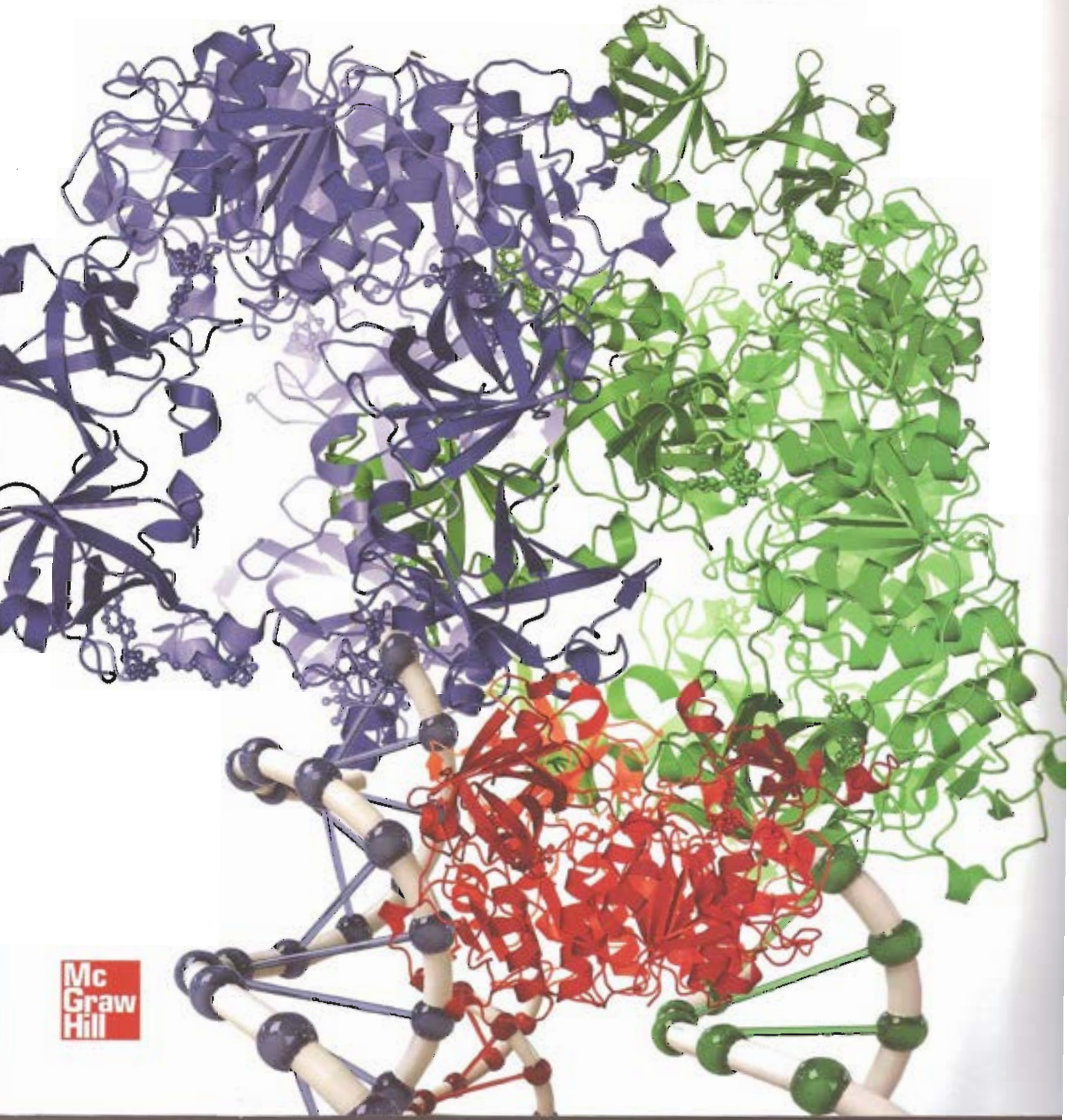


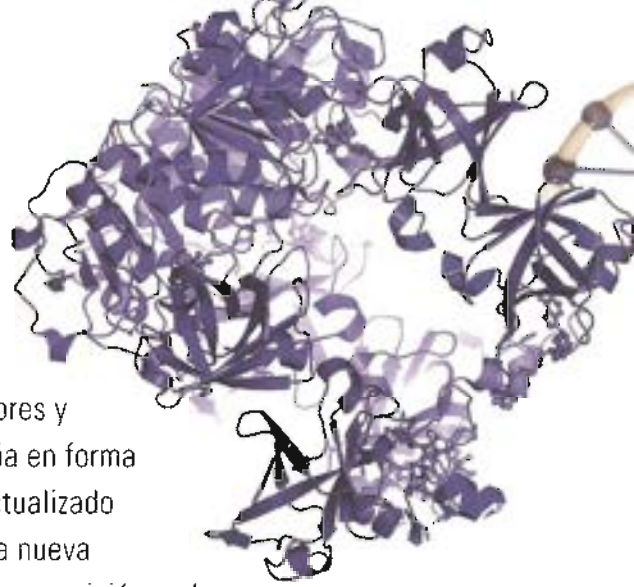
GENES IX

BENJAMIN LEWIN



Mc
Graw
Hill

Cuando se publicó la primera edición de *GENES*, Benjamín Lewin estableció el estándar de la enseñanza de la biología y genética moleculares con un abordaje unificado. La novena edición de este libro clásico continúa con dicha tradición, presenta la estructura y función de los genes en organismos eucariotas y procariotas. El Dr. Lewin mantiene el compromiso de proporcionar a los estudiantes, investigadores y educadores los conceptos actuales en este campo que cambia en forma constante. La novena edición de *GENES* incluye contenido actualizado y una cobertura más amplia de temas fundamentales con una nueva organización que le permite al estudiante enfocarse con mayor precisión en los genes y en su expresión. *GENES IX* también ostenta un diseño moderno, nuevo y un programa contemporáneo.



Características nuevas y principales de *GENES IX*

- Mayor cobertura en numerosas áreas, que incluye:
 - Replicación del DNA
 - Recombinación y reparación
 - El replicón
 - Regulación de la cromatina y regulación génica
 - Evolución de los genes
 - El cromosoma Y
- Reorganización para permitir a los instructores construir conceptos críticos a lo largo del curso.
- Nuevo diseño contemporáneo y un impactante programa de arte de cuatro colores.
- Nuevas actualizaciones de principio a fin, que incluyen información actual referente a la organización del genoma, replicación del DNA, regulación génica y mucho más.

 **Educación**

The McGraw-Hill Companies

ISBN-13: 978-970-10-6685-0
ISBN-10: 970-10-6685-5



Visite nuestra página WEB
www.mcgraw-hill-educacion.com

GENES IX

Editado por

Benjamin Lewin, PhD

Geneticist
Cambridge University

Novena edición

Traducción:

Héctor Barrera Villa Zevallos
Félix García Roig



MÉXICO • BOGOTÁ • BUENOS AIRES • CARACAS • GUATEMALA
LISBOA • MADRID • NUEVA YORK • SAN JUAN • SANTIAGO • SAO PAULO
AUCKLAND • LONDRES • MILÁN • MONTREAL • NUEVA DELHI
SAN FRANCISCO • SIDNEY • SINGAPUR • ST. LOUIS • TORONTO

Editor sponsor: Camilo Heras Martínez
Corrección de estilo: Guillermina Cuevas Meza y Elia Olvera Martínez
Supervisión de edición: Leonora Véliz Salazar
Supervisión de producción: José Luis González Huerta
Composición y formación: By Color Soluciones Gráficas

NOTA

La medicina es una ciencia en constante desarrollo. Conforme surjan nuevos conocimientos, se requerirán cambios de la terapéutica. El (los) autor(es) y los editores se han esforzado para que los cuadros de dosificación medicamentosa sean precisos y acordes con lo establecido en la fecha de publicación. Sin embargo, ante los posibles errores humanos y cambios en la medicina, ni los editores ni cualquier otra persona que haya participado en la preparación de la obra garantizan que la información contenida en ella sea precisa o completa, tampoco son responsables de errores u omisiones, ni de los resultados que con dicha información se obtengan. Convendría recurrir a otras fuentes de datos, por ejemplo, y de manera particular, habrá que consultar la hoja informativa que se adjunta con cada medicamento, para tener certeza de que la información de esta obra es precisa y no se han introducido cambios en la dosis recomendada o en las contraindicaciones para su administración. Esto es de particular importancia con respecto a fármacos nuevos o de uso no frecuente. También deberá consultarse a los laboratorios para recabar información sobre los valores normales.

GENES IX

Prohibida la reproducción total o parcial de esta obra, por cualquier medio, sin autorización escrita del editor.



DERECHOS RESERVADOS © 2008 respecto a la primera edición en español por
McGRAW-HILL INTERAMERICANA EDITORES, S.A. de C.V.
A subsidiary of The McGraw-Hill Companies, Inc.
Prolongación Paseo de la Reforma 1015, Torre A, Piso 17, Col. Desarrollo Santa Fe,
Delegación Álvaro Obregón
C.P. 01376, México, D.F.
Miembro de la Cámara Nacional de la Industria Editorial Mexicana, Reg. núm. 736

ISBN 10: 970-10-6685-5
ISBN 13: 978-970-10-6685-0

Translated from the ninth English edition of: *Genes IX*
Edited by Benjamin Lewin
Copyright © 2008 by Jones & Bartlett Publishers, Inc., 40 Tall Pine Drive, Sudbury, MA 01776
All Rights Reserved

ISBN 10: 0763740632
ISBN 13: 9780763740634

1234567890 09765432108
Impreso en México Printed in Mexico

Impreso en México en Abril del 2008 Printed in Mexico in April 2008
Impreso por Editorial Impresora Apolo Printed by Editorial Impresora Apolo

Contenido abreviado

Contenido vi

Prefacio xvi

- 1 Los genes son DNA 1
 - 2 Los genes codifican proteínas 23
 - 3 El gen interrumpido 37
 - 4 El contenido del genoma 55
 - 5 Secuencias genómicas y números de genes 76
 - 6 Agrupamientos y repeticiones 98
 - 7 El RNA mensajero 127
 - 8 Síntesis proteínica 151
 - 9 Utilización del código genético 189
 - 10 Localización de las proteínas 218
 - 11 Transcripción 256
 - 12 El operón 300
 - 13 RNA regulador 331
 - 14 Estrategias de los fagos 349
 - 15 El replicón 376
 - 16 Replicones extracromosómicos 392
 - 17 La replicación bacteriana está conectada con el ciclo celular 408
 - 18 Replicación del DNA 428
 - 19 Recombinación homóloga y específica de sitio 457
 - 20 Sistemas de reparación 499
 - 21 Transposones 521
 - 22 Retrovirus y retroposones 550
 - 23 Diversidad inmunitaria 570
 - 24 Promotores y potenciadores 609
 - 25 Activación de la transcripción 640
 - 26 Corte, empalme y procesamiento del RNA 667
 - 27 RNA catalítico 706
 - 28 Cromosomas 729
 - 29 Nucleosomas 757
 - 30 Control de la estructura de la cromatina 796
 - 31 Los efectos epigenéticos son heredados 818
- Glosario 845
- Índice alfabético 867

Contenido

Prefacio xvi

1 Los genes son DNA 1

- 1.1 Introducción 2
- 1.2 El DNA es el material genético de las bacterias 3
- 1.3 El DNA es el material genético de los virus 4
- 1.4 El DNA es el material genético de las células animales 5
- 1.5 Los polinucleótidos tienen bases nitrogenadas ligadas a un esqueleto de azúcar-fosfato 6
- 1.6 El DNA es una hélice dúplex 6
- 1.7 La duplicación del DNA es semiconservadora 8
- 1.8 Las cadenas de DNA se separan en la horquilla de duplicación 9
- 1.9 La información genética puede ser proporcionada por el DNA o por el RNA 10
- 1.10 Los ácidos nucleicos se hibridan por apareamiento de bases 12
- 1.11 Las mutaciones cambian la secuencia del DNA 14
- 1.12 Las mutaciones pueden afectar a pares de bases individuales o a secuencias más largas 15
- 1.13 Los efectos de las mutaciones pueden ser revertidos 16
- 1.14 Las mutaciones se concentran en puntos calientes 17
- 1.15 Numerosos puntos calientes son resultado de bases modificadas 18
- 1.16 Algunos agentes hereditarios son extremadamente pequeños 19
- 1.17 Resumen 20

2 Los genes codifican proteínas 23

- 2.1 Introducción 24
- 2.2 Un gen codifica a un solo polipéptido 24
- 2.3 Las mutaciones que se presentan en el mismo gen no se pueden complementar 25
- 2.4 Las mutaciones pueden provocar pérdida o ganancia de función 26

- 2.5 Un locus puede tener numerosos alelos mutantes diferentes 27
- 2.6 Un locus puede tener más de un alelo de tipo silvestre 28
- 2.7 La recombinación ocurre por al intercambio físico de DNA 28
- 2.8 El código genético se lee en tripletes 30
- 2.9 Cada secuencia tiene tres marcos de lectura posibles 31
- 2.10 Los genes procarióticos son colineales con sus proteínas 32
- 2.11 Numerosos procesos son necesarios para expresar el producto proteínico de un gen 33
- 2.12 Las proteínas actúan en *trans*, pero los sitios del DNA, en *cis* 35
- 2.13 Resumen 36

3 El gen interrumpido 37

- 3.1 Introducción 38
- 3.2 Un gen interrumpido está formado por intrones y exones 38
- 3.3 Las endonucleasas de restricción son una herramienta esencial en el mapeo del DNA 39
- 3.4 La organización de los genes interrumpidos puede conservarse 40
- 3.5 Las secuencias de los exones se conservan, pero las de los intrones varían 42
- 3.6 La distribución de tamaños de los genes es amplia 43
- 3.7 Algunas secuencias de DNA codifican a más de una proteína 45
- 3.8 ¿Cómo evolucionaron los genes interrumpidos? 47
- 3.9 Algunos exones pueden equipararse con funciones proteínicas 49
- 3.10 Los miembros de una familia de genes tienen una organización común 51
- 3.11 ¿Se encuentra toda la información genética contenida en el DNA? 53
- 3.12 Resumen 53

4 El contenido del genoma 55

- 4.1 Introducción 56
- 4.2 Pueden trazarse mapas de los genomas por ligamiento, por restricción, por división o por secuencia de DNA 56
- 4.3 Los genomas individuales son muy variables 57
- 4.4 Los RFLP y los SNP pueden ser utilizados para el mapeo genético 58
- 4.5 ¿Por qué los genomas son tan grandes? 60
- 4.6 Los genomas eucarióticos contienen secuencias de DNA repetitivas y no repetitivas 61
- 4.7 Los genes pueden ser aislados por la conservación de los exones 63
- 4.8 La conservación de la organización del genoma ayuda a identificar genes 65
- 4.9 Los organelos contienen DNA 67
- 4.10 Los genomas de los organelos son moléculas circulares de DNA que codifican proteínas de los organelos 69
- 4.11 La organización del DNA mitocondrial es variable 70
- 4.12 El genoma de los cloroplastos codifica a numerosas proteínas y moléculas de RNA 71
- 4.13 Las mitocondrias evolucionaron por endosimbiosis 72
- 4.14 Resumen 73

5 Secuencias genómicas y números de genes 76

- 5.1 Introducción 77
- 5.2 El número de genes bacterianos abarca un rango superior de un orden de magnitud 77
- 5.3 En numerosas eucariotas se conoce el número total de genes 79
- 5.4 ¿Cuántos tipos diferentes de genes hay? 81
- 5.5 El genoma humano tiene menos genes que los esperados 83
- 5.6 ¿Cómo se distribuyen los genes y otras secuencias en el genoma? 85
- 5.7 El cromosoma Y tiene varios genes específicos de la masculinidad 86
- 5.8 Las especies más complejas evolucionan agregando nuevas funciones génicas 87
- 5.9 ¿Cuántos genes son esenciales? 89
- 5.10 Los genes se expresan en niveles muy diferentes 92
- 5.11 ¿Cuántos genes se expresan? 93
- 5.12 El número de genes expresados puede medirse en masa 93
- 5.13 Resumen 94

6 Agrupamientos y repeticiones 98

- 6.1 Introducción 99
- 6.2 La duplicación de los genes es una fuerza importante en la evolución 100
- 6.3 Los agrupamientos de las globinas son formados por duplicación y por divergencia 101
- 6.4 La divergencia secuencial es la base del reloj evolutivo 104
- 6.5 La velocidad de sustitución neutral puede ser medida a partir de la divergencia de secuencias repetidas 107
- 6.6 Los pseudogenes son callejones sin salida de la evolución 108
- 6.7 Mediante entrecruzamiento desigual se reestructuran los agrupamientos de genes 109
- 6.8 Los genes del RNAr forman repeticiones en tándem 112
- 6.9 Los genes repetidos de RNAr mantienen una secuencia constante 114
- 6.10 La fijación entrecruzada podría mantener repeticiones idénticas 115
- 6.11 Los DNA satélite a menudo se encuentran en la heterocromatina 117
- 6.12 Los satélites de los artrópodos tienen repeticiones idénticas muy cortas 119
- 6.13 Los satélites de los mamíferos consisten de repeticiones jerárquicas 120
- 6.14 Los minisatélites facilitan el mapeo genético 123
- 6.15 Resumen 125

7 El RNA mensajero 127

- 7.1 Introducción 128
- 7.2 El RNAm se produce por transcripción y se traduce 129
- 7.3 El RNA de transferencia forma una hoja de trébol 130
- 7.4 El tallo aceptor y el anticodón se encuentran en los extremos de la estructura terciaria 131
- 7.5 El RNA mensajero es traducido por los ribosomas 132
- 7.6 A una molécula de RNAm se unen numerosos ribosomas 133
- 7.7 El ciclo vital del RNA mensajero de las bacterias 135
- 7.8 El RNAm eucariótico es modificado durante su transcripción o después de ésta 137
- 7.9 El extremo 5' del RNAm eucariótico posee un casquete 138
- 7.10 El extremo 3' está poliadenilado 139
- 7.11 La degradación del RNAm bacteriano involucra a múltiples enzimas 140
- 7.12 La estabilidad del RNAm depende de su estructura y de su secuencia 141

- 7.13 La degradación del RNAm involucra a múltiples actividades 143
- 7.14 Las mutaciones sin sentido activan un sistema de vigilancia 144
- 7.15 Las moléculas de RNA eucariótico se transportan 145
- 7.16 El RNAm puede localizarse específicamente 146
- 7.17 Resumen 147

