyaratne & J.M. Marita, 1999. Confirmation of QTL associated with common bacterial blight resistance in four different genetic backgrounds in common bean. *Crop Sci* 39: 1448–1455.
- Kantety, R.V., M. La Rota, D.E. Matthews & M.E. Sorrells, 2002. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Mol Biol* 48: 501–510.
- Karakousis, A., J.P. Gustafson, K.J. Chalmers, A.R. Barr & P. Langridge, 2003. A consensus map of barley integrating SSR, RFLP, and AFLP markers. *Aust J Agric Res* 54: 1173–1185.
- Kasha, K.J., 1999. Biotechnology and world food supply. *Genome* 42: 642–645.
- Kearsey, M. & H. Pooni, 1996. *The genetical analysis of quantitative traits*. Chapman & Hall, London.
- Kelly, J.D. & P.N. Miklas, 1998. The role of RAPD markers in breeding for disease resistance in common bean. *Mol Breed* 4: 1–11.
- Kelly, J.D., P. Gepts, P.N. Miklas & D.P. Coyne, 2003. Tagging and mapping of genes and QTL and molecular marker-assisted selection for traits of economic importance in bean and cowpea. *Field Crops Res* 82: 135–154.
- Khairallah, M., M. Bohn, C. Jiang, J. Deutsch, D. Jewell, J. Mihm, A. Melchinger, D. Gonzalez-De-Leon & D. Hoisington, 1998. Molecular mapping of QTL for southwestern corn borer resistance, plant height and flowering in tropical maize. *Plant Breed* 117: 309–318.
- Kochert, G., 1994. RFLP technology, p. 8–38. In: R.L. Phillips & I.K. Vasil (Eds.), *DNA-based Markers in Plants*. Kluwer Academic Publishers, Dordrecht.
- Koebner, R.M.D. & R.W. Summers, 2003. 21st century wheat breeding: Plot selection or plate detection? *Trends Biotechnol* 21: 59–63.
- Kunzel, G., L. Korzun & A. Meister, 2000. Cytologically integrated physical restriction fragment length polymorphism maps for the barley genome based on translocation breakpoints. *Genetics* 154: 397–412.
- Lander, E. & D. Botstein, 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121: 185–199.
- Lander, E.S. & L. Kruglyak, 1995. Genetic dissection of complex traits: Guidelines for interpreting and reporting linkage results. *Nat Genet* 11: 241–247.
- Lander, E.S., P. Green, J. Abrahamson, A. Barlow, M.J. Daly, S.E. Lincoln & L. Newburg, 1987. Mapmaker an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1: 174–181.
- Langridge, P., E. Lagudah, T. Holton, R. Appels, P. Sharp & K. Chalmers, 2001. Trends in genetic and genome analyses in wheat: A review. *Aust J Agric Res* 52: 1043–1077.
- Laurie, D. & K. Devos, 2002. Trends in comparative genetics and their potential impacts on wheat and barley research. *Plant Mol Biol* 48: 729–740.
- Lecomte, L., P. Duffe, M. Buret, B. Servin, F. Hospital & M. Causse, 2004. Marker-assisted introgression of five QTLs controlling fruit quality traits into three tomato lines revealed interactions between QTLs and genetic backgrounds. *Theor Appl Genet* 109: 658–668.
- Lem, P. & J. Lallemand, 2003. Grass consensus STS markers: An efficient approach for detecting polymorphism in *Lolium*. *Theor Appl Genet* 107: 1113–1122.
- Lee, M., 1995. *DNA Markers and Plant Breeding Programs*. Adv Agron 55: 265–344.
- Lefebvre, V., S. Pflieger, A. Thabuis, C. Caranta, A. Blattes, J.C. Chauvet, A.M. Daubeze & A. Palloix, 2002. Towards the saturation of the pepper linkage map by alignment of three intraspecific maps including known-function genes. *Genome* 45: 839–854.
- Lehmsiek, A., A. Esterhuizen, D. van Staden, S. Nelson & A. Retief, 2001. Genetic mapping of gray leaf spot (GLS) resistance genes in maize. *Theor Appl Genet* 103: 797–803.
- Li, L., S. Lu, D. O'Halloran, D. Garvin & J. Vrebal, 2003. High-resolution genetic and physical mapping of the cauliflower high-beta-carotene gene *Or* (Orange). *Mol Genet Genomic* 270: 132–138.
- Li, Z., L. Jakkula, R.S. Hussey, J.P. Tamulonis & H.R. Boerma, 2001. SSR mapping and confirmation of the QTL from PI96354 conditioning soybean resistance to southern root-knot nematode. *Theor Appl Genet* 103: 1167–1173.
- Lincoln, S., M. Daly & E. Lander, 1993a. Constructing genetic linkage maps with MAPMAKER/EXP. Version 3.0. Whitehead Institute for Biomedical Research Technical Report, 3rd Edn. Lincoln, S., M. Daly & E. Lander, 1993b. Mapping genes controlling quantitative traits using MAPMAKER/QTL. Version 1.1. Whitehead Institute for Biomedical Research Technical Report, 2nd Edn.
- Lindhout, P., 2002. The perspectives of polygenic resistance in breeding for durable disease resistance. *Euphytica* 124: 217–226.
- Liu, B., 1998. *Statistical Genomics: Linkage, Mapping and QTL Analysis*. CRC Press, Boca Raton.
- Liu, S. & J.A. Anderson, 2003. Targeted molecular mapping of a major wheat QTL for *Fusarium* head blight resistance using wheat ESTs and synteny with rice. *Genome* 46: 817–823.
- Lombard, V. & R. Delourme, 2001. A consensus linkage map for rapeseed (*Brassica napus* L.): Construction and integration of three individual maps from DH populations. *Theor Appl Genet* 103: 491–507.

- Ma, X.F., K. Ross & J.P. Gustafson, 2001. Physical mapping of restriction fragment length polymorphism (RFLP) markers in homeologous groups 1 and 3 chromosomes of wheat by in situ hybridization. *Genome* 44: 401–412.
- Mackill, D.J., H.T. Nguyen & J. Zhan, 1999. Use of molecular markers in plant improvement programs for rainfed lowland rice. *Field Crops Res* 64: 177–185.
- Manly, K.F., H. Cudmore Robert, Jr. & J.M. Meer, 2001. Map Manager QTX, cross-platform software for genetic mapping. *Mamm Genome* 12: 930–932.
- Marquez-Cedillo, L., P. Hayes, A. Kleinohs, W. Legge, B. Rosnagel, K. Sato, S. Ullrich & D. Wesenberg, 2001. QTL analysis of agronomic traits in barley based on the doubled haploid progeny of two elite North American varieties representing different germplasm groups. *Theor Appl Genet* 103: 625–637.
- Masi, P., P.L.S. Zeuli & P. Donini, 2003. Development and analysis of multiplex microsatellite markers sets in common bean (*Phaseolus vulgaris* L.). *Mol Breed* 11: 303–313.
- McCouch, S.R. & R.W. Doerge, 1995. QTL mapping in rice. *Trends Genet* 11: 482–487.
- McCouch, S.R., X. Chen, O. Panaud, S. Temnykh, Y. Xu, Y. Cho, N. Huang, T. Ishii & M. Blair, 1997. Microsatellite marker development, mapping and applications in rice genetics and breeding. *Plant Mol Biol* 35: 89–99.
- Mehlenbacher, S.A., 1995. Classical and molecular approaches to breeding fruit and nut crops for disease resistance. *HortScience* 30: 466–477.
- Melchinger, A.E., H.F. Utz & C.C. Schon, 1998. Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. *Genetics* 149: 383–403.
- Meyer, K., G. Benning & E. Grill, 1996. Cloning of plant genes based on genetic map position. In: A.H. Paterson (Ed.), *Genome mapping in plants*, pp. 137–154. R G Landes Company, San Diego, California Academic Press, Austin, Texas.
- Michelmore, R., 1995. Molecular approaches to manipulation of disease resistance genes. *Annu Rev Phytopathol* 33: 393–427.
- Michelmore, R., I. Paran & R. Kesseli, 1991. Identification of markers linked to disease-resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. *Proc Natl Acad Sci USA* 88: 9828–9832.
- Mohan, M., S. Nair, A. Bhagwat, T.G. Krishna, M. Yano, C.R. Bhatia & T. Sasaki, 1997. Genome mapping, molecular markers and marker-assisted selection in crop plants. *Mol Breed* 3: 87–103.