8 Síntesis proteínica 151

- 8.1 Introducción 151
- 8.2 La síntesis proteínica ocurre por iniciación, elongación y terminación 153
- 8.3 La precisión de la síntesis proteínica es controlada por mecanismos especiales 156
- 8.4 La iniciación en las bacterias requiere de subunidades 30S y de factores accesorios 157
- 8.5 Un RNAt iniciador especial comienza la cadena polipeptídica 158
- 8.6 El uso del fMet-RNAt, está controlado por el IF-2 y por el ribosoma 160
- 8.7 La iniciación implica el apareamiento de bases entre el RNAm y el RNAr 161
- 8.8 Las subunidades pequeñas buscan sitios de iniciación en el RNAm eucariótico 162
- 8.9 Las eucariotas utilizan un complejo formado por numerosos factores de iniciación 164
- 8.10 El factor de elongación Tu carga al aminoacil-RNAt en el sitio A 167
- 8.11 La cadena polipeptídica se transfiere al aminoacil-RNAt 168
- 8.12 La translocación mueve al ribosoma 169
- 8.13 Los factores de elongación se unen alternativamente al ribosoma 170
- 8.14 La síntesis proteínica termina con tres codones 172
- 8.15 Los codones de terminación son reconocidos por factores proteínicos 173
- 8.16 El RNA ribosómico se extiende en ambas subunidades ribosómicas 175
- 8.17 Los ribosomas tienen numerosos centros activos 177
- 8.18 El RNAr 16S desempeña un papel activo en la síntesis proteínica 179
- 8.19 El RNAr 23S tiene actividad peptidil transferasa 182
- 8.20 Las estructuras ribosómicas cambian cuando se unen las subunidades 183
- 8.21 Resumen 183

9 Utilización del código genético 189

- 9.1 Introducción 190

- 9.2 Los codones relacionados representan a aminoácidos relacionados 190
- 9.3 El reconocimiento codón-anticodón involucra balanceo 192
- 9.4 Los RNAt son procesados a partir de precursores más largos 194
- 9.5 El RNAt contiene bases modificadas 194
- 9.6 Las bases modificadas afectan el apareamiento codón-anticodón 196
- 9.7 El código universal tiene alteraciones esporádicas 197
- 9.8 En ciertos codones de terminación pueden insertarse aminoácidos nuevos 199
- 9.9 Los RNAt son cargados con aminoácidos por medio de sintetetasas 200
- 9.10 Las aminoacil-RNAt sintetetasas se clasifican en dos grupos 201
- 9.11 Las sintetetasas utilizan mecanismos de corrección para incrementar la precisión 203
- 9.12 Los RNAt supresores tienen anticodones mutados que leen a codones nuevos 206
- 9.13 Hay supresores de mutaciones sin sentido para cada codón de terminación 207
- 9.14 Los supresores pueden competir con la lectura de tipo silvestre del código 208
- 9.15 El ribosoma influye en la precisión de la traducción 209
- 9.16 La modificación de la codificación cambia el significado de los codones 211
- 9.17 Los cambios del marco de lectura ocurren en las secuencias resbaladizas 213
- 9.18 La evasión implica el movimiento del ribosoma 214
- 9.19 Resumen 215

10 Localización de las proteínas 218

- 10.1 Introducción 220
- 10.2 El desplazamiento a través de una membrana requiere de un mecanismo especial 220
- 10.3 La translocación de las proteínas puede ser posterior a la traducción o durante ésta 221
- 10.4 Los chaperones pueden ser necesarios para el plegamiento de las proteínas 223
- 10.5 Las proteínas desnaturalizadas y las recién sintetizadas necesitan chaperones 224
- 10.6 La familia Hsp70 es ubicua 226
- 10.7 Las secuencias de señal inician la translocación 227
- 10.8 La secuencia de señal interactúa con la SRP 228
- 10.9 La SRP interactúa con el receptor de SRP 229
- 10.10 El traslocón forma un poro 231

- 10.11 La translocación requiere de inserción en el traslocón y (en ocasiones) de un trinquete en el ER 233
- 10.12 La translocación inversa envía proteínas al citosol para que sean degradadas 234
- 10.13 Las proteínas residen en las membranas por medio de regiones hidrófobas 235
- 10.14 Las secuencias de anclaje determinan la orientación de las proteínas 236
- 10.15 ¿Cómo se insertan las proteínas en las membranas? 238
- 10.16 La inserción en la membrana después de la traducción depende de las secuencias líder 240
- 10.17 Una jerarquía de secuencias determina la localización dentro de los organelos 241
- 10.18 Las membranas mitocondriales interna y externa tienen traslocones distintos 243
- 10.19 Los peroxisomas emplean otro tipo de sistema de translocación 245
- 10.20 Las bacterias utilizan tanto translocación cotraduccional como translocación postraduccional 246
- 10.21 El sistema Sec transporta proteínas al interior de la membrana interna y a través de ella 247
- 10.22 Sistemas de translocación independientes de Sec en *E. coli* 249
- 10.23 Resumen 250

11 Transcripción 256

- 11.1 Introducción 258
- 11.2 La transcripción ocurre por medio de apareamiento de bases en una "burbuja" de DNA no apareado 259
- 11.3 La reacción de la transcripción consiste en tres etapas 260
- 11.4 La polimerasa de RNA del fago T7 es un sistema de modelos útil 261
- 11.5 La estructura cristalina sugiere un modelo de desplazamiento enzimático 262
- 11.6 La polimerasa de RNA bacteriana está formada por múltiples subunidades 265
- 11.7 La polimerasa de RNA está formada por la enzima central y por un factor σ 267
- 11.8 La asociación con el factor σ cambia en la iniciación 267
- 11.9 Una polimerasa de RNA varada puede reiniciar la transcripción 269
- 11.10 ¿Cómo encuentra una polimerasa de RNA las secuencias promotoras? 270
- 11.11 El factor σ controla la unión al DNA 271
- 11.12 El reconocimiento del promotor depende de las secuencias de consenso 272

- 11.13 La eficiencia de los promotores puede incrementar o disminuir por medio de mutaciones 274
- 11.14 La polimerasa de RNA se une a una cara del DNA 275
- 11.15 El superenrollamiento es una característica importante de la transcripción 277
- 11.16 La sustitución de los factores σ puede controlar la iniciación 278
- 11.17 Los factores σ entran en contacto de manera directa con el DNA 280
- 11.18 Los factores σ pueden organizarse en cascadas 282
- 11.19 La esporulación es controlada por factores σ 283
- 11.20 La polimerasa de RNA bacteriana termina en sitios discretos 286
- 11.21 Hay dos tipos de terminadores en *E. coli* 287
- 11.22 ¿Cómo funciona el factor ρ ? 288
- 11.23 La antiterminación es un episodio regulador 291
- 11.24 La antiterminación requiere sitios que son independientes de los terminadores 292
- 11.25 Los factores de terminación y de antiterminación interactúan con la polimerasa de RNA 293
- 11.26 Resumen 295

12 El operón 300

- 12.1 Introducción 302
- 12.2 La regulación puede ser positiva o negativa 303
- 12.3 Los agrupamientos de genes estructurales son controlados de manera coordinada 304
- 12.4 Los genes *lac* son controlados por un represor 305
- 12.5 El operón *lac* puede ser inducido 305
- 12.6 El represor es controlado por una pequeña molécula inductora 307
- 12.7 Las mutaciones constitutivas de actuación en *cis* identifican al operador 308
- 12.8 Las mutaciones de actuación en *trans* identifican al gen regulador 309
- 12.9 Las proteínas multiméricas tienen propiedades genéticas especiales 309
- 12.10 El monómero represor tiene numerosos dominios 310
- 12.11 Un represor es un tetrámero formado por dos dímeros 311
- 12.12 La unión al DNA es regulada por un cambio alostérico en la conformación 312
- 12.13 Los fenotipos mutantes se correlacionan con la estructura del dominio 312
- 12.14 La proteína represora se une al operador 313
- 12.15 La unión del inductor libera al represor del operador 314
- 12.16 El represor se une a tres operadores e interactúa con la polimerasa de RNA 315
- 12.17 El represor siempre está unido al DNA 316

- 12.18 El operador compete con los sitios de baja afinidad para unirse al represor 317
- 12.19 La represión puede suceder en múltiples loci 319
- 12.20 El AMP cíclico es un efector que activa al factor CRP para que actúe en múltiples operones 320
- 12.21 El factor CRP funciona de formas diferentes en operones diana distintos 321
- 12.22 La traducción puede ser regulada 323
- 12.23 La síntesis de las proteínas r es controlada por medio de regulación autógena 325
- 12.24 La p32 del fago T4 es controlada por un circuito autógeno 326
- 12.25 La regulación autógena se utiliza frecuentemente para controlar la síntesis de ensamblajes macromoleculares 327
- 12.26 Resumen 328

13 RNA regulador 331

- 13.1 Introducción 332
- 13.2 Las estructuras secundarias alternativas controlan la atenuación 333
- 13.3 La terminación de los genes *trp* de *Bacillus subtilis* es controlada por el triptófano y por el RNAT^{trp} 333
- 13.4 El operón triptófano de *Escherichia coli* es controlado por medio de atenuación 335
- 13.5 La atenuación puede ser controlada por la traducción 336
- 13.6 El RNA antisentido puede ser utilizado para desactivar la expresión génica 338
- 13.7 Las moléculas pequeñas de RNA son capaces de regular la traducción 339
- 13.8 Las bacterias contienen RNA reguladores 341
- 13.9 Los microRNA son reguladores en numerosas eucariotas 342
- 13.10 La interferencia de RNA está relacionada con el silenciamiento de los genes 343
- 13.11 Resumen 345

14 Estrategias de los fagos 349

- 14.1 Introducción 350
- 14.2 El desarrollo lítico se divide en dos periodos 352
- 14.3 El desarrollo lítico es controlado por una cascada 353
- 14.4 La cascada lítica es controlada por dos tipos de sucesos reguladores 354
- 14.5 Los genomas de los fagos T4 y T7 presentan agrupación funcional 355
- 14.6 Los genes λ tempranos inmediatos y tempranos retrasados son indispensables para la lisogenia y para el ciclo lítico 356

- 14.7 El ciclo lítico depende de la antiterminación 357
- 14.8 La lisogenia la mantiene una proteína represora 359
- 14.9 El represor y sus operadores definen la región de inmunidad 360
- 14.10 La forma de unión al DNA del represor es un dímero 361
- 14.11 El represor utiliza un elemento hélice-giro-hélice para unirse al DNA 362
- 14.12 La hélice de reconocimiento determina la especificidad por el DNA 363
- 14.13 Los dímeros represores se unen en colaboración al operador 364
- 14.14 El represor en O_{p2} interactúa con la polimerasa de RNA en el P_{RM} 365
- 14.15 El represor mantiene un circuito autógeno 366
- 14.16 Las interacciones en colaboración incrementan la sensibilidad de la regulación 367
- 14.17 Los genes *cII* y *cIII* son necesarios para establecer la lisogenia 368
- 14.18 Un mal promotor requiere proteína *cII* 369
- 14.19 La lisogenia requiere numerosos sucesos 369
- 14.20 El represor *Cro* es necesario para la infección lítica 371
- 14.21 ¿Qué determina el balance entre la lisogenia y el ciclo lítico? 373
- 14.22 Resumen 374

15 El replicón 376

- 15.1 Introducción 377
- 15.2 Los replicones pueden ser lineales o circulares 378
- 15.3 Es posible elaborar el mapa de los orígenes con autoradiografía y electroforesis 379
- 15.4 ¿Regula la metilación del origen la iniciación? 380
- 15.5 Los orígenes pueden ser secuestrados después de la replicación 381
- 15.6 Cada cromosoma eucariótico contiene numerosos replicones 383
- 15.7 En las levaduras pueden aislarse los orígenes de replicación 384
- 15.8 El factor de competencia controla a la replicación eucariótica 385
- 15.9 El factor de competencia está formado por proteínas MCM 386
- 15.10 Los lazos D mantienen a los orígenes mitocondriales 388
- 15.11 Resumen 389

16 Replicones extracromosómicos 392

- 16.1 Introducción 393

- 16.2 Los extremos del DNA lineal representan un problema para la replicación 393
- 16.3 Las proteínas terminales permiten la iniciación en los extremos de los DNA víricos 394
- 16.4 Los círculos rodantes producen multímeros de un replicón 396
- 16.5 Los círculos rodantes se utilizan para replicar los genomas de los fagos 397
- 16.6 El plásmido F es transferido por conjugación entre bacterias 398
- 16.7 La conjugación transfiere DNA de cadena individual 400
- 16.8 El plásmido bacteriano Ti provoca la enfermedad de agalla de Crown en las plantas 401
- 16.9 El T-DNA porta los genes necesarios para la infección 402
- 16.10 La transferencia del T-DNA es semejante a la conjugación bacteriana 405
- 16.11 Resumen 407

17 La replicación bacteriana está conectada con el ciclo celular 408

- 17.1 Introducción 409
- 17.2 La replicación está conectada con el ciclo celular 410
- 17.3 El septo divide a una bacteria en bacterias hijas que contienen cada una un cromosoma 411
- 17.4 Las mutaciones de la división o la segregación alteran la forma de la célula 412
- 17.5 El producto FtsZ es necesario para la formación del septo 413
- 17.6 Los genes *min* regulan la localización del septo 415
- 17.7 La segregación cromosómica puede requerir de recombinación específica de sitio 415
- 17.8 La partición implica a la separación de los cromosomas 417
- 17.9 Los plásmidos de una sola copia tienen un sistema de partición 419
- 17.10 La incompatibilidad de los plásmidos depende del replicón 421
- 17.11 El sistema de compatibilidad ColE1 es controlado por un regulador de RNA 422
- 17.12 ¿Cómo se replican y segregan las mitocondrias? 424
- 17.13 Resumen 425

18 Replicación del DNA 428

- 18.1 Introducción 429
- 18.2 Las polimerasas de DNA son enzimas que sintetizan DNA 430

- 18.3 Las polimerasas de DNA tienen varias actividades de nucleasa 431
- 18.4 Las polimerasas de DNA controlan la fidelidad de la replicación 432
- 18.5 Las polimerasas de DNA tienen una estructura común 433
- 18.6 La síntesis de DNA es semidiscontinua 434
- 18.7 El modelo ϕ X muestra cómo se genera el DNA de cadena individual para la replicación 435
- 18.8 La actividad cebadora es necesaria para iniciar la síntesis de DNA 437
- 18.9 La holoenzima polimerasa de DNA tiene tres subcomplejos 439
- 18.10 La pinza controla la asociación de la enzima central con el DNA 440
- 18.11 Coordinación de la síntesis de la cadena líder y de la cadena retrasada 442
- 18.12 Los fragmentos de Okazaki están unidos por la ligasa 443
- 18.13 Polimerasas de DNA eucarióticas independientes realizan la iniciación y la elongación 444
- 18.14 El fago T4 proporciona su propio mecanismo de replicación 445
- 18.15 Creación de las horquillas de replicación en el origen 448
- 18.16 Sucesos comunes en el cebamiento de la replicación en el origen 450
- 18.17 El primosoma es necesario para reiniciar la replicación 451
- 18.18 Resumen 453

19 Recombinación homóloga y específica de sitio 457

- 19.1 Introducción 459
- 19.2 Ocorre recombinación homóloga entre cromosomas en sinapsis 460
- 19.3 Rotura y reunión afectan el DNA heterodúplex 462
- 19.4 Las roturas de la doble cadena inician la recombinación 464
- 19.5 Los cromosomas en recombinación se conectan por el complejo sinaptonémico 465
- 19.6 El complejo sinaptonémico se forma después de roturas de la doble cadena 467
- 19.7 El apareamiento y la formación del complejo sinaptonémico son independientes 469
- 19.8 Las secuencias *chi* estimulan el sistema RecBCD bacteriano 470
- 19.9 Las proteínas de transferencia de cadena catalizan la asimilación de una sola cadena 471

- 19.10 El sistema Ruv resuelve las uniones Holliday 473
- 19.11 La conversión génica contribuye a la recombinación interalélica 475
- 19.12 El superenrollamiento afecta la estructura del DNA 476
- 19.13 Las topoisomerasas relajan o introducen superhélices en el DNA 478
- 19.14 Las topoisomerasas rompen y resellan cadenas 480
- 19.15 La girasa funciona por la inversión de hélice 481
- 19.16 La recombinación especializada involucra sitios específicos 482
- 19.17 La recombinación específica de sitio comprende rotura y reunión 484
- 19.18 La recombinación específica de sitio simula la actividad de la topoisomerasa 484
- 19.19 La recombinación λ ocurre en un intasoma 486
- 19.20 Las levaduras pueden cambiar loci silentes y activos por el tipo de apareamiento 488
- 19.21 El locus *MAT* codifica proteínas reguladoras 490
- 19.22 Se reprimen los cartuchos silentes en *HML* y *HMR* 492
- 19.23 El locus *MAT* receptor inicia la transposición unidireccional 493
- 19.24 La regulación de la expresión de *H0* controla el cambio 494
- 19.25 Resumen 496

20 Sistemas de reparación 499

- 20.1 Introducción 500
- 20.2 Los sistemas de reparación corrigen el daño del DNA 502
- 20.3 Sistemas de reparación por escisión en *E. coli* 503
- 20.4 Vías de reparación por escisión en células de mamíferos 504
- 20.5 Las metilasas y las glucosilasas "lanzan" las bases 506
- 20.6 Reparación susceptible de error y fenotipos mutadores 507
- 20.7 Control de la dirección de la reparación de apareamientos erróneos 507
- 20.8 Sistemas de reparación por recombinación en *E. coli* 510
- 20.9 La recombinación es un mecanismo importante de recuperación ante errores de replicación 511
- 20.10 La proteína RecA desencadena el sistema SOS 513
- 20.11 Las células eucarióticas tienen sistemas de reparación conservados 515
- 20.12 Un sistema común repara roturas de la doble cadena 516
- 20.13 Resumen 518

21 Transposones 521

- 21.1 Introducción 522
- 21.2 Las secuencias de inserción son simples módulos de transposición sencillos 524
- 21.3 Los transposones compuestos tienen módulos IS 525
- 21.4 La transposición ocurre por mecanismos replicativos y no replicativos 527
- 21.5 Los transposones causan reestructuración del DNA 528
- 21.6 Intermediarios comunes para la transposición 530
- 21.7 La transposición replicativa avanza a través de un cointegrado 531
- 21.8 La transposición no replicativa avanza por rotura y reunión 533
- 21.9 La transposición de TnA requiere transposasa y resolvasa 534
- 21.10 La transposición de Tn10 tiene múltiples controles 534
- 21.11 Los elementos de control en el maíz causan roturas y reestructuraciones 538
- 21.12 Los elementos de control forman familias de transposones 540
- 21.13 Los elementos Spm influyen en la expresión génica 542
- 21.14 Intervención de los elementos susceptibles de transposición en la disgenesia híbrida 544
- 21.15 Los elementos P se activan en la línea germinal 545
- 21.16 Resumen 546

22 Retrovirus y retroposones 550

- 22.1 Introducción 551
- 22.2 El ciclo vital de los retrovirus comprende sucesos similares a la transposición 551
- 22.3 Los genes retrovíricos codifican poliproteínas 552
- 22.4 El DNA vírico se genera por transcripción inversa 554
- 22.5 El DNA vírico se integra al cromosoma 556
- 22.6 Los retrovirus pueden hacer transducción de secuencias celulares 558
- 22.7 Los elementos *Ty* de las levaduras se asemejan a los retrovirus 559
- 22.8 Muchos elementos susceptibles de transposición residen en la *Drosophila melanogaster* 561
- 22.9 Los retroposones son de tres clases 562
- 22.10 La familia Alu tiene muchos miembros muy dispersos 564
- 22.11 Los pseudogenes procesados se originaron como sustratos para la transposición 565
- 22.12 Las LINES utilizan una endonucleasa para generar un extremo con actividad cebadora 566
- 22.13 Resumen 567

23 Diversidad inmunitaria 570

- 23.1 Introducción 572
- 23.2 La selección clonal amplifica linfocitos que responden a antígenos individuales 574
- 23.3 Los genes de las inmunoglobulinas se ensamblan en los linfocitos a partir de sus constituyentes 575
- 23.4 Las cadenas ligeras se ensamblan por recombinación simple 577
- 23.5 Las cadenas pesadas se ensamblan mediante dos recombinaciones 579
- 23.6 La recombinación genera una gran diversidad 580
- 23.7 La recombinación inmunitaria utiliza dos tipos de secuencias de consenso 581
- 23.8 La recombinación genera deleciones o inversiones 582
- 23.9 La exclusión alélica es desencadenada por rearreglo productivo 582
- 23.10 Las proteínas RAG catalizan la rotura y la nueva unión 584
- 23.11 La expresión temprana de las cadenas pesadas puede cambiar por procesamiento del RNA 586
- 23.12 El cambio de clase es producto de la recombinación del DNA 587
- 23.13 El cambio se debe a una nueva reacción de recombinación 589
- 23.14 La mutación somática genera otras diferencias entre el ratón y el ser humano 590
- 23.15 La desaminasa de citidina y la glucosilasa de uracilo inducen la mutación somática 591
- 23.16 Las inmunoglobulinas aviaarias se ensamblan a partir de pseudogenes 593
- 23.17 La célula B de memoria permite una respuesta secundaria rápida 594
- 23.18 Los receptores de las células T tienen relación con las inmunoglobulinas 595
- 23.19 El receptor de la célula T actúa en unión con el MHC 597
- 23.20 El locus de histocompatibilidad mayor codifica muchos genes del sistema inmunitario 599
- 23.21 La inmunidad innata hace uso de vías de señal conservadas 602
- 23.22 Resumen 604

24 Promotores y potenciadores 609

- 24.1 Introducción 610
- 24.2 Las polimerasas de RNA eucarióticas constan de muchas subunidades 612
- 24.3 Los elementos del promotor se definen por mutaciones y vestigios 613
- 24.4 La polimerasa de RNA I tiene un promotor bipartido 614

- 24.5 La polimerasa de RNA III utiliza promotores en flujo ascendente y descendente 615
- 24.6 $TF_{III}B$ es el factor de encargo de los promotores de Pol III 616
- 24.7 El punto de inicio de la polimerasa de RNA II 618
- 24.8 La TBP es un factor universal 619
- 24.9 La TBP se une al DNA de forma inusual 620
- 24.10 El aparato basal se ensambla en el promotor 621
- 24.11 El inicio va seguido de la depuración del promotor 623
- 24.12 Una conexión entre transcripción y reparación 625
- 24.13 Los elementos de secuencia corta se unen a activadores 627
- 24.14 La estructura del promotor es flexible, pero el contexto puede ser importante 628
- 24.15 Los potenciadores contienen elementos bidireccionales que facilitan el inicio 629
- 24.16 Los potenciadores contienen los mismos elementos encontrados en los promotores 630
- 24.17 Los potenciadores actúan aumentando la concentración de activadores cerca del promotor 631
- 24.18 La expresión genética se vincula con la desmetilación 632
- 24.19 Los islotes CpG son dianas reguladoras 634
- 24.20 Resumen 635

25 Activación de la transcripción 640

- 25.1 Introducción 641
- 25.2 Hay varios tipos de factores de transcripción 642
- 25.3 Dominios independientes se unen con el DNA y activan la transcripción 643
- 25.4 El análisis de dos híbridos detecta interacciones proteína-proteína 645
- 25.5 Los activadores interactúan con el aparato basal 646
- 25.6 Algunas proteínas de unión con promotor corresponden a represores 648
- 25.7 Los elementos de respuesta son reconocidos por los activadores 649
- 25.8 Hay muchos tipos de dominios de unión con DNA 651
- 25.9 El segmento de dedo de zinc es un dominio de unión con DNA 652
- 25.10 Los receptores de esteroides son activadores 653
- 25.11 Los receptores de esteroides tienen dedos de zinc 655
- 25.12 La unión con el elemento de respuesta se activa por unión con un ligando 656
- 25.13 Los receptores de esteroides reconocen a los elementos de respuesta por un código combinatorio 657
- 25.14 Los homeodominos se unen con dianas relacionadas en el DNA 658

- 25.15 Las proteínas hélice-asa-hélice interactúan por vínculo combinatorio 658
- 25.16 Las cremalleras de Levana participan en la formación de dímeros 658
- 25.17 Resumen 658

26 Corte, empalme y procesamiento del RNA 667

- 26.1 Introducción 669
- 26.2 Las uniones de corte y empalme nuclear son secuencias cortas 670
- 26.3 Las uniones de corte y empalme se leen por pares 671
- 26.4 El corte y empalme del pre-RNA_m procede a través de un lazo 673
- 26.5 Se requieren RNAs_n para el corte y empalme 674
- 26.6 La RNPSn U1 inicia el corte y empalme 676
- 26.7 El complejo E puede formarse por definición de intrón o exón 678
- 26.8 Cinco RNPSn forman el empalmosoma 679
- 26.9 Un aparato alternativo de corte y empalme utiliza diferentes RNPSn 681
- 26.10 El corte y empalme está relacionado con la exportación del RNA_m 682
- 26.11 Los intrones del grupo II se cortan y empalman a sí mismos por la formación de un lazo 683
- 26.12 El proceso de corte y empalme alternativo implica el uso diferencial de uniones de corte y empalme 865
- 26.13 Las reacciones de corte y empalme en configuración *trans* utilizan RNA pequeños 688
- 26.14 El corte y empalme del RNA_t de las levaduras implica escisión y reunión 690
- 26.15 La endonucleasa de corte y empalme reconoce al RNA_t 691
- 26.16 La escisión y ligadura del RNA_t son reacciones separadas 692
- 26.17 La respuesta de la proteína no plegada tiene relación con el corte y empalme del RNA_t 693
- 26.18 Los extremos 3' de los productos de transcripción pol I y pol III se generan por terminación 694
- 26.19 Los extremos 3' del RNA_m se generan por escisión y poliadenilación 695
- 26.20 La escisión del extremo 3' del RNA_m de histonas puede requerir de un RNA pequeño 697
- 26.21 La producción de RNA_r requiere de sucesos de escisión 697
- 26.22 Se requieren RNA pequeños para el procesamiento del RNA_r 699
- 26.23 Resumen 700

27 RNA catalítico 706

- 27.1 Introducción 707
- 27.2 Los intrones del grupo I realizan el autocorte y empalme por transesterificación 707
- 27.3 Los intrones del grupo I forman una estructura secundaria característica 709
- 27.4 Las ribozimas tienen varias actividades catalíticas 711
- 27.5 Algunos intrones del grupo I codifican a endonucleasas que fomentan la movilidad 715
- 27.6 Los intrones del grupo II pueden codificar proteínas multifuncionales 716
- 27.7 Algunos intrones de autocorte y empalme requieren de madurasas 717
- 27.8 La actividad catalítica de la ribonucleasa P se debe al RNA 718
- 27.9 Los viroides tienen actividad catalítica 718
- 27.10 La edición del RNA ocurre en bases individuales 720
- 27.11 La edición del RNA puede ser dirigida por RNA guías 721
- 27.12 El corte y empalme de las proteínas son autocatalíticos 724
- 27.13 Resumen 725

28 Cromosomas 729

- 28.1 Introducción 730
- 28.2 Los genomas víricos están empaquetados en sus cubiertas 731
- 28.3 El genoma bacteriano es un nucleoide 734
- 28.4 El genoma bacteriano está superenrollado 735
- 28.5 El DNA eucariótico tiene asas y dominios unidos a un bastidor 736
- 28.6 Secuencias específicas unen el DNA a una matriz de interfase 737
- 28.7 La cromatina se divide en eucromatina y heterocromatina 738
- 28.8 Los cromosomas tienen patrones de bandas 740
- 28.9 Los cromosomas en escobillón se extienden 741
- 28.10 Los cromosomas politénicos forman bandas 742
- 28.11 Los cromosomas politénicos se expanden en sitios de expresión génica 743
- 28.12 El cromosoma eucariótico es un dispositivo de segregación 744
- 28.13 Los centrómeros pueden contener DNA repetitivo 746
- 28.14 Las secuencias de DNA de los centrómeros de *S. cerevisiae* son cortas 747
- 28.15 El centrómero se une a un complejo proteínico 748
- 28.16 Los telómeros tienen secuencias de repetición sencillas 748
- 28.17 Los telómeros sellan los extremos de los cromosomas 749

- 28.18 Los telómeros se sintetizan mediante una enzima ribonucleoproteínica 750
- 28.19 Los telómeros son esenciales para la supervivencia 752
- 28.20 Resumen 753

29 Nucleosomas 757

- 29.1 Introducción 758
- 29.2 El nucleosoma es la subunidad de la cromatina 759
- 29.3 El DNA está enrollado en la estructura de los nucleosomas 761
- 29.4 Los nucleosomas tienen una estructura común 762
- 29.5 La estructura del DNA varía en la superficie del nucleosoma 763
- 29.6 La periodicidad del DNA cambia en el nucleosoma 766
- 29.7 Organización del octámero de histonas 767
- 29.8 La vía de los nucleosomas en la fibra de cromatina 769
- 29.9 La replicación de la cromatina implica el ensamblaje de los nucleosomas 771
- 29.10 ¿Los nucleosomas yacen en posiciones específicas? 774
- 29.11 ¿Los genes transcritos se organizan en nucleosomas? 777
- 29.12 Los octámeros de histonas son desplazados por la transcripción 779
- 29.13 El desplazamiento del nucleosoma y su reensamblado requieren factores especiales 781
- 29.14 Los aislantes impiden la acción de los potenciadores y de la heterocromatina 781
- 29.15 Los aislantes pueden definir un dominio 783
- 29.16 los aislantes pueden actuar en una dirección 784
- 29.17 Los aislantes pueden variar en potencia 785
- 29.18 Los sitios hipersensibles a la desoxirribonucleasa reflejan cambios en la estructura de la cromatina 786
- 29.19 Los dominios definen regiones que contienen genes activos 788
- 29.20 Una LCR puede controlar a un dominio 789
- 29.21 ¿Qué constituye un dominio regulador? 790
- 29.22 Resumen 791

30 Control de la estructura de la cromatina 796

- 30.1 Introducción 797
- 30.2 La cromatina puede tener estados alternativos 797
- 30.3 El remodelado de la cromatina es un proceso activo 798
- 30.4 La organización de los nucleosomas puede cambiar en el promotor 801

- 30.5 La modificación de las histonas es un episodio clave 802
- 30.6 Ocurre acetilación de histonas en dos circunstancias 805
- 30.7 Las acetilasas se relacionan con activadores 806
- 30.8 Las desacetilasas se relacionan con los represores 808
- 30.9 Las metilaciones de histonas y de DNA están relacionadas 808
- 30.10 Los estados de la cromatina se interconvierten por modificación 809
- 30.11 La activación del promotor comprende una serie ordenada de episodios 809
- 30.12 La fosforilación de las histonas afecta la estructura de la cromatina 810
- 30.13 Se encuentran algunos segmentos comunes en las proteínas que modifican a la cromatina 811
- 30.14 Resumen 812

31 Los efectos epigenéticos son heredados 818

- 31.1 Introducción 819
- 31.2 La heterocromatina se propaga a partir de un episodio de nucleación 820
- 31.3 La heterocromatina depende de interacciones con las histonas 822
- 31.4 Polycomb y Trithorax son represores y activadores antagonistas 824
- 31.5 Los cromosomas X experimentan cambios globales 826
- 31.6 Las condensinas producen la condensación de los cromosomas 828
- 31.7 Una metilasa de mantenimiento perpetúa la metilación del DNA 830
- 31.8 La metilación del DNA se encarga de la impresión 832
- 31.9 Un solo centro puede controlar los genes con impresión opuesta 834
- 31.10 Los efectos epigenéticos pueden heredarse 835
- 31.11 Los priones de levaduras muestran herencia desusada 836
- 31.12 Los priones causan enfermedades en los mamíferos 839
- 31.13 Resumen 840

Glosario 845

Índice alfabético 867

Prefacio

La ciencia es una aventura extraordinariamente difícil. En cada revisión de esta obra hay nuevos e interesantes acontecimientos de qué reportar, de modo que ésta incluye mucho material actualizado que da cuenta de nuevos hallazgos de la investigación molecular. La organización general del material de esta edición ha sido revisada según los criterios de *Genes esenciales* para facilitar el uso conjunto de las dos obras. Como el aumento de tamaño representaba un problema, el contenido se ha enfocado con mayor precisión en los genes y su expresión, de modo que se eliminaron los capítulos que tratan de las consecuencias de la expresión génica en la biología celular. La primera parte del libro incluye cambios sobresalientes relacionados con el genoma derivados de exitosos proyectos de secuenciación genómica; resalta la importancia del RNA como regulador, y ahora es evidente que cubre todos los niveles de expresión génica, tanto en procariontes como en eucariontes. Algo así como un "eslabón perdido", arroja más luz sobre la forma en que el mecanismo actual de la expresión génica debió haber evolucionado desde el mundo primitivo regido por el RNA.

En esta obra me he propuesto citar artículos de revisión y de investigación, que creo que serán de fácil acceso para los lectores; preferí aquellos que después de seis meses no representan ningún costo, y cuando no fue posible, la publicación debería ser de amplia difusión.

Agradezco a las siguientes personas el haber leído las pruebas y el haberme asesorado en esta revisión:

Elliott Goldstein	University of Arizona, Tempe
Jocelyn Krebs	University of Alaska, Anchorage
Kathleen Matthews	Rice University, Houston

Benjamin Lewin
Enero de 2007

Organización

La nueva organización de *GENES IX* permite a instructores y estudiantes enfocarse con mayor especificidad en los genes y su expresión, con énfasis en los temas principales. El número de capítulos y el orden de los temas permanecen igual; sin embargo, numerosos capítulos se convirtieron en dos o más. Los cambios son los siguientes:

El capítulo 1 de *GENES VIII*, Los genes son DNA, se convirtió en dos capítulos en *GENES IX*. La información básica de la estructura del DNA, la replicación y las mutaciones permanecen en el capítulo 1, mientras que la descripción de la función de los genes como unidad de la herencia constituyen el nuevo capítulo 2, Los genes codifican proteínas.

El capítulo 3 de *GENES VIII*, El contenido del genoma, se divide en dos capítulos en *GENES IX*.

El capítulo 4, El contenido del genoma, incluye información sobre secuencias de DNA, mapeo genómico y DNA de organelos.

El capítulo 5, Secuencias genómicas y números de genes, ahora contiene información sobre las dimensiones y la expresión de los genomas de gran número de organismos, así como material nuevo sobre los genes del cromosoma Y.

El nuevo capítulo 12, El operón, incluye el capítulo 10 de *GENES VIII*, así como información sobre la regulación de la transcripción y de la traducción del capítulo 11 de *GENES VIII*, Circuitos reguladores. El material que corresponde al RNA regulador se encuentra ahora en el capítulo 13.

El material del capítulo 13 de *GENES VIII*, El replicón, en *GENES IX* se encuentra en tres capítulos. El capítulo 15, El replicón, abarca la estructura y la función del replicón, así como los orígenes de la replicación. El capítulo 16, Replicones extracromosómicos, contiene material sobre las proteínas terminales, el círculo de re-

plicación, los plásmidos y el T-DNA. La información sobre la forma en que la replicación bacteriana se vincula con el ciclo celular se encuentra en el capítulo 17.

Recombinación y reparación, capítulo 15 de *GENES VIII*, son dos capítulos de *GENES IX*. El capítulo 19 cubre la recombinación homóloga y específica de sitio, y el 20, Sistemas de reparación, incluye información novedosa sobre las rutas de reparación por escisión en las células de mamíferos.

El capítulo 23, Control de la estructura de la cromatina, de *GENES VIII* ahora es el capítulo 30, en el cual

se analiza la relación entre la estructura de la cromatina y la expresión génica.

El capítulo 31, Los efectos epigenéticos son heredados, detalla las causas y los mecanismos de la herencia epigenética.

Programa de arte y diseño

GENES IX luce ahora más moderno. El diseño y la parte artística de esta edición se actualizaron y revisaron para facilitar el aprendizaje de los estudiantes.

Los genes son DNA

ESQUEMA DEL CAPÍTULO

1.1 Introducción

1.2 El ácido desoxirribonucleico (DNA, *deoxyribonucleic acid*) es el material genético de las bacterias

- La transformación bacteriana proporcionó la primera prueba de que el DNA es el material genético de las bacterias. Las propiedades genéticas pueden ser transferidas de una cepa bacteriana a otra extrayendo el DNA de la primera cepa y añadiéndolo a la segunda.

1.3 El DNA es el material genético de los virus

- La infección por fagos demostró que el DNA es el material genético de los virus. Cuando el DNA y los componentes proteínicos de los bacteriófagos son marcados con isótopos radioactivos diferentes, sólo el DNA es transmitido a la progenie de fagos producida por las bacterias infectuosas.