- Mohapatra, T., K.S. Singh, S. Swain, R. Sharma & N. Singh, 2003. STMS-based DNA fingerprints of the new plant type wheat lines. *Curr Sci* 84: 1125–1129.
- Morgante, M. & F. Salamini, 2003. From plant genomics to breeding practice. *Curr Opin Biotechnol* 14: 214–219.
- Morris, M., K. Dreher, J.M. Ribaut & M. Khairallah, 2003. Money matters (II): Costs of maize inbred line conversion schemes at CIMMYT using conventional and marker-assisted selection. *Mol Breed* 11: 235–247.
- Muehlbauer, F., W. Kaiser & C. Simon, 1994. Potential for wild species in cool season food legume breeding. *Euphytica* 73: 109–114.
- Nelson, J.C., 1997. Qgene—software for marker-based genomic analysis and breeding. *Mol Breed* 3: 239–245.
- Ogonnaya, F.C., N.C. Subrahmanyam, O. Moullet, J. de Majnik, H.A. Eagles, J.S. Brown, R.F. Eastwood, J. Kollmorgen, R. Appels & E.S. Lagudah, 2001. Diagnostic DNA markers for cereal cyst nematode resistance in bread wheat. *Aust J Agric Res* 52: 1367–1374.
- Ortiz, R., 1998. Critical role of plant biotechnology for the genetic improvement of food crops: Perspectives for the next millennium. *Electron J Biotechnol [online]* 1(3): Issue of August 15.
- Paran, I. & R. Michelmore, 1993. Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. *Theor Appl Genet* 85: 985–993.
- Paterson, A., S. Tanksley & M.E. Sorrells, 1991a. DNA markers in plant improvement. *Adv Agron* 44: 39–90.
- Paterson, A.H., 1996a. Making genetic maps. In: A.H. Paterson (Ed.), *Genome Mapping in Plants*, pp. 23–39. R. G. Landes Company, San Diego, California; Academic Press, Austin, Texas.
- Paterson, A.H., 1996b. Mapping genes responsible for differences in phenotype. In: A.H. Paterson (Ed.), *Genome Mapping in Plants*, pp. 41–54. R. G. Landes Company, San Diego, California; Academic Press; Austin, Texas.
- Paterson, A.H., E.S. Lander, J.D. Hewitt, S. Peterson, S.E. Lincoln & S.D. Tanksley, 1988. Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature* 335: 721–726.
- Paterson, A.H., S. Damon, J.D. Hewitt, D. Zamir, H.D. Rabinowitch, S.E. Lincoln, E.S. Lander & S.D. Tanksley, 1991b. Mendelian factors underlying quantitative traits in tomato comparison across species generations and environments. *Genetics* 127: 181–198.
- Penner, G., 1996. RAPD analysis of plant genomes. In: P.P. Jauhar (Ed.), *Methods of Genome Analysis in Plants*, pp. 251–268. CRC Press, Boca Raton.
- Perovic, D., N. Stein, H. Zhang, A. Drescher, M. Prasad, R. Kota, D. Kopahnke & A. Graner, 2004. An integrated approach for comparative mapping in rice and barley with special reference to the Rph16 resistance locus. *Funct Integr Genomics* 4: 74–83.
- Pflieger, S., V. Lefebvre & M. Causse, 2001. The candidate gene approach in plant genetics: A review. *Mol Breed* 7: 275–291.
- Pilet-Nayel, M.L., F.J. Muehlbauer, R.J. McGee, J.M. Kraft, A. Baranger & C.J. Coyne, 2002. Quantitative trait loci for partial resistance to *Aphanomyces* root rot in pea. *Theor Appl Genet* 106: 28–39.
- Polacco, M., E. Coe, Z. Fang, D. Hancock, H. Sanchez-Villeda & S. Schroeder, 2002. MaizeDB—a functional genomics perspective. *Comp Funct Genomics* 3: 128–131.
- Powell, W., G. Machray & J. Provan, 1996. Polymorphism revealed by simple sequence repeats. *Trends Plant Sci* 1: 215–222.
- Price, A. & B. Courtois, 1999. Mapping QTLs associated with drought resistance in rice: Progress, problems and prospects. *Plant Growth Reg* 29: 123–133.