1.4 El DNA es el material genético de las células animales

- El DNA puede ser utilizado para introducir nuevas características genéticas en células animales o en animales completos.
- En algunos virus, el material genético es el RNA.

1.5 Los polinucleótidos tienen bases nitrogenadas ligadas a un esqueleto de azúcar-fosfato

- Un nucleósido está formado por una base púrica o pirimidica ligada a una pentosa (azúcar de cinco carbonos) en la posición 1.
- Las posiciones en el anillo de ribosa se describen con el símbolo matemático de primo (¹), para distinguirlas.
- La diferencia entre el DNA y el RNA radica en el grupo localizado en la posición 2' del azúcar. El DNA tiene un azúcar desoxirriboso (2'-H); el RNA tiene un azúcar riboso (2'-OH).
- Un nucleótido consiste en un nucleósido ligado a un grupo fosfato en la posición 5' o 3' de la (desoxi)ribosa.
- Los residuos sucesivos de (desoxi)ribosa de una cadena de polinucleótidos están unidos por un grupo fosfato entre la posición 3' de un azúcar y la 5' del siguiente.
- Un extremo de la cadena (convencionalmente el izquierdo) tiene una punta 5' libre y, en el otro, la punta libre es 3'.
- El DNA contiene las cuatro bases, adenina, guanina, citosina y timina; el RNA posee uracilo en vez de timina.

1.6 El DNA es una hélice dúplex

- La forma B del DNA es una hélice dúplex formada por dos cadenas polinucleótídicas que corren de forma antiparalela.

- Las bases nitrogenadas de cada cadena son anillos planos de purina o de pirimidina orientados hacia el interior que se aparean uno a otro a través de puentes de hidrógeno para formar únicamente pares de A-T o G-C.
- El diámetro de la hélice dúplex es de 20 Å, además de una vuelta completa cada 34 Å, con diez pares de bases por vuelta.
- La hélice dúplex forma un surco profundo (ancho) y un surco superficial (angosto).

1.7 La duplicación del DNA es semiconservadora

- En el experimento de Meselson-Stahl se utilizó marcaje de densidad para demostrar que la cadena polinucleotídica individual es la unidad del DNA que se conserva durante la duplicación.
- Cada cadena de un DNA dúplex actúa como molde para sintetizar a una cadena hija.
- Las secuencias de las cadenas hijas están determinadas por apareamiento complementario de bases con las cadenas progenitoras separadas.

1.8 Las cadenas de DNA se separan en la horquilla de duplicación

- La duplicación de DNA es emprendida por un complejo de enzimas que separan a las cadenas progenitoras y que sintetizan a las cadenas hijas.
- La horquilla de duplicación es el punto en el cual las cadenas progenitoras se separan.
- Las enzimas que sintetizan DNA se denominan DNA polimerasas; las enzimas que sintetizan RNA son conocidas como RNA polimerasas.
- Las nucleasas son enzimas que degradan a los ácidos nucleicos; entre otras, DNAsas y RNAsas, que pueden dividirse en endonucleasas y en exonucleasas.

1.9 La información genética puede ser proporcionada por el DNA o el RNA

- Los genes celulares son DNA, pero los virus y los viroides pueden tener genomas de RNA.
- El DNA se convierte en RNA por transcripción, y el RNA puede convertirse en DNA por transcripción inversa.
- La traducción del RNA a proteínas es unidireccional.

1.10 Los ácidos nucleicos se hibridan por apareamiento de bases

- El calentamiento provoca que las dos cadenas de un DNA dúplex se separen.
- La T_m es el punto medio del rango de temperatura para la desnaturalización.
- Las cadenas complementarias únicas pueden renaturalizarse cuando se reduce la temperatura.

Continúa en la siguiente página

- La desnaturalización y la renaturalización/hibridación pueden ocurrir con combinaciones de DNA-DNA, de DNA-RNA o de RNA-RNA y pueden ser intermoleculares o intramoleculares.
- La capacidad de dos preparaciones de ácidos nucleicos de una sola cadena para hibridarse es indicio de su complementariedad.

1.11 Las mutaciones cambian la secuencia del DNA

- Todas las mutaciones consisten en cambios de la secuencia del DNA.
- Las mutaciones pueden ser espontáneas o inducidas por mutágenos.

1.12 Las mutaciones pueden afectar a pares de bases individuales o a secuencias más largas

- Una mutación puntual cambia a un solo par de bases.
- Las mutaciones puntuales pueden ser provocadas por la conversión química de una base en otra o por errores durante la duplicación.
- Una transición reemplaza a un par de bases G-C por un par de bases A-T o viceversa.
- Una transversión reemplaza a una purina por una pirimidina, por ejemplo, cambiando A-T a T-A.
- Las inserciones son el tipo más común de mutación, y son resultado del movimiento de elementos transponibles.

1.13 Los efectos de las mutaciones pueden ser revertidos

- Las mutaciones primarias desactivan a un gen, mientras que las inversas (o retrorrotaciones) revierten sus efectos.

- Las inserciones pueden revertirse por delección del material insertado, pero las delecciones no pueden revertirse.
- La supresión ocurre cuando una mutación de un segundo gen anula el efecto de la mutación del primer gen.

1.14 Las mutaciones se concentran en puntos calientes

- La frecuencia de la mutación en cualquier par de bases en particular depende de fluctuaciones estadísticas, excepto en los puntos calientes, en los cuales la frecuencia se incrementa cuando menos en un orden de magnitud.

1.15 Numerosos puntos calientes son resultado de bases modificadas

- Una causa común de los puntos calientes es la base modificada 5-metilcitosina, la cual experimenta desaminación espontánea y se convierte en timina.

1.16 Algunos agentes hereditarios son extremadamente pequeños

- Algunos agentes hereditarios muy pequeños no codifican proteínas, sino que constan de RNA o proteínas que poseen propiedades hereditarias.

1.17 Resumen

1.1 Introducción

La naturaleza hereditaria de todo organismo viviente es definida por su **genoma**, el cual consiste en una secuencia larga de **ácido nucleico** que proporciona la **información** necesaria para construir el organismo. Se utiliza el término "información" debido a que el genoma por sí mismo no desempeña ninguna función activa en la construcción del organismo, más bien es la secuencia de las subunidades individuales (bases) del ácido nucleico la que determina las características hereditarias. Por una compleja serie de interacciones, esta secuencia se utiliza para producir todas las proteínas del organismo en el momento y el lugar apropiados. Las proteínas forman parte de la estructura del organismo o tienen la capacidad de construir estructuras o llevar a cabo las reacciones metabólicas necesarias para la vida.

El genoma contiene el conjunto completo de información hereditaria para cualquier organismo. Físicamente, el genoma puede ser dividido en varias moléculas diferentes de ácido nucleico y funcionalmente, en **genes**. Cada gen es una secuencia que dentro del ácido nucleico representa a una sola proteína. Cada una de las moléculas de ácido nucleico que componen el genoma puede contener

gran número de genes. Los genomas de los organismos vivos pueden contener incluso <500 genes (como un micoplasma, que es un tipo de bacteria) o >25 000 en el ser humano.

En este capítulo se analizan las propiedades del gen en cuanto a su construcción molecular básica. En la **FIGURA 1.1** se resumen las etapas de transición del concepto histórico del gen a la definición moderna del genoma.

Un genoma consiste en el conjunto completo de cromosomas de cualquier organismo específico, de modo que comprende una serie de moléculas de ácido desoxirribonucleico (DNA, *deoxyribonucleic acid*) (una para cada cromosoma), cada una de las cuales contiene numerosos genes. La principal definición de un genoma es determinar la secuencia del DNA de cada cromosoma.

La primera definición del gen como unidad funcional se derivó del descubrimiento de que cada gen es responsable de la producción de proteínas específicas. La diferencia en las características químicas del DNA del gen y su producto proteínico condujo al concepto de que un gen *codifica* a una proteína. A su vez, esto llevó al descubrimiento del aparato complejo que permite que la secuencia de DNA de un gen genere la secuencia de aminoácidos de una proteína.



FIGURA 1.1 Breve historia de la genética.

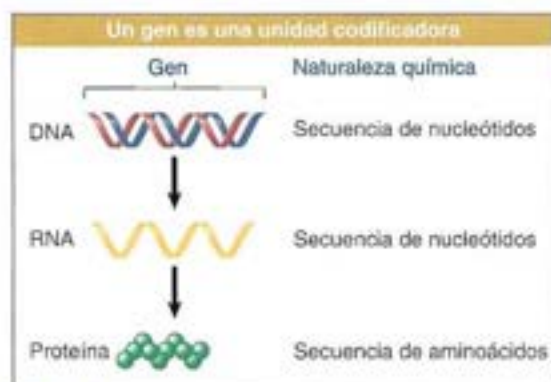


FIGURA 1.2 Un gen codifica una molécula de RNA, la cual codifica una proteína.

Entendiendo el proceso por medio del cual un gen se expresa, permite una definición más rigurosa de su naturaleza. En la FIGURA 1.2 se ilustra el tema básico de esta obra. Un gen es una secuencia de DNA que produce otro ácido nucleico, el ácido ribonucleico (RNA, *ribonucleic acid*), que cuenta con dos cadenas de ácido nucleico, mientras que el RNA tiene una sola. La secuencia del RNA es determinada por la secuencia del DNA. (De hecho, es idéntica a una de las cadenas de DNA.) En muchos casos, pero no en todos, el RNA se utiliza a su vez para dirigir la producción de una proteína. *Por lo tanto,*

un gen es una secuencia de DNA que codifica a un RNA; en los genes codificadores de proteínas, el RNA a su vez codifica a una proteína.

A partir de la demostración de que un gen consiste en DNA, y de que un cromosoma consiste en un tramo largo de DNA que representa a numerosos genes, se llega a la organización global del genoma en función de su secuencia de DNA. En el Capítulo 3, El gen interrumpido, se retoma en más detalle la organización del gen y su representación en proteínas. En el Capítulo 4, El contenido del genoma, se analiza el número total de genes, y en el Capítulo 6, Agrupamientos y repeticiones, se describen otros componentes del genoma y el mantenimiento de su organización.

1.2 El DNA es el material genético de las bacterias

Concepto principal

- La transformación bacteriana proporcionó la primera prueba de que el DNA es el material genético de las bacterias. Las propiedades genéticas pueden ser transferidas de una cepa bacteriana a otra extrayendo el DNA de la primera cepa y sumándolo a la segunda.

La idea de que el material genético es ácido nucleico tiene sus raíces en el descubrimiento de la **transformación**, en 1928. La bacteria *Pneumococcus* mata de neumonía a los ratones. La virulencia de la bacteria depende de su *polisacárido capsular*, componente de la superficie que permite que la bacteria escape de la destrucción provocada por su hospedador. Diversos tipos de *Pneumococcus* (I, II y III) tienen polisacáridos capsulares diferentes; su aspecto es liso (*S. smooth*).

Todos los tipos de *Pneumococcus* lisos puede dar lugar a variantes incapaces de producir el polisacárido capsular, en cuyo caso, la superficie de la bacteria es *rugosa* (*R. rough*) (formada por el material que estaba debajo del polisacárido capsular). Estas bacterias son **avirulentas**, no matan a los ratones porque la ausencia del polisacárido permite que el animal destruya a la bacteria.

Cuando por medio de tratamiento calórico se matan bacterias lisas, pierden la capacidad de dañar al animal, sin embargo, la combinación de bacterias S desactivadas con calor y la variante ineficaz R produce un efecto considerablemente diferente del de cada bacteria por separado. En la FIGURA 1.3 se muestra que cuando se inyectan conjuntamente en un animal, el ratón muere como resultado de una infección por *Pneumococcus*. Las bacterias S virulentas pueden ser recuperadas del ratón muerto.

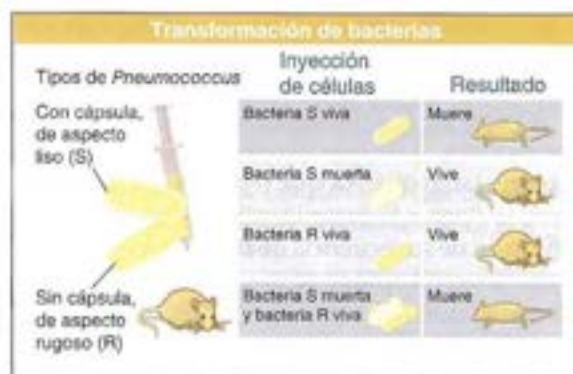


FIGURA 1.3 Ni las bacterias tipo S asesinadas con calor ni las bacterias tipo R vivas pueden matar a los ratones, pero la inyección simultánea de ambos tipos de bacterias puede matar a los ratones de forma tan efectiva como la bacteria tipo S viva.

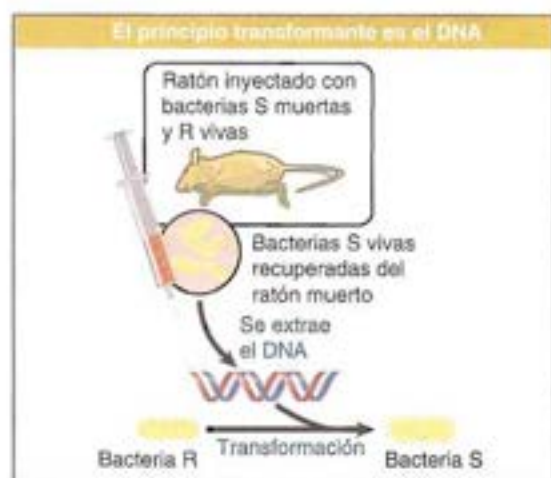


FIGURA 1.4 El DNA de la bacteria de tipo S puede transformar a la bacteria de tipo R en el mismo tipo S.

En este experimento, las bacterias S muertas fueron de tipo III y las R vivas, de tipo II. La cubierta de las bacterias virulentas recuperadas de la infección mixta era lisa, de tipo III, por lo tanto, alguna propiedad de las bacterias S tipo III, muertas, puede *transformar* a las bacterias R, induciendo la formación del polisacárido capsular tipo III, que las torna virulentas.

En la **FIGURA 1.4** se muestra la identificación del componente de las bacterias muertas responsable de la transformación, el cual se denominó **principio transformante**, que se purificó desarrollando un sistema libre de células, en donde los extractos de las bacterias S muertas podían agregarse a las bacterias R vivas antes de ser inyectadas al animal. En 1994, mediante la purificación del principio trans-

formante, se demostró que éste es **ácido desoxirribonucleico (DNA, deoxyribonucleic acid)**.

1.3 El DNA es el material genético de los virus

Concepto principal

- La infección por fagos demostró que el DNA es el material genético de los virus. Cuando el DNA y los componentes proteínicos de los bacteriófagos son marcados con diferentes isótopos radioactivos, sólo el DNA es transmitido a la progenie de fagos producida por las bacterias infecciosas.

Una vez que se comprobó que el DNA es el material genético de las bacterias, el siguiente paso fue demostrar que el DNA proporciona el material genético en un sistema bastante diferente. El fago T2 es un virus que infecta a la bacteria *Escherichia coli*. Cuando las partículas del fago (virus) se agregan a las bacterias, éstas se adsorben en la superficie exterior, parte del material entra en la bacteria y, al cabo de ~20 min, cada bacteria explota (se lisa) para liberar una numerosa progenie de fagos.

En la **FIGURA 1.5** se ilustran los resultados de un experimento realizado en 1952 en el cual se infectaron bacterias con fagos T2, los cuales habían sido marcados radioactivamente en su componente de DNA (con ^{32}P) o en su componente proteínico (con ^{35}S). Las bacterias infectadas se mezclaron en una licuadora y dos fracciones se separaron por centrifugación. Una de éstas contenía las cubiertas de fagos vacías liberadas de la superficie de la bacteria, en tanto que la otra estaba formada por las bacterias infectadas.*

Gran parte del marcador ^{32}P se encontró en las bacterias infectadas. Las partículas de la progenie de fagos producidas por la infección contenían ~30% del marcador ^{32}P original. La progenie recibió una cantidad muy pequeña, menos del 1%, de la proteína contenida en la población original de fagos. Las cubiertas de fago estaban formadas por proteínas, y por lo tanto portaban el marcador radioactivo ^{35}S . Así pues, con este experimento se demostró directamente que sólo el DNA de los fagos progenitores entra en las bacterias y que después formará parte de los fagos de la progenie, exactamente el patrón de herencia que se espera del material genético.

* N del T: este párrafo implica que los virus fueron marcados en un componente o en otro (pero no en los dos), mientras que la descripción de los resultados sugiere que tanto un componente como el otro fueron marcados.

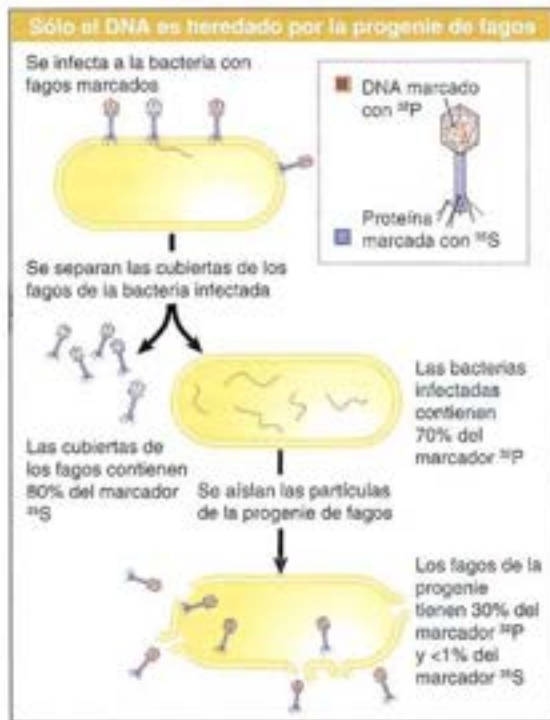


FIGURA 1.5 El material genético del fago T2 es el DNA.

Un fago se reproduce ordenando a la maquinaria de una célula hospedadora infectada que produzca más copias de sí misma. El fago posee material genético cuyo comportamiento es análogo al de los genomas celulares: sus rasgos son fielmente reproducidos y están sujetos a las mismas reglas que gobiernan la herencia. El caso de los T2 refuerza la conclusión general de que el material genético es el DNA, ya sea que forme parte del genoma de una célula o de un virus.

1.4 El DNA es el material genético de las células animales

Conceptos principales

- El DNA puede ser utilizado para introducir nuevas características genéticas en células animales o en animales completos.
- En algunos virus, el material genético es RNA.