- Rafalski, A., 2002. Applications of single nucleotide polymorphisms in crop genetics. *Curr Opin Plant Biol* 5: 94–100.
- Rafalski, J. & S. Tingey, 1993. Genetic diagnostics in Plant Breed: RAPDs, microsatellites and machines. *Trends Genet* 9: 275–280.
- Rampling, L.R., N. Harker, M.R. Shariflou & M.K. Morell, 2001. Detection and analysis systems for microsatellite markers in wheat. *Aust J Agric Res* 52: 1131–1141.
- Reyna, N. & C.H. Sneller, 2001. Evaluation of marker-assisted introgression of yield QTL alleles into adapted soybean. *Crop Sci* 41: 1317–1321.
- Ribaut, J.M. & J. Betran, 1999. Single large-scale marker-assisted selection (SLS-MAS). *Mol Breed* 5: 531–541.

-
- Ribaut, J.-M. & D. Hoisington, 1998. Marker-assisted selection: New tools and strategies. *Trends Plant Sci* 3: 236–239.
 - Ribaut, J.-M., C. Jiang & D. Hoisington, 2002. Simulation experiments on efficiencies of gene introgression by backcrossing. *Crop Sci* 42: 557–565.
 - Ribaut, J.-M., X. Hu, D. Hoisington & D. Gonzalez-De-Leon, 1997. Use of STSs and SSRs as rapid and reliable preselection tools in marker-assisted selection backcross scheme. *Plant Mol Biol Report* 15: 156–164.
 - Ripol, M., G.A. Churchill, J. da Silva & M.E. Sorrells, 1999. Statistical aspects of genetic mapping in autopolyploids. *Gene* 235: 31–41.
 - Risch, N., 1992. Genetic linkage: Interpreting LOD scores. *Science* 255: 803–804.
 - Sayed, H., H. Kayyal, L. Ramsey, S. Ceccarelli & M. Baum, 2002. Segregation distortion in doubled haploid lines of barley (*Hordeum vulgare* L.) detected by simple sequence repeat markers. *Euphytica* 225: 265–272.
 - Serquen, F., J. Bacher & J. Staub, 1997. Mapping and QTL analysis of horticultural traits in a narrow cross in cucumber (*Cucumis sativus* L.) using random-amplified polymorphic DNA markers. *Mol Breed* 3: 257–268.
 - Shan, X., T.K. Blake & L.E. Talbert, 1999. Conversion of AFLP markers to sequence-specific PCR markers in barley and wheat. *Theor Appl Genet* 98: 1072–1078.
 - Sharp, P.J., S. Johnston, G. Brown, R.A. McIntosh, M. Pallotta, M. Carter, H.S. Bariana, S. Khartkar, E.S. Lagudah, R.P. Singh, M. Khairallah, R. Potter & M.G.K. Jones, 2001. Validation of molecular markers for wheat breeding. *Aust J Agric Res* 52: 1357–1366.
 - Skiba, B., R. Ford & E.C.K. Pang, 2004. Construction of a linkage map based on a *Lathyrus sativus* backcross population and preliminary investigation of QTLs associated with resistance to ascochyta blight. *Theor Appl Genet* 109: 1726–1735.
 - Snowdon, R. & W. Friedt, 2004. Molecular markers in Brassica oilseeds breeding: Current status and future possibilities. *Plant Breed* 123: 1–8.
 - Spielmeyer, W., P.J. Sharp & E.S. Lagudah, 2003. Identification and validation of markers linked to broad-spectrum stem rust resistance gene *Sr2* in wheat (*Triticum aestivum* L.). *Crop Sci* 43: 333–336.
 - Stam, P., 1993. Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. *Plant J* 3: 739–744.
 - Staub, J.E., F. Serquen & M. Gupta, 1996. Genetic markers, map construction and their application in Plant Breed. *HortScience* 31: 729–741.
 - Stuber, C.W., M. Polacco & M.L. Senior, 1999. Synergy of empirical breeding, marker-assisted selection, and genomics to increase crop yield potential. *Crop Sci* 39: 1571–1583.
 - Svetleva, D., M. Velcheva & G. Bhowmik, 2003. Biotechnology as a useful tool in common bean (*Phaseolus vulgaris* L.) improvement: A review. *Euphytica* 131: 189–200.