Cuando se agrega DNA a poblaciones de células eucariotas individuales en cultivo, el ácido nucleico entra a las células, lo cual, en algunas, resulta en la producción de proteínas nuevas. Cuando se utiliza DNA purificado, su incorporación da lugar a la producción de una proteína en particular. En la FIGURA 1.6 se representa uno de los sistemas estándar.

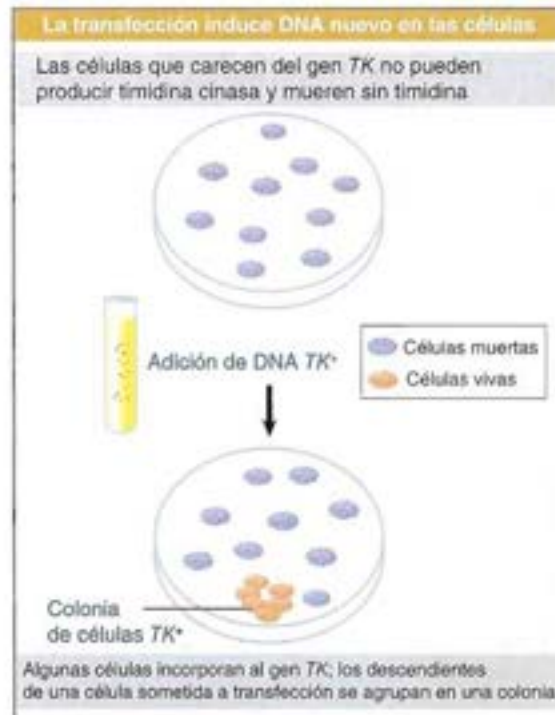


FIGURA 1.6 Las células eucariotas pueden adquirir un fenotipo nuevo como resultado de la transfección por adición de DNA.

Aunque por razones históricas estos experimentos se describen como **transfección** cuando son realizados con células eucariotas, constituyen una contraparte directa de la transformación bacteriana. El DNA introducido en la célula receptora se incorpora a su material genético y es heredado en la misma forma que cualquier otra parte. Su expresión confiere un nuevo rasgo a las células (síntesis de timidina-cinasa en el ejemplo de la Figura 1.6). En un principio, estos experimentos sólo tuvieron éxito con células individuales adaptadas para crecer en un medio de cultivo, pero desde entonces, el DNA ha sido introducido en óvulos de ratón por microinyección, y puede llegar a ser una parte estable del material genético del animal.

Dichos experimentos muestran directamente no sólo que el DNA es el material genético de las eucariotas, sino también, que *puede ser transferido entre especies diferentes y seguir siendo funcional*.

El material genético de todos los organismos conocidos y de numerosos virus es el DNA. No obstante, algunos virus usan un tipo alternativo de ácido nucleico, el **ribonucleico (RNA, ribonucleic acid)**, como material genético. El principio general de la naturaleza del material genético, por lo tanto, es que siempre es ácido nucleico; de hecho, es DNA, excepto en los virus de RNA.

Los polinucleótidos tienen bases nitrogenadas ligadas a un esqueleto de azúcar-fosfato

Conceptos principales

- Un nucleósido está formado por una base púrica o pirimídica ligada a un azúcar pentosa en la posición 1.
- Las posiciones en el anillo de ribosa se describen con el símbolo (') para distinguirlas.
- La diferencia entre el DNA y el RNA radica en el grupo localizado en la posición 2' del azúcar. El DNA tiene un azúcar desoxirribosa (2'-H) y el RNA, un azúcar ribosa (2'-OH).
- Un nucleótido consta de un nucleósido ligado a un grupo fosfato en la posición 5' o 3' de la (desoxi)ribosa.
- Los residuos sucesivos de (desoxi)ribosa de un polinucleótido están unidos por un grupo fosfato entre la posición 3' de un azúcar y la posición 5' de la siguiente.
- Un extremo de la cadena (convencionalmente el izquierdo) tiene una punta 5' libre y en el otro, una punta 3' libre.
- El DNA contiene las cuatro bases, adenina, guanina, citosina y timina; el RNA posee uracilo en vez de timina.

El componente básico de los ácidos nucleicos es el nucleótido, que tiene tres componentes:

- una base nitrogenada,
- un azúcar y
- un fosfato.

La base nitrogenada es un anillo de purina o de pirimidina. La base está ligada a la posición 1 en una pentosa por medio de un enlace glucosídico del N₁ de las pirimidinas o el N₉ de la purinas. Para evitar ambigüedades entre los sistemas de numeración de los anillos heterocíclicos y del azúcar, a las posiciones de la pentosa se agrega un símbolo (').

Los ácidos nucleicos son nombrados según el tipo de azúcar que contienen; el DNA posee 2'-desoxirribosa, mientras que el RNA tiene ribosa. La diferencia radica en que el azúcar del RNA tiene un grupo OH en la posición 2' del anillo de pentosa. El azúcar puede ligarse en las posiciones 5' o 3' a un grupo fosfato.

Un ácido nucleico consiste en una cadena larga de nucleótidos. En la FIGURA 1.7 se muestra que el esqueleto de la cadena polinucleotídica consta de series alternas de residuos de fosfato y de pentosa (azúcar) a través de la unión de la posición 5' de un anillo de pentosa con la posición 3' del siguiente anillo de pentosa mediante un grupo fosfato. Por

eso se dice que el esqueleto de azúcar-fosfato consiste en enlaces 5'-3' fosfodiéster. Las bases nitrogenadas "sobresalen" del esqueleto.

Cada ácido nucleico contiene cuatro tipos de bases. Las mismas dos purinas, adenina y guanina, están presentes tanto en el DNA como en el RNA. Las dos pirimidinas del DNA son citosina y timina; en el RNA se encuentra uracilo, en vez de timina. La única diferencia entre el uracilo y la timina es un grupo metilo suplantador en la posición C₅. Las bases suelen nombrarse por sus iniciales. El DNA contiene A, G, C y T; el RNA contiene A, G, C y U.

El nucleótido terminal de un extremo de la cadena tiene un grupo 5' libre; el nucleótido terminal del otro extremo tiene un grupo 3' libre. Se ha convenido en escribir las secuencias de ácido nucleico en dirección 5' a 3', es decir, del extremo 5' de la izquierda al extremo 3' de la derecha.

El DNA es una hélice dúplex

Conceptos principales

- La forma B del DNA es una hélice dúplex formada por dos cadenas polinucleotídicas antiparalelas.
- Las bases nitrogenadas de cada una son anillos planos de purina o pirimidina orientados hacia el interior y que se aparean mediante puentes de hidrógeno para formar únicamente pares de A-T y G-C.
- El diámetro de la hélice dúplex es de 20 Å, y cada 34 Å hay una vuelta completa, con diez pares de bases por vuelta.
- La hélice dúplex forma un surco principal profundo (ancho) y uno menor (angosto).

La observación de que las cantidades de bases presentes en los DNA varía según la especie, condujo al concepto de que *la secuencia de bases es la forma en que se porta la información genética*. Hacia la década de 1950, el concepto de información genética era común: el par de problemas que planteaba era determinar la estructura del ácido nucleico y explicar cómo una secuencia de bases del DNA podía representar la secuencia de aminoácidos de una proteína.

Tres nociones convergieron en la construcción de un modelo de hélice dúplex del DNA, desarrollado por Watson y Crick en 1953:

- Los datos de la difracción de los rayos X demostraron que el DNA tiene la forma de una hélice regular que da un giro completo cada 34 Å (3.4 nm), con un diámetro de ~20 Å (2 nm). Como la distancia entre nucleótidos adyacentes es de 3.4 Å, debe haber 10 nucleótidos por vuelta.

- La densidad del DNA sugiere que la hélice debe contener dos cadenas de polinucleótidos. El diámetro constante de la hélice puede explicarse si las bases de cada cadena apuntan hacia adentro y son restringidas de manera que una purina siempre se oponga a una pirimidina, evitando asociaciones purina-purina (demasiado ancha) o pirimidina-pirimidina (demasiado angosta).
- Independientemente de las cantidades absolutas de cada base, la proporción de G es siempre igual a la de C en el DNA, y la proporción de A siempre es la misma que la de T, de manera que la composición de cualquier DNA puede describirse por la proporción de sus bases, es decir, G + C, que fluctúa entre 26 y 74% en diferentes especies.

Watson y Crick propusieron que las dos cadenas polinucleotídicas de la hélice dúplex se relacionan mediante *puentes de hidrógeno entre las bases nitrogenadas*. La G puede formar puentes de hidrógeno específicamente sólo con la C, y la A, sólo con la T. Estas reacciones se describen como **apareamiento de bases**, y se dice que las bases apareadas (G con C o A con T) son **complementarias**.

El modelo proponía que las dos cadenas polinucleotídicas corren en direcciones opuestas (**antiparalelas**), como se ilustra en la **FIGURA 1.8**. Observando a lo largo de la hélice, una cadena va en dirección 5' a 3', mientras que su pareja corre de 3' a 5'.

El esqueleto de azúcar-fosfato está en la parte exterior y porta cargas negativas en los grupos fosfato. En solución *in vitro*, las cargas del DNA se neutralizan por la unión de iones metálicos, típicamente Na⁺. En la célula, las proteínas con carga positiva proporcionan parte de la fuerza neutralizante. Estas proteínas desempeñan una función importante en la determinación de la organización del DNA en la célula.

Las bases se encuentran en el interior: son estructuras planas, en pares perpendiculares al eje de la hélice. Considérese la hélice dúplex como una escalera espiral: los pares de bases forman los escalones, como se ilustra esquemáticamente en la **FIGURA 1.9**. Al ir avanzando por la hélice, las bases están una sobre otra, como una pila de platos.

Cada par de bases gira ~36° en torno al eje de la hélice, respecto del siguiente par de bases, de modo que ~10 pares de bases forman una vuelta completa de 360°. La torsión de las cadenas sobre sí mismas forma una hélice dúplex con un **surco menor** (~12 Å de ancho) y un **surco mayor** (~22 Å de ancho), como puede observarse en el modelo a escala de la **FIGURA 1.10**. La hélice dúplex es **dextrógira**, es decir, que gira en la dirección de las manecillas del

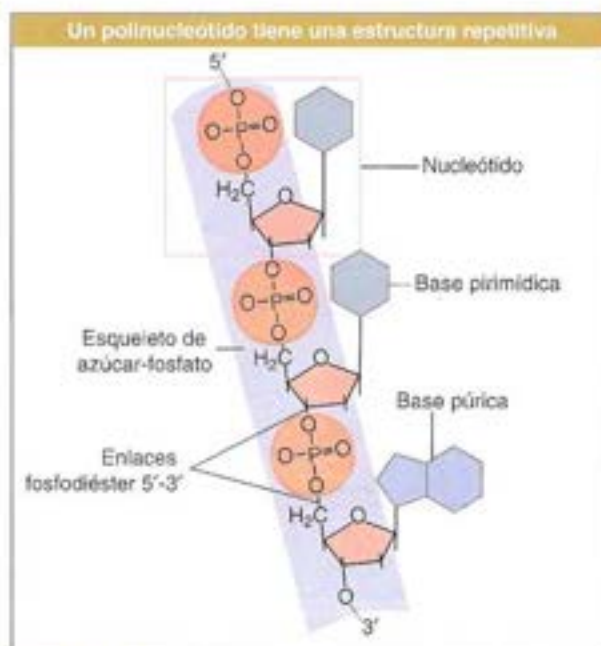


FIGURA 1.7 Una cadena polinucleotídica está formada por una serie de enlaces azúcar-fosfato 5'-3' que forman un esqueleto con las bases prominentes.

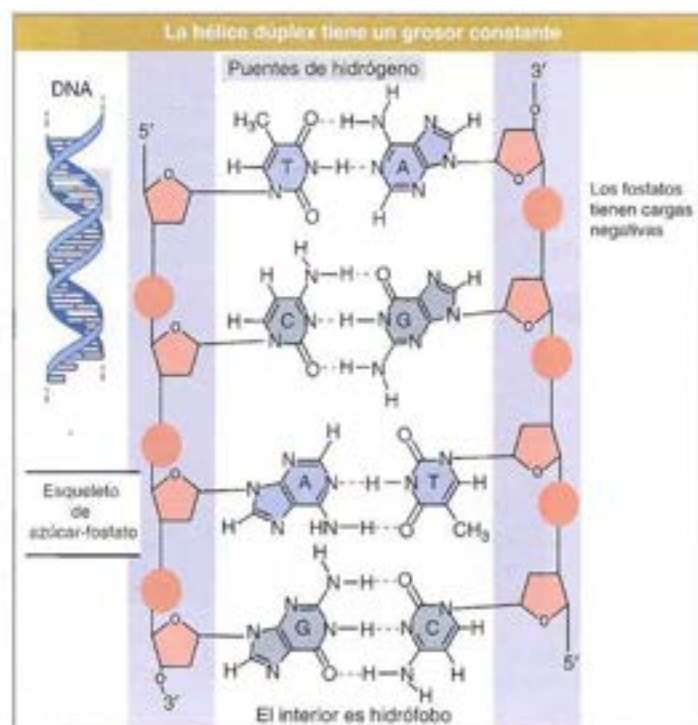


FIGURA 1.8 La hélice dúplex mantiene un grosor constante porque las purinas siempre enfrentan a las pirimidinas en los pares de bases complementarios A-T y G-C. La secuencia que se ilustra en la figura es T-A, C-G, A-T, G-C.

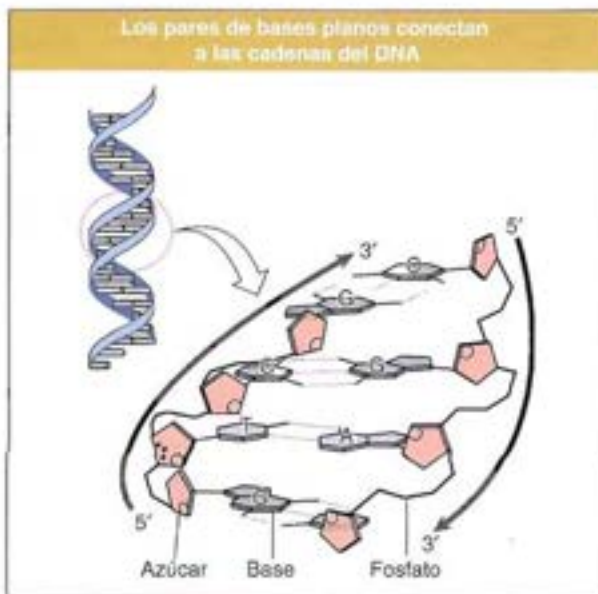


FIGURA 1.9 Los pares de bases planos yacen perpendicularmente al esqueleto de azúcar-fosfato.

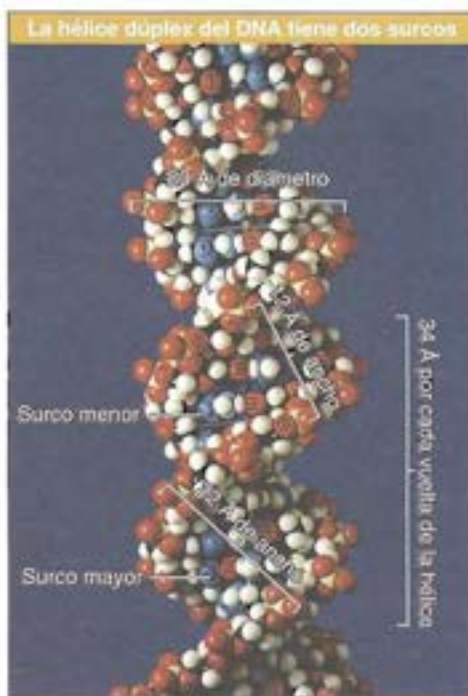


FIGURA 1.10 Las dos cadenas del DNA forman una hélice dúplex. Fotografía © Photodisc.

reloj observada a lo largo del eje helicoidal. Estas características representan el modelo aceptado para lo que se conoce como **forma B** del DNA.

Es importante percatarse de que la forma B representa un *promedio*, no una estructura específica con toda precisión, pues la estructura del DNA puede cambiar localmente. Si tiene más pares de bases por vuelta se dice que está **sobregirada**, de lo contrario, será **laxa**. La torción local puede ser afectada por la conformación global de la hélice dúplex del DNA en el espacio o por la unión de proteínas en sitios específicos.

1.7 La duplicación del DNA es semiconservadora

Conceptos principales

- En el experimento de Meselson-Stahl se utilizó marcaje de densidad para demostrar que la cadena polinucleotídica individual es la unidad del DNA que se conserva durante la duplicación.
- Cada cadena de un DNA dúplex (o DNA duplohelicoidal, DNA de doble cadena o DNA de doble banda) hace las veces de molde para sintetizar una cadena hija.
- Las secuencias de las cadenas hijas dependen del apareamiento complementario de las bases con las cadenas progenitoras separadas.

Es crucial que el material genético se reproduzca con exactitud. Las dos cadenas polinucleotídicas están unidas sólo por puentes de hidrógeno, de modo que pueden separarse sin necesidad de romper enlaces covalentes. La especificidad del apareamiento de bases sugiere que cada una de las cadenas **progenitoras** separadas puede hacer las veces de **cadena molde** para la síntesis de una **cadena hija** complementaria. En la **FIGURA 1.11** se representa el principio de que una nueva cadena hija es ensamblada en cada cadena progenitora. La secuencia de la cadena hija es dictada por la cadena progenitora; una **A** en la cadena progenitora ocasiona la colocación de una **T** en la cadena hija, en tanto que una **G** progenitora provoca la incorporación de una **C** hija, y así sucesivamente.

En la parte superior de la Figura 1.11 se muestra un dúplex progenitor (no replicado) formado por las dos cadenas progenitoras originales. La parte inferior muestra las dos dúplex hijas que están siendo producidas por el apareamiento complementario de bases. Cada una de las dúplex hijas es idéntica en secuencia original y contiene una cadena progenitora y una cadena recientemente sintetizada. La



FIGURA 1.11 El apareamiento de bases proporciona el mecanismo necesario para la duplicación del DNA.

estructura del DNA porta la información necesaria para perpetuar su secuencia.

Las consecuencias de este modo de duplicación se ilustran en la FIGURA 1.12. El dúplex progenitor se replica para formar dos dúplex hijos, cada uno de los cuales está formado por una cadena progenitora y una cadena hija (de síntesis reciente). La unidad que se conserva de una generación a la siguiente es una de las dos cadenas individuales que forman el dúplex progenitor. Este comportamiento se conoce como **duplicación semiconservadora**.