 - Tanksley, S.D., 1993. Mapping polygenes. *Annu Rev Genet* 27: 205–233.
 - Tanksley, S.D., M.W. Ganal, J.P. Prince, M.C. De Vicente, M.W. Bonierbale, P. Broun, T.M. Fulton, J.J. Giovannoni & S. Grandillo, 1992. High density molecular linkage maps of the tomato and potato genomes. *Genetics* 132: 1141–1160.
 - Tanksley, S.D. & J.C. Nelson, 1996. Advanced backcross QTL analysis: A method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. *Theor Appl Genet* 92: 191–203.
 - Tanksley, S.D., M. Ganal & G.B. Martin, 1995. Chromosome landing — a paradigm for map-based cloning in plants with large genomes. *Trends Genet* 11: 63–68.
 - Tanksley, S.D., N.D. Young, A.H. Paterson & M. Bonierbale, 1989. RFLP mapping in plant breeding: New tools for an old science. *Biotechnology* 7: 257–264.
 - Tanksley, S.D., S. Grandillo, T. Fulton, D. Zamir, Y. Eshed, V. Petiard, J. Lopez & T. Beck-Bunn, 1996. Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L. pimpinellifolium*. *Theor Appl Genet* 92: 213–224.
 - Taramino, G. & S. Tingey, 1996. Simple sequence repeats for germplasm analysis and mapping in maize. *Genome* 39: 277–287.
 - Taran, B., T.E. Michaels & K.P. Pauls, 2002. Genetic mapping of agronomic traits in common bean. *Crop Sci* 42: 544–556.
 - Thomas, W., 2003. Prospects for molecular breeding of barley. *Ann Appl Biol* 142: 1–12.
 - Tuberosa, R., S. Salvi, M.C. Sanguineti, M. Maccaferri, S. Giuliani & P. Landi, 2003. Searching for quantitative trait loci controlling root traits in maize: A critical appraisal. *Plant Soil* 255: 35–54.
 - Utz, H. & A. Melchinger, 1996. PLABQTL: A program for composite interval mapping of QTL. *J Quant Trait Loci* 2(1).
 - Utz, H.F., A.E. Melchinger & C.C. Schon, 2000. Bias and sampling error of the estimated proportion of genotypic variance explained by quantitative trait loci determined from experimental data in maize using cross validation and validation with independent samples. *Genetics* 154: 1839–1849.
 - Van Berloo, R., H. Aalbers, A. Werkman & R.E. Niks, 2001. Resistance QTL confirmed through development of QTL-NILs for barley leaf rust resistance. *Mol Breed* 8: 187–195.
 - Van Sanford, D., J. Anderson, K. Campbell, J. Costa, P. Cregan, C. Griffey, P. Hayes & R. Ward, 2001. Discovery and deployment of molecular markers linked to fusarium head blight resistance: An integrated system for wheat and barley. *Crop Sci* 41: 638–644.
 - Visscher, P., R. Thompson & C. Haley, 1996. Confidence intervals in QTL mapping by bootstrapping. *Genetics* 143: 1013–1020.
 - Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hoernes, A. Frijters, J. Pot, J. Peleman, M. Kuiper & M. Zabeau, 1995. AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Res* 23: 4407–4414.
 - Wang, G. & A.H. Paterson, 1994. Assessment of DNA pooling strategies for mapping of QTLs. *Theor Appl Genet* 88: 355–361.
 - Wang, Z., G. Taramino, D. Yang, G. Liu, S. Tingey, G.-H. Miao & G. Wang, 2001. Rice ESTs with disease-resistance gene- or defense-response gene-like sequences mapped to regions containing major resistance genes or QTLs. *Mol Genet Genomics* 265: 302–310.
 - Warburton, M., X. Xianchun, J. Crossa, J. Franco, A.E. Melchinger, M. Frisch, M. Bohn & D. Hoisington, 2002. Genetic characterization of CIMMYT inbred maize lines and open pollinated population using large scale fingerprinting methods. *Crop Sci* 42: 1832–1840.
 - Weeden, N., G. Timmerman & J. Lu, 1994. Identifying and mapping genes of economic significance. *Euphytica* 73: 191–198.

- Weising, K., H. Nybom, K. Wolff & W. Meyer, 1995. Applications of DNA Fingerprinting in Plants and Fungi DNA Fingerprinting in Plants and Fungi, CRC Press, Boca Raton.
- Welsh, J. & M. McClelland, 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res* 18: 7213–7218.
- Williams, J., A. Kubelik, K. Livak, J. Rafalski & S. Tingey, 1990.
- DNA Polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res* 18: 6531–6535.
- Williams, K.J., 2003. The molecular genetics of disease resistance in barley. *Aust J Agric Res* 54: 1065–1079.
- Winter, P. & G. Kahl, 1995. Molecular marker technologies for plant improvement. *World Journal of Microbiology & Biotechnology* 11: 438–448.
- Winter, P., T. Pfaff, S. Udupa, B. Huttel, P. Sharma, S. Sahi, R. Arreguin-Espinoza, F. Weigand, F.J. Muehlbauer & G.
- Kahl, 1999. Characterisation and mapping of sequence-tagged microsatellite sites in the chickpea (*Cicer arietinum* L.) genome. *Mol Gen Genet* 262: 90–101.
- Wissuwa, M. & N. Ae, 2001. Further characterization of two QTLs that increase phosphorus uptake of rice (*Oryza sativa* L.) under phosphorus deficiency. *Plant Soil* 237: 275–286.
- Witcombe, J.R. & D.S. Virk, 2001. Number of crosses and population size for participatory and classical plant breeding. *Euphytica* 122: 451–462.
- Wu, K., W. Burnquist, M.E. Sorrells, T. Tew, P. Moore & S.D. Tanksley, 1992. The detection and estimation of linkage in polyploids using single-dose restriction fragments. *Theor Appl Genet* 83: 294–300.
- Xu, Y., L. Zhu, J. Xiao, N. Huang & S.R. McCouch, 1997. Chromosomal regions associated with segregation distortion of molecular markers in F₂, backcross, doubled haploid, and recombinant inbred populations in rice (*Oryza sativa* L.). *Mol Gen Genet* 253: 535–545.
- Yamamoto, K. & T. Sasaki, 1997. Large-scale EST sequencing in rice. *Plant Mol Biol* 35: 135–144.
- Yao, H., Q. Zhou, J. Li, H. Smith, M. Yandeu, B.J. Nikolau & P.S. Schnable, 2002. Molecular characterization of meiotic recombination across the 140-kb multigenic *al-sh2* interval of maize. *Proc Natl Acad Sci USA* 99: 6157–6162.
- Young, N.D., 1994. Constructing a plant genetic linkage map with DNA markers, p. 39–57, In: I. K.V. Ronald & L. Phillips (Eds.), *DNA-based markers in plants*. Kluwer, Dordrecht/ Boston/London.
- Young, N.D., 1996. QTLmapping and quantitative disease resistance in plants. *Annu Rev Phytopathol* 34: 479–501.
- Young, N.D., 1999. A cautiously optimistic vision for markerassisted breeding. *Mol Breed* 5: 505–510.
- Yu, K., S. Park & V. Poysa, 2000. Marker-assisted selection of common beans for resistance to common bacterial blight: Efficacy and economics. *Plant Breed* 119: 411–415.
- Yu, L.-X. & H. Nguyen, 1994. Genetic variation detected with RAPD markers among upland and lowland rice cultivars (*Oryza sativa* L.). *Theor Appl Genet* 87: 668–672.
- Zeng, Z.-B., 1994. Precision mapping of quantitative trait loci. *Genetics* 136: 1457–1468.
- Zeng, Z.-B., 1993. Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. *Proc Natl Acad Sci USA* 90: 10972–10976.
- Zhang, L.P., G.Y. Lin, D. Nino-Liu & M.R. Foolad, 2003. Mapping QTLs conferring early blight (*Alternaria solani*) resistance in a *Lycopersicon esculentum* × *L. hirsutum* cross by selective genotyping. *Mol Breed* 12: 3–19.
- Zhang, W.K., Y.J. Wang, G.Z. Luo, J.S. Zhang, C.Y. He, X.L. Wu, J.Y. Gai & S.Y. Chen, 2004. QTL mapping of ten agronomic traits on the soybean (*Glycine max* L. Merr.) genetic map and their association with EST markers. *Theor Appl Genet* 108: 1131–1139.