En la Figura 1.12 se ilustra una predicción de este modelo. Si el DNA progenitor porta un marcaje de densidad "pesada" porque el organismo ha sido cultivado en un medio que contiene un isótopo adecuado (como ^{15}N) (N del T: en otras bibliografías aparece como N^{15} , el resto de los marcadores mencionados en este capítulo también aparecen con el número después de la letra p. ej., P^{32} y S^{35}), sus cadenas pueden ser distinguidas de las sintetizadas cuando el organismo es transferido a un medio que contenga isótopos "ligeros" normales.

El DNA progenitor consiste en un dúplex de dos cadenas pesadas (rojas). Después de una generación de crecimiento en medio ligero, el DNA dúplex es "híbrido" en cuanto a densidad, o sea que está formado por una cadena progenitora pesada (roja) y por una cadena hija ligera (azul). Después de la segunda generación, las dos cadenas de cada dúplex híbrido se han separado. Cada una adquiere a una pareja ligera, de tal forma que la mitad de los DNA dúplex

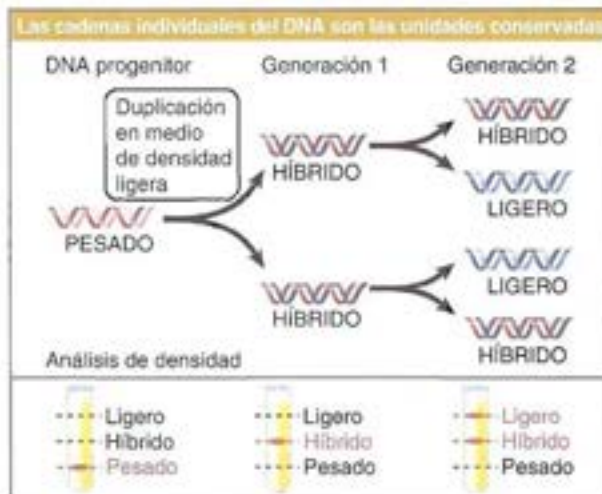


FIGURA 1.12 La duplicación del DNA es semiconservadora.

siguen siendo híbridos y la otra mitad es completamente ligera (ambas cadenas son azules).

Las cadenas individuales de estos dúplex son completamente pesadas o completamente ligeras. Este patrón se confirmó experimentalmente en la prueba de Meselson-Stahl en 1958, que siguió a la duplicación semiconservadora del DNA a través de tres generaciones de crecimiento de *E. coli*. Cuando el DNA fue extraído de las bacterias y su densidad se midió por centrifugación, el DNA formó bandas correspondientes a su densidad, pesada en las progenitoras, híbrida en la primera generación, y mitad híbrida y mitad ligera en la segunda generación.

1.8 Las cadenas del DNA se separan en la horquilla de duplicación

Conceptos principales

- La duplicación del DNA es llevada a cabo por un complejo de enzimas que separan las cadenas progenitoras y sintetizan las cadenas hijas.
- La horquilla de duplicación es el punto en que se separan las cadenas progenitoras.
- Las enzimas que sintetizan el DNA se llaman DNA polimerasas; las enzimas que sintetizan el RNA se conocen como RNA polimerasas.
- Las nucleasas son enzimas que degradan a los ácidos nucleicos; incluyen DNAsas y RNAsas, y pueden dividirse en endonucleasas y en exonucleasas.

La duplicación exige la separación de las dos cadenas de un dúplex progenitor; sin embargo, el rompimiento de la estructura es transitorio, y se revierte conforme se forma el dúplex hijo. En algún mo-

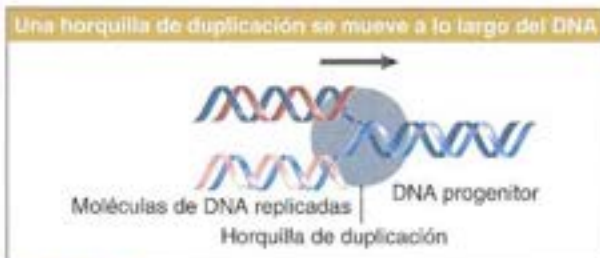


FIGURA 1.13 La horquilla de duplicación es la región del DNA en la cual hay una transición del dúplex progenitor desenrollado a los dúplex hijos recientemente replicados.



FIGURA 1.14 Una endonucleasa divide a un enlace en un ácido nucleico. En este ejemplo se muestra una enzima que ataca a una cadena de un DNA dúplex.



FIGURA 1.15 Una exonucleasa elimina bases, una a la vez, dividiendo el último enlace de una cadena polinucleotídica.

mento, sólo un tramo pequeño del DNA dúplex se separa en cadenas individuales.

La estructura helicoidal de una molécula de DNA dedicada a la duplicación se ilustra en la **FIGURA 1.13**. La región no replicada es el dúplex progenitor que se abre hacia la región replicada, donde se han formado los dos dúplex hijos. La estructura helicoidal doble se rompe en la unión entre las dos regiones, llamada **horquilla de duplicación**. La duplicación implica el movimiento de la horquilla de duplicación a lo largo del DNA progenitor, de modo que hay un desenrollamiento continuo de las cadenas progenitoras y un enrollamiento de los dúplex hijos.

La síntesis de ácidos nucleicos es catalizada por enzimas específicas, las cuales reconocen el molde y emprenden la tarea de catalizar la adición de subunidades a la cadena polinucleotídica que está sien-

do sintetizada. Las enzimas se nombran según el tipo de cadena que sintetizan: las **DNA polimerasas** sintetizan DNA, las **RNA polimerasas**, RNA.

La degradación de los ácidos nucleicos también requiere de enzimas específicas: las **desoxirribonucleasas (DNAsas)** degradan al DNA y las **ribonucleasas (RNAsas)** degradan al RNA. Las nucleasas se incluyen en las clases generales de **exonucleasas** y **endonucleasas**:

- Las endonucleasas cortan enlaces individuales *en el interior* de las moléculas de DNA y de RNA y generan fragmentos discretos. Algunas DNAsas dividen ambas cadenas de un DNA dúplex en el sitio diana, mientras que otras dividen sólo una de ellas. Las endonucleasas participan en reacciones de corte, como se observa en la **FIGURA 1.14**.
- Las exonucleasas eliminan los residuos, uno a uno, a partir del extremo de la molécula, y generan mononucleótidos. Siempre actúan en una sola cadena de ácido nucleico, y cada exonucleasa avanza en una dirección específica, es decir, empieza en un extremo 5' o 3' para dirigirse hacia el otro. Están involucradas en reacciones de ajuste, como se muestra en la **FIGURA 1.15**.

1.9

La información genética puede ser proporcionada por el DNA o el RNA

Conceptos principales

- Los genes celulares son DNA, pero los virus y los viroides pueden tener genomas de RNA.
- El DNA se convierte en RNA por transcripción, y el RNA puede convertirse en DNA por transcripción inversa.
- La traducción del RNA a proteína es unidireccional.

El **dogma central** define el paradigma de la biología molecular. Los genes se perpetúan como secuencias de ácido nucleico, si bien actúan al ser expresados en forma de proteínas. La duplicación es responsable de la herencia de la información genética. La transcripción y la traducción son responsables de la conversión de una forma a otra.

En la **FIGURA 1.16** se ilustran las funciones de la duplicación, la transcripción y la traducción desde la perspectiva del dogma central:

- *La perpetuación del ácido nucleico suele implicar al DNA o el RNA como el material genético.* Las células utilizan únicamente DNA, en tanto que algunos virus usan RNA, y la duplica-

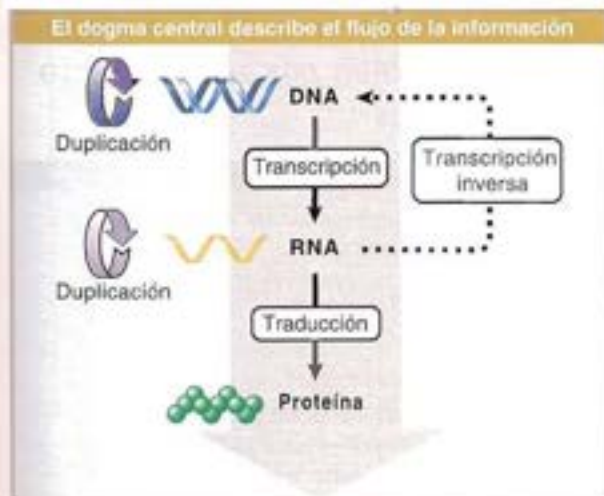


FIGURA 1.16 El dogma central manifiesta que la información del ácido nucleico puede ser perpetuada o transferida, sin embargo la transferencia de la información a proteínas es irreversible.

ción del RNA vírico ocurre en la célula infectada.

- *La expresión de la información genética celular suele ser unidireccional.* La transcripción del DNA genera moléculas de RNA que *únicamente* pueden ser utilizadas otra vez para generar secuencias de proteínas; en general no pueden recuperarse para usarlas como fuente de información genética. La traducción del RNA a proteína siempre es irreversible.

Estos mecanismos son igualmente efectivos para la información genética celular de procariontas o eucariotas y para la información transportada por los virus. Los genomas de todos los organismos vivos están formados por DNA dúplex. Los virus poseen genomas de DNA o RNA, y de cada tipo hay ejemplos de doble cadena (ds, *double stranded*) o de cadena única (ss, *single stranded*). Los detalles del mecanismo de replicación del ácido nucleico varían entre los diferentes sistemas víricos, pero el principio de duplicación por síntesis de cadenas complementarias sigue siendo el mismo, según se ilustra en la **FIGURA 1.17**.

Los genomas celulares reproducen el DNA por el mecanismo de duplicación semiconservadora. Los genomas víricos de doble cadena, sean de DNA o RNA, también se replican utilizando como molde las cadenas individuales del dúplex para sintetizar así las cadenas complementarias.

Los virus con genomas de una sola cadena la utilizan como molde para sintetizar una cadena complementaria, la cual, a su vez, es utilizada para

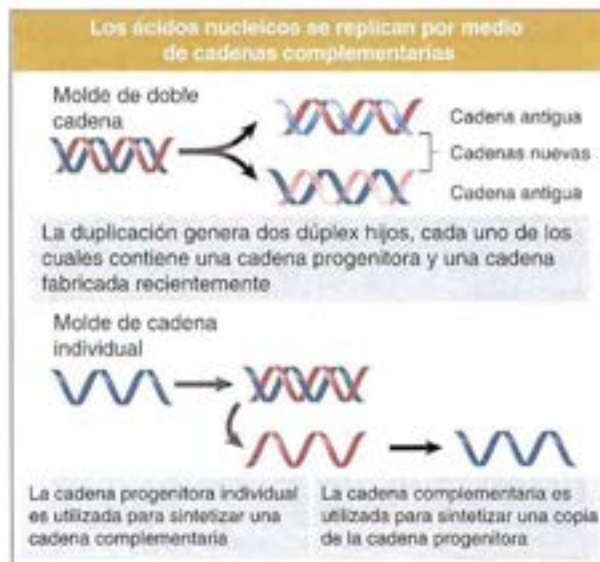


FIGURA 1.17 Los ácidos nucleicos, tanto los de doble cadena como los de cadena individual, se replican por la síntesis de cadenas complementarias regida por las reglas del apareamiento de bases.

sintetizar a su complemento, que, por supuesto, es idéntico a la cadena original inicial. La duplicación puede implicar la formación de intermediarios estables de doble cadena o utilizar ácido nucleico de doble cadena sólo como una fase transitoria.

La restricción de la transferencia unidireccional de DNA a RNA no es absoluta, ha sido superada por los **retrovirus**, cuyos genomas están formados por moléculas de RNA de una sola cadena. Durante el ciclo infeccioso, el RNA se convierte en una cadena única de DNA por el proceso de **transcripción inversa**; dicha cadena, a su vez, es convertida en un DNA de doble banda. Este DNA dúplex se convierte en parte del genoma de la célula y es heredado como cualquier otro gen, *de manera que la transcripción inversa permite que una secuencia de RNA sea recuperada y utilizada como información genética.*

La existencia de la duplicación del RNA y de la transcripción inversa establece el principio general de que *la información en cualquier tipo de secuencia de ácido nucleico puede convertirse en el otro tipo.* Sin embargo, en el curso normal de los hechos, la célula depende de los procesos de duplicación, transcripción y traducción del DNA. No obstante, alguna vez la información de un RNA celular se convierte en DNA y se inserta en el genoma (evento posiblemente mediado por un virus RNA). Aunque la transcripción inversa no participa en las operaciones regulares de la célula, llega a ser un mecanismo potencialmente importante cuando se analiza la evolución del genoma.

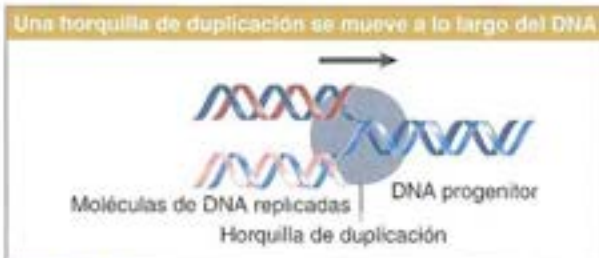


FIGURA 1.13 La horquilla de duplicación es la región del DNA en la cual hay una transición del dúplex progenitor desenrollado a los dúplex hijos recientemente replicados.



FIGURA 1.14 Una endonucleasa divide a un enlace en un ácido nucleico. En este ejemplo se muestra una enzima que ataca a una cadena de un DNA dúplex.

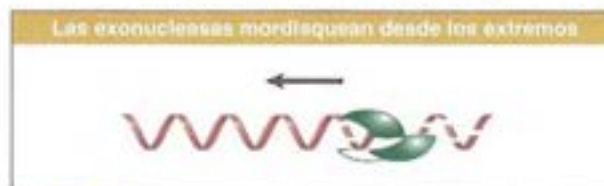


FIGURA 1.15 Una exonucleasa elimina bases, una a la vez, dividiendo el último enlace de una cadena polinucleotídica.

mento, sólo un tramo pequeño del DNA dúplex se separa en cadenas individuales.

La estructura helicoidal de una molécula de DNA dedicada a la duplicación se ilustra en la **FIGURA 1.13**. La región no replicada es el dúplex progenitor que se abre hacia la región replicada, donde se han formado los dos dúplex hijos. La estructura helicoidal doble se rompe en la unión entre las dos regiones, llamada **horquilla de duplicación**. La duplicación implica el movimiento de la horquilla de duplicación a lo largo del DNA progenitor, de modo que hay un desenrollamiento continuo de las cadenas progenitoras y un enrollamiento de los dúplex hijos.

La síntesis de ácidos nucleicos es catalizada por enzimas específicas, las cuales reconocen el molde y emprenden la tarea de catalizar la adición de subunidades a la cadena polinucleotídica que está sien-

do sintetizada. Las enzimas se nombran según el tipo de cadena que sintetizan: las **DNA polimerasas** sintetizan DNA, las **RNA polimerasas**, RNA.

La degradación de los ácidos nucleicos también requiere de enzimas específicas: las **desoxirribonucleasas (DNAs)** degradan al DNA y las **ribonucleasas (RNAs)** degradan al RNA. Las nucleasas se incluyen en las clases generales de **exonucleasas** y **endonucleasas**:

- Las endonucleasas cortan enlaces individuales *en el interior* de las moléculas de DNA y de RNA y generan fragmentos discretos. Algunas DNAs dividen ambas cadenas de un DNA dúplex en el sitio diana, mientras que otras dividen sólo una de ellas. Las endonucleasas participan en reacciones de corte, como se observa en la **FIGURA 1.14**.
- Las exonucleasas eliminan los residuos, uno a uno, a partir del extremo de la molécula, y generan mononucleótidos. Siempre actúan en una sola cadena de ácido nucleico, y cada exonucleasa avanza en una dirección específica, es decir, empieza en un extremo 5' o 3' para dirigirse hacia el otro. Están involucradas en reacciones de ajuste, como se muestra en la **FIGURA 1.15**.

1.9 La información genética puede ser proporcionada por el DNA o el RNA

Conceptos principales

- Los genes celulares son DNA, pero los virus y los viroídeos pueden tener genomas de RNA.
- El DNA se convierte en RNA por transcripción, y el RNA puede convertirse en DNA por transcripción inversa.
- La traducción del RNA a proteína es unidireccional.

El **dogma central** define el paradigma de la biología molecular. Los genes se perpetúan como secuencias de ácido nucleico, si bien actúan al ser expresados en forma de proteínas. La duplicación es responsable de la herencia de la información genética. La transcripción y la traducción son responsables de la conversión de una forma a otra.

En la **FIGURA 1.16** se ilustran las funciones de la duplicación, la transcripción y la traducción desde la perspectiva del dogma central:

- *La perpetuación del ácido nucleico suele implicar al DNA o el RNA como el material genético.* Las células utilizan únicamente DNA, en tanto que algunos virus usan RNA, y la duplica-

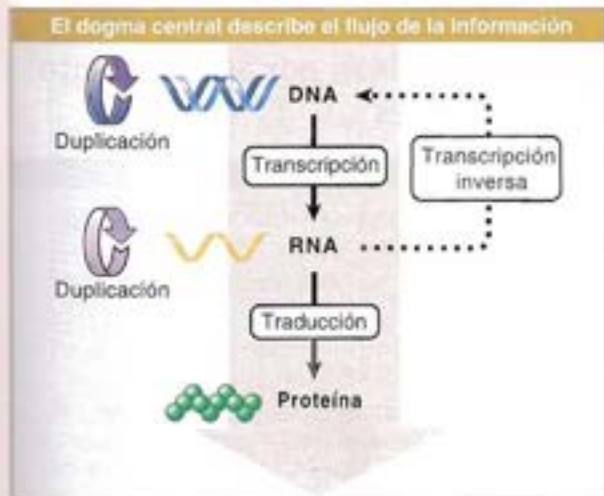


FIGURA 1.16 El dogma central manifiesta que la información del ácido nucleico puede ser perpetuada o transferida, sin embargo la transferencia de la información a proteínas es irreversible.

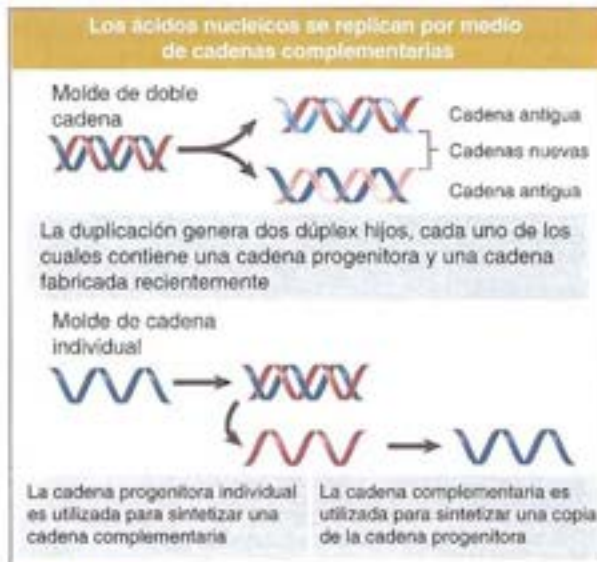


FIGURA 1.17 Los ácidos nucleicos, tanto los de doble cadena como los de cadena individual, se replican por la síntesis de cadenas complementarias regida por las reglas del apareamiento de bases.

ción del RNA vírico ocurre en la célula infectada.

- *La expresión de la información genética celular suele ser unidireccional.* La transcripción del DNA genera moléculas de RNA que *únicamente* pueden ser utilizadas otra vez para generar secuencias de proteínas; en general no pueden recuperarse para usarlas como fuente de información genética. La traducción del RNA a proteína siempre es irreversible.

Estos mecanismos son igualmente efectivos para la información genética celular de procariontas o eucariotas y para la información transportada por los virus. Los genomas de todos los organismos vivos están formados por DNA dúplex. Los virus poseen genomas de DNA o RNA, y de cada tipo hay ejemplos de doble cadena (ds, *double stranded*) o de cadena única (ss, *single stranded*). Los detalles del mecanismo de replicación del ácido nucleico varían entre los diferentes sistemas víricos, pero el principio de duplicación por síntesis de cadenas complementarias sigue siendo el mismo, según se ilustra en la **FIGURA 1.17**.

Los genomas celulares reproducen el DNA por el mecanismo de duplicación semiconservadora. Los genomas víricos de doble cadena, sean de DNA o RNA, también se replican utilizando como molde las cadenas individuales del dúplex para sintetizar así las cadenas complementarias.

Los virus con genomas de una sola cadena la utilizan como molde para sintetizar una cadena complementaria, la cual, a su vez, es utilizada para

sintetizar a su complemento, que, por supuesto, es idéntico a la cadena original inicial. La duplicación puede implicar la formación de intermediarios estables de doble cadena o utilizar ácido nucleico de doble cadena sólo como una fase transitoria.

La restricción de la transferencia unidireccional de DNA a RNA no es absoluta, ha sido superada por los **retrovirus**, cuyos genomas están formados por moléculas de RNA de una sola cadena. Durante el ciclo infeccioso, el RNA se convierte en una cadena única de DNA por el proceso de **transcripción inversa**; dicha cadena, a su vez, es convertida en un DNA de doble banda. Este DNA dúplex se convierte en parte del genoma de la célula y es heredado como cualquier otro gen, *de manera que la transcripción inversa permite que una secuencia de RNA sea recuperada y utilizada como información genética.*

La existencia de la duplicación del RNA y de la transcripción inversa establece el principio general de que *la información en cualquier tipo de secuencia de ácido nucleico puede convertirse en el otro tipo.* Sin embargo, en el curso normal de los hechos, la célula depende de los procesos de duplicación, transcripción y traducción del DNA. No obstante, alguna vez la información de un RNA celular se convierte en DNA y se inserta en el genoma (evento posiblemente mediado por un virus RNA). Aunque la transcripción inversa no participa en las operaciones regulares de la célula, llega a ser un mecanismo potencialmente importante cuando se analiza la evolución del genoma.

Los genomas varían enormemente en cuanto a tamaño		
Genoma	Número de genes	Pares de bases
Organismos		
Plantas	<50 000	<10 ¹¹
Mamíferos	30 000	~3 x 10 ⁹
Gusanos	14 000	~10 ⁸
Moscas	12 000	1,6 x 10 ⁸
Hongos	6 000	1,3 x 10 ⁷
Bacterias	2-4 000	<10 ⁷
Micoplasma	500	<10 ⁶
Virus DNAds		
Vaccinia	<300	187 000
Papova (SV40)	-6	5 226
Fago T4	-200	165 000
Virus DNAss		
Parvovirus	5	5 000
Fago φX174	11	5 387
Virus RNAds		
Reovirus	22	23 000
Virus RNAss		
Coronavirus	7	20 000
Influenza	12	13 500
TMV	4	6 400
Fago MS2	4	3 569
STNV	1	1 300
Viroides		
PSTV RNA	0	359

FIGURA 1.18 La cantidad de ácido nucleico del genoma varía en un rango enorme. ds, doble cadena; ss, cadena única.

Los mismos principios se aplican a la perpetuación de la información genética tanto en los genomas enormes de plantas o anfibios como en los genomas diminutos del micoplasma y en la información genética aún más pequeña de los virus de DNA o RNA. En la **FIGURA 1.18** se resumen algunos ejemplos de la gama de tipos y tamaños de los genomas.

En toda la gama de organismos, cuyo contenido total de genomas varía en un rango superior a las 100 000 veces, prevalece un principio común: *el DNA codifica a todas las proteínas que la(s) célula(s) del organismo debe(n) sintetizar, y las proteínas, a su vez (directa o indirectamente), proporcionan las funciones necesarias para la supervivencia.* Mediante un principio similar se describe la función de la información genética de los virus, ya sea de DNA o RNA: *el ácido nucleico codifica a la(s) proteína(s) necesaria(s) para empaquetar al genoma y también para cualquier función adicional a las proporcionadas por la célula hospedadora necesaria para la reproducción del virus durante su ciclo infeccioso.* (El virus más pequeño, el virus satélite de la necrosis del tabaco [STNV, *satellite tobacco necrosis virus*], no puede replicarse de forma independiente, requiere de la presencia simultánea de un virus "cooperador", el virus de la necrosis del tabaco [TNV, *tobacco necrosis virus*], que por sí mismo es un virus normalmente infeccioso.)

1.10 Los ácidos nucleicos se hibridan por apareamiento de bases

Conceptos principales

- El calentamiento provoca que las dos cadenas de un DNA dúplex se separen.
- La T_m es el punto medio del rango de temperatura de desnaturalización.
- Las cadenas complementarias únicas pueden renaturalizarse cuando se reduce la temperatura.
- Puede haber desnaturalización y renaturalización/hibridación con combinaciones de DNA-DNA, DNA-RNA o RNA-RNA, y pueden ser intermoleculares o intramoleculares.
- La capacidad de dos preparaciones de ácido nucleico de una sola cadena para hibridarse demuestra su complementariedad.

Una propiedad clave de la hélice dúplex es su capacidad para separar las dos cadenas sin afectar los enlaces covalentes; esto hace posible dicha separación, y que se reformen en condiciones fisiológicas a la velocidad necesaria (muy rápida) para mantener las funciones genéticas. La especificidad del proceso depende del apareamiento complementario de las bases.

El concepto de apareamiento de las bases es fundamental para todos los procesos relacionados con los ácidos nucleicos. La separación de los pares de bases es un aspecto crucial de la función de una molécula de doble cadena, mientras que la capacidad para formar pares de bases lo es para la actividad de un ácido nucleico de cadena única. En la **FIGURA 1.19** se muestra que el apareamiento de bases permite que los ácidos nucleicos complementarios de cadena única formen una estructura dúplex.

- Una región dúplex intramolecular puede formarse por apareamiento de las bases entre dos secuencias complementarias que forman parte de una molécula de cadena única.
- Una molécula de cadena única puede aparearse a través de las bases con una molécula de cadena única, complementaria e independiente, para formar un dúplex intermolecular.

La formación de dominios dúplex a partir de ácidos nucleicos de cadena única es más importante para el RNA, pero también existe un DNA de cadena única (en forma de genomas víricos). El apareamiento de las bases entre cadenas complementarias únicas e independientes no está restringido a combinaciones DNA-DNA o RNA-RNA, también puede ocurrir entre una molécula de DNA y una de RNA.

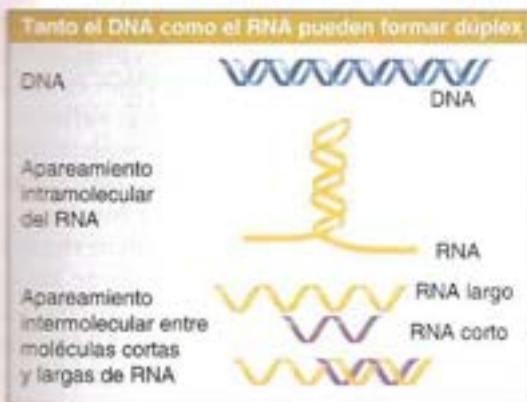


FIGURA 1.19 El apareamiento de bases ocurre en los DNA dúplex e incluso en las interacciones intermoleculares e intramoleculares del RNA (o del DNA) de cadena individual.

La carencia de enlaces covalentes entre cadenas complementarias permite manipular al DNA *in vitro*. Las fuerzas no covalentes que estabilizan a la hélice dúplex se interrumpen por calentamiento o exposición a concentraciones bajas de sales. Las dos cadenas de la hélice dúplex se separan completamente cuando se rompen todos los puentes de hidrógeno que las unen.

El proceso de separación de cadenas se llama **desnaturalización** o (coloquialmente) *fusión*. (El término "desnaturalización" también se utiliza para describir la pérdida de la estructura auténtica de una proteína; es un término general que implica que la conformación natural de una macromolécula se ha transformado.)

La desnaturalización del DNA se presenta en un rango limitado de temperatura y da lugar a cambios sorprendentes en muchas de sus propiedades físicas. El punto medio del rango de temperatura en que las cadenas del DNA se separan se conoce como *temperatura de fusión* (T_m , *melting temperature*), la cual depende de la proporción de pares de bases G-C. Como cada uno de estos pares tiene tres puentes de hidrógeno, es más estable que un par de bases formado por A-T, el cual sólo cuenta con dos de dichos puentes. Entre más pares de G-C contenga el DNA, mayor será la cantidad de energía necesaria para separar las cadenas. En solución y en condiciones fisiológicas, un DNA con 40% de puentes G-C, valor típico de los genomas de los mamíferos, se desnaturaliza a una T_m aproximada de 87°C, de modo que el DNA dúplex es estable a la temperatura de la célula.

La desnaturalización del DNA es reversible en condiciones apropiadas. La capacidad de las dos cadenas complementarias separadas para volver a formar una hélice dúplex se llama **renaturalización**, la cual depende del apareamiento específico de las bases entre las cadenas complementarias. En la **FIGURA 1.20** se observa que la reacción tiene lugar

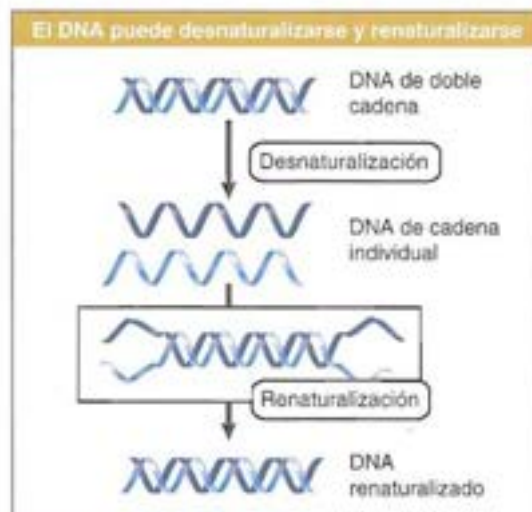


FIGURA 1.20 Las cadenas individuales de DNA desnaturalizadas pueden renaturalizarse para formar una estructura dúplex.

en dos etapas. Primero, las cadenas individuales de DNA que están en la solución se encuentran por casualidad; si sus secuencias son complementarias, aparearán sus bases para generar una región corta de hélice dúplex. Posteriormente, dicha región se extiende a lo largo de la molécula como una cremallera para formar una molécula dúplex larga. La renaturalización de la hélice dúplex restaura las propiedades originales perdidas con la desnaturalización del DNA.

La renaturalización describe la reacción entre dos secuencias complementarias separadas por desnaturalización. Sin embargo, la técnica puede ampliarse para permitir que cualesquiera dos secuencias complementarias de ácidos nucleicos reaccionen entre sí para formar una estructura dúplex. A este proceso se le llama **asociación**, si bien la reacción se describe de forma más general como **hibridación** cuando participan ácidos nucleicos de diferentes orígenes, como sucede cuando una preparación es DNA y la otra, RNA. *La capacidad de dos preparaciones de ácidos nucleicos para hibridarse demuestra con precisión su complementariedad, pues sólo las secuencias complementarias pueden formar una estructura dúplex.*

El principio de la reacción de la hibridación es poner frente a frente dos preparaciones de ácidos nucleicos de cadena individual y después calcular la cantidad de materia de doble cadena que se forma. En la **FIGURA 1.21** se ilustra un procedimiento por el cual se desnaturaliza una preparación de DNA y se adsorben en un filtro las cadenas unitarias. Posteriormente se añade una segunda preparación de DNA (o de RNA) desnaturalizado. El filtro es tratado de tal forma que la segunda preparación se adsorba sólo si puede formar pares de bases con el DNA adsorbido en un principio. Usualmente la segunda

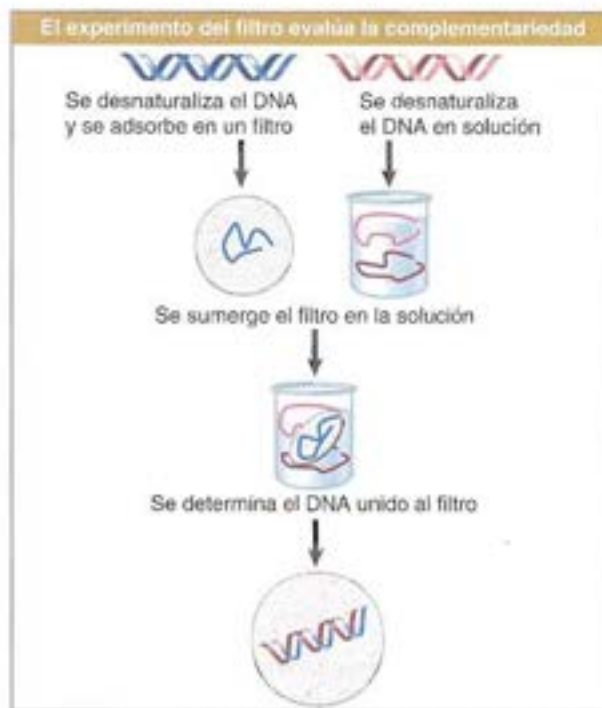


FIGURA 1.21 La hibridación en filtro establece si la solución de DNA (o RNA) desnaturalizado contiene secuencias complementarias a las cadenas inmobilizadas en el filtro.

preparación se marca radioactivamente, de manera que se pueda medir la reacción como la cantidad de marcador radioactivo retenido por el filtro.

La extensión de la hibridación entre dos ácidos nucleicos de cadena individual depende de su complementariedad. No es necesario que dos secuencias sean *perfectamente* complementarias para hibridarse. Si están estrechamente relacionadas, pero no son idénticas, se forma un dúplex imperfecto en el cual el apareamiento de bases se interrumpe en las posiciones en que las cadenas individuales no corresponden.

1.11 Las mutaciones cambian la secuencia del DNA

Conceptos principales

- Todas las mutaciones consisten en cambios en la secuencia del DNA.
- Las mutaciones pueden ser espontáneas o inducidas por mutágenos.

Las mutaciones proporcionan evidencias determinantes de que el DNA es el material genético. Cuando un cambio en la secuencia del DNA provoca una modificación en la secuencia de una proteína, se puede concluir que el DNA codifica a dicha proteína. Ade-

más, un cambio en el fenotipo del organismo puede permitir que se identifique la función de la proteína. La existencia de numerosas mutaciones en un gen permite que se comparen numerosas variantes de una proteína, y mediante un análisis detallado se identificarán las regiones de la proteína responsables de una función enzimática o de otras funciones.

Todos los organismos experimentan cierto número de mutaciones como resultado de las operaciones celulares normales o de las interacciones aleatorias con el ambiente a las cuales se les llama **mutaciones espontáneas**; la velocidad a la que se producen es característica del organismo específico, y en ocasiones se le llama **nivel de fondo**. Las mutaciones son eventos raros y, por supuesto, las que dañan a un gen son descartadas con la evolución, de modo que es difícil obtener grandes cantidades de mutantes espontáneos para estudiarlos en las poblaciones naturales.

La incidencia de las mutaciones suele incrementarse mediante tratamientos con ciertos compuestos; se les llama **mutágenos**, y los cambios que provocan se conocen como **mutaciones inducidas**. La mayoría de los mutágenos actúan directamente en virtud de su capacidad para modificar una base específica del DNA o para incorporarse en el ácido nucleico. La efectividad de un mutágeno se juzga por la medida en que incrementa la tasa de mutaciones respecto del nivel de fondo. Utilizando mutágenos se pueden inducir numerosos cambios en un gen.

Las mutaciones espontáneas que desactivan el funcionamiento de un gen tienen lugar en los bacteriófagos y en las bacterias a una tasa relativamente constante de 3 a 4×10^{-3} por genoma, por generación. Dada la gran variación en el tamaño de los genomas entre bacteriófagos y bacterias, esta tasa corresponde a amplias diferencias en el ritmo de las mutaciones por par de bases, lo cual sugiere que la tasa global de mutación ha estado sujeta a fuerzas selectivas, que han equilibrado los efectos nocivos de la mayoría de las mutaciones respecto de los favorables de algunas mutaciones. Esta conclusión se refuerza por la observación de que una arqueobacteria que vive en condiciones adversas de altas temperaturas y acidez (que supuestamente dañan el DNA) no muestra una tasa elevada de mutaciones, de hecho la tasa de mutación global apenas es menor al rango promedio.

En la **FIGURA 1.22** se muestra que en las bacterias, la tasa de mutaciones corresponde a $\sim 10^{-6}$ eventos por locus por generación, o a una tasa promedio de cambio por par de bases de 10^{-9} a 10^{-12} por generación. La tasa en pares de bases individuales varía considerablemente, en un rango de 10 000 veces. No hay un cálculo preciso de la tasa de mutaciones en

Las tasas de mutación se incrementan con el tamaño de la diana

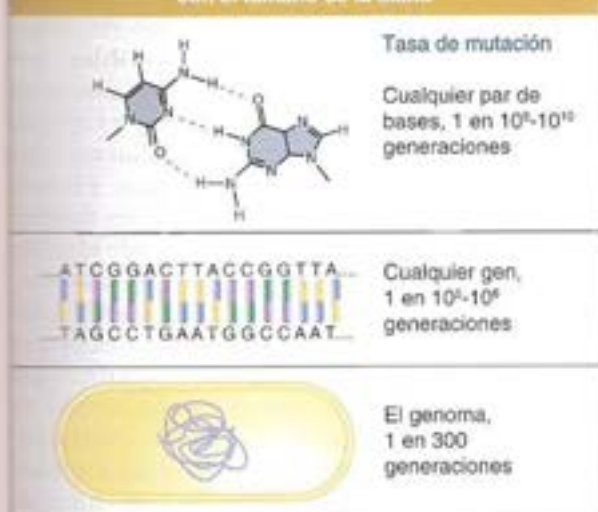


FIGURA 1.22 Un par de bases muta a una tasa de 10^{-9} a 10^{-10} por generación, un gen de 1 000 pares de bases muta a una tasa de $\sim 10^{-6}$ por generación, en tanto que un genoma bacteriano muta a una tasa de 3×10^{-3} por generación.

eucariotas, aunque en general se piensa que hasta cierto punto es similar a la de las bacterias, calculada por locus y por generación.

1.12 Las mutaciones pueden afectar a pares de bases individuales o a secuencias más largas

Conceptos principales

- Una mutación puntual modifica a un solo par de bases.
- Las mutaciones puntuales pueden ser provocadas por la conversión química de una base en otra o por errores durante la duplicación.
- Una transición reemplaza un par de bases G-C con un par de bases A-T, o viceversa.
- Una transversión reemplaza una purina con una pirimidina, por ejemplo, A-T a T-A.
- Las inserciones son el tipo más común de mutación, son resultado del movimiento de elementos transponibles.

Cualquier par de bases del DNA puede mutar. Una **mutación puntual** cambia sólo un par de bases, y puede ser provocada por dos tipos de eventos:

- La modificación química del DNA convierte directamente una base en otra diferente.
- Un mal funcionamiento durante la duplicación del DNA provoca la inserción de una base equivocada en una cadena polinucleotídica durante la síntesis del DNA.

El ácido nitroso desamina a la citosina formando uracilo

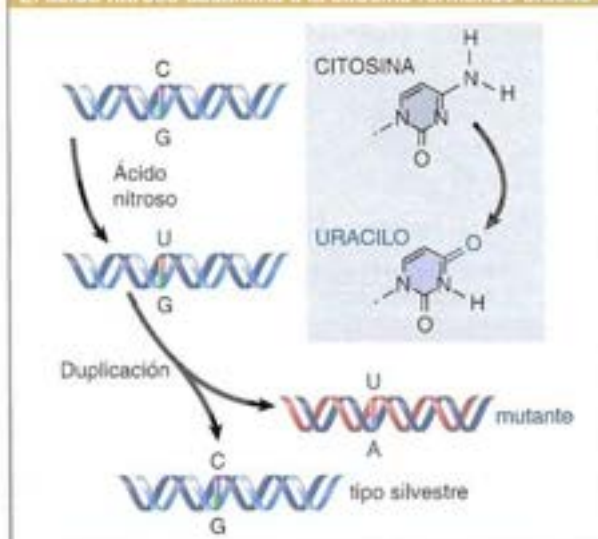


FIGURA 1.23 Las mutaciones pueden ser inducidas por modificación química de una base.

Las mutaciones puntuales pueden ser de dos tipos, dependiendo de la naturaleza del cambio cuando una base es sustituida por otra:

- La clase más común es la **transición**, que consiste en la sustitución de una pirimidina por la otra, o de una purina por la otra, en cuyo caso, un par G-C es reemplazado por un par A-T, o viceversa.
- La clase menos común es la **transversión**, en la cual una purina es reemplazada por una pirimidina o viceversa, de tal forma que un par A-T se convierte en T-A o en C-G.

Los efectos del ácido nitroso constituyen un ejemplo clásico de transición por la conversión química de una base en otra. En la **FIGURA 1.23** se muestra que el ácido nitroso lleva a cabo una desaminación oxidativa que convierte a la citosina en uracilo. En el ciclo de duplicación que sigue a la transición, el U se aparea con una A, y no con la G, con la cual la C original se habría apareado. De esta manera, el par C-G es reemplazado por un par T-A cuando la A se aparea con la T en el siguiente ciclo de duplicación. (El ácido nitroso también desamina a la adenina, provocando la transición inversa de A-T a G-C.)

Las transiciones también son provocadas por **apareamiento erróneo de bases**, cuando parejas inusuales se aparean desafiando la restricción usual de los pares de Watson y Crick. En general, el apareamiento erróneo de bases es una aberración, resultado de la incorporación en el DNA de una base anormal que tiene propiedades ambiguas de apareamiento. En la **FIGURA 1.24** se muestra el ejemplo del bromouracilo (BrdU), molécula análoga a la timina

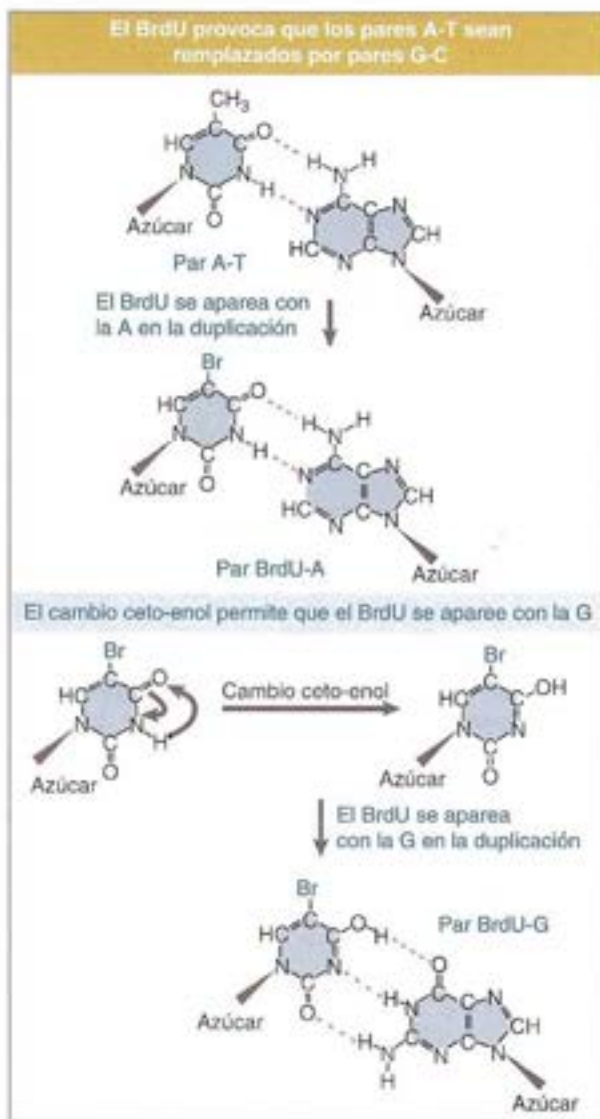


FIGURA 1.24 Se pueden inducir mutaciones al incorporar análogos de bases en el DNA.

que contiene un átomo de bromina en lugar del grupo metilo de la timina. El BrdU se incorpora al DNA en lugar de la timina. No obstante, sus propiedades de apareamiento son ambiguas porque el átomo de bromina permite que se realice un cambio en el cual la base modifica su estructura, de una forma ceto (=O) a una forma enol (-OH). Esta última puede aparearse con la guanina, lo cual conduce a la sustitución del par original A-T por un par G-C.

El apareamiento erróneo puede ocurrir durante la incorporación original de la base o en un ciclo subsiguiente de duplicación. La transición es inducida con cierta probabilidad en cada ciclo de duplicación, de modo que la incorporación de BrdU tiene efectos continuos en la secuencia del DNA.

Durante mucho tiempo se pensó que las mutaciones puntuales eran la vía principal de cambio

en genes individuales, pero ahora sabemos que las **inserciones** de fragmentos de material adicional son muy frecuentes. La fuente del material insertado yace en los **elementos transponibles**, que son secuencias de DNA susceptibles de cambiar de lugar (véase Capítulo 21, Transposones, y Capítulo 22, Retrovirus y retroposones). Normalmente, una inserción suprime la actividad de un gen, y donde han ocurrido dichas inserciones, después pueden producirse **delecciones** de una parte o todo el material insertado, y hasta de los dominios adyacentes.

Una diferencia significativa entre las mutaciones puntuales y las inserciones/delecciones es que la frecuencia de las primeras puede incrementarse por los mutágenos, mientras que la incidencia de los cambios provocados por elementos transponibles no se ve afectada. Sin embargo, las inserciones y delecciones también pueden ser producto de mecanismos, por ejemplo, los que implican errores cometidos durante la duplicación o la recombinación, aunque probablemente sean menos comunes. Por otra parte, una clase de mutágenos, las acridinas, dan lugar a inserciones y delecciones (muy pequeñas).

1.13 Los efectos de las mutaciones pueden ser revertidos

Conceptos principales

- Las mutaciones directas desactivan a un gen, en tanto que inversas (o revertientes) revierten sus efectos.
- Las inserciones pueden revertir por delección del material insertado, pero las delecciones no pueden revertirse.
- La supresión ocurre cuando una mutación de un segundo gen anula el efecto de la mutación del primero.

En la **FIGURA 1.25** se muestra que el aislamiento de los **revertientes** es una característica importante que distingue las mutaciones puntuales y las inserciones de las delecciones:

- Una mutación puntual puede revertirse al restaurar la secuencia original o al adquirir una mutación compensadora en alguna otra parte del gen.
- Una inserción de material adicional puede ser revertida por delección del material insertado.
- Una delección de una parte de un gen no puede revertirse.

Las mutaciones que desactivan a un gen son llamadas **mutaciones directas**, y sus efectos son revertidos por **mutaciones inversas**, las cuales son de dos tipos, reversión verdadera y reversión supresora intragénica.

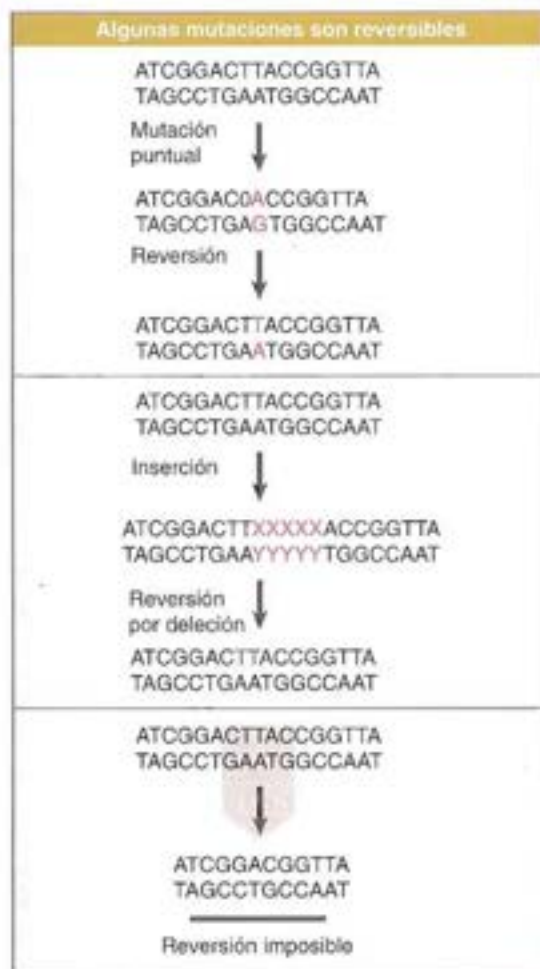


FIGURA 1.25 Las mutaciones puntuales y las inserciones pueden ser revertidas, pero las delecciones son irreversibles.

Una reversión exacta de la mutación original se llama **reversión verdadera**, de tal manera que si un par A-T ha sido remplazado por un par G-C, otra mutación que restituya el par A-T regenerará exactamente la secuencia de tipo silvestre.

El segundo tipo de mutación inversa, la **reversión supresora intragénica**, puede tener lugar en otro sitio del gen, y sus efectos compensan a la primera mutación. Por ejemplo, un cambio de aminoácido en una proteína puede abolir la función del gen, pero una segunda alteración puede compensar por la primera y restaurar la actividad de la proteína.

Una mutación directa resulta de cualquier cambio que desactive un gen, mientras que una mutación inversa debe restaurar la función de una proteína dañada por una mutación directa en particular, de modo que las demandas de una mutación inversa son mucho más específicas que las de una mutación directa. En proporción, la tasa de retromutación es menor que la de mutación directa, normalmente por un factor de ~10 veces.

Las mutaciones también pueden ocurrir en otros genes para contrarrestar los efectos de una mutación en el gen original, efecto que se conoce como **supresión**. Se llama **supresor** al locus en que una mutación suprime el efecto de una mutación en otro.

1.14 Las mutaciones se concentran en puntos calientes

Concepto principal

- La frecuencia de las mutaciones en cualquier par de bases en particular depende de una fluctuación estadística, excepto en los puntos calientes, en los cuales la frecuencia se incrementa cuando menos en un orden de magnitud.

Hasta ahora se han descrito las mutaciones en función de los cambios individuales en la secuencia del DNA que influyen en la actividad de la unidad genética en la cual ocurren. Cuando se analizan las mutaciones desde la perspectiva de la desactivación del gen, la mayoría de los genes de una especie muestran tasas más o menos similares de mutación respecto de su tamaño, lo cual sugiere que el gen puede ser considerado como un objetivo de la mutación, y que un daño en cualquiera de sus partes puede abolir su función, por consiguiente, la susceptibilidad a la mutación es aproximadamente proporcional al tamaño del gen. Pero considerando los sitios de mutación de la secuencia del DNA, ¿todos los pares de bases de un gen son igualmente susceptibles o algunos son más propensos que otros a sufrir mutaciones?

¿Qué sucede cuando se aísla un número importante de mutaciones independientes del mismo gen? Se obtienen numerosos mutantes, cada uno de los cuales es resultado de un evento mutacional individual. Posteriormente se determina el sitio de cada mutación. La mayoría de las mutaciones radicarán en sitios diferentes, si bien algunas estarán en la misma posición. Dos mutaciones aisladas de manera independiente en el mismo sitio pueden constituir exactamente el mismo cambio en el DNA (en cuyo caso el mismo evento mutacional ha sucedido en más de una ocasión), o pueden constituir cambios diferentes (en cada par de bases son posibles tres mutaciones puntuales diferentes).

En el histograma de la FIGURA 1.26 se muestra la frecuencia con la cual se encuentran mutaciones en cada par de bases del gen *lacI* de *E. coli*. La probabilidad estadística de que ocurra más de una mutación en un sitio particular se determina por cinética de impactos aleatorios (como se observa en



FIGURA 1.26 Las mutaciones espontáneas ocurren a lo largo del gen *locI* de la bacteria *E. coli*, pero están concentradas en un punto caliente.

la distribución de Poisson). Por lo tanto, algunos sitios adquirirán una, dos o tres mutaciones, mientras que otros no obtendrán ninguna. Algunos sitios adquieren muchas más mutaciones de las esperadas en una distribución aleatoria; pueden presentar 10 e incluso 100 veces más mutaciones que las predichas por los impactos aleatorios. Estos sitios se llaman **puntos calientes**. En los puntos calientes pueden ocurrir mutaciones espontáneas, y mutágenos diferentes pueden tener diferentes puntos calientes.

1.15 Numerosos puntos calientes son resultado de bases modificadas

Concepto principal

- Una causa común de puntos calientes es la base modificada 5-metilcitosina, la cual sufre desaminación espontánea y se convierte en timina.

Una causa importante de mutación espontánea resulta de la presencia de una base inusual en el DNA. Además de las cuatro bases que se insertan en éste cuando es sintetizado, en ocasiones se pueden observar **bases modificadas** cuyo nombre refleja su origen; son producidas al modificarse químicamente una de las cuatro bases ya presentes en el DNA. La base modificada más común es la 5-metilcitosina, generada por una enzima metilasa que agrega un grupo metilo a ciertos residuos de citosina en sitios específicos del DNA.

Los sitios que contienen 5-metilcitosina proporcionan puntos calientes para que se desarrollen mutaciones puntuales espontáneas en *E. coli*. En cada caso, la mutación adquiere la forma de una transición de G-C a A-T. Las cepas de *E. coli* incapaces de metilar la citosina carecen de puntos calientes.

La razón de la existencia de los puntos calientes es que las bases de citosina sufren desaminación



FIGURA 1.27 La desaminación de la citosina produce uracilo, y la de 5-metilcitosina, timina.

espontánea con frecuencia considerable. En esta reacción, el grupo amino es reemplazado por un grupo ceto. Recuérdese que la desaminación de la citosina genera uracilo (véase Figura 1.23). En la **FIGURA 1.27** se compara esta reacción con la desaminación de 5-metilcitosina, en donde la desaminación genera timina. El efecto en el DNA es la generación de pares de bases G-U y G-T, respectivamente, donde hay un **apareamiento erróneo** entre parejas.

Todos los organismos cuentan con sistemas de reparación que corrigen pares de base apareados erróneamente eliminando y reemplazando una de las bases. La operación de estos sistemas determina si los pares apareados erróneamente, como G-U y G-T, resultan en mutaciones.

En la **FIGURA 1.28** se muestra que las consecuencias de la desaminación son diferentes para la 5-metilcitosina y para la citosina. La desaminación (poco frecuente) de la 5-metilcitosina provoca una mutación, en tanto que la desaminación de la citosina más común no tiene este efecto porque los sistemas de reparación son mucho más efectivos para reconocer G-U que G-T.

La *E. coli* contiene una enzima, uracilo-DNA-glucosidasa, que elimina los residuos de uracilo del DNA (véase la sección 20.5, La inversión de bases es utilizada por las metilasas y las glucosilasas). Esta acción deja sin aparear un residuo de G, y posteriormente "un sistema de reparación" inserta una base C para aparearlo. El resultado final de estas reacciones es la restauración de la secuencia original del DNA. Este sistema protege al DNA de las consecuencias de la desaminación espontánea de la citosina (pero no es lo suficientemente activo como para prevenir los efectos del alto nivel de desaminación provocado por el ácido nitroso; véase la Figura 1.23).

Nótese que la desaminación de la 5-metilcitosina produce timina, lo cual da lugar a un par de bases apareadas erróneamente, G-T. Si el apareamiento

La remoción del uracilo evita las mutaciones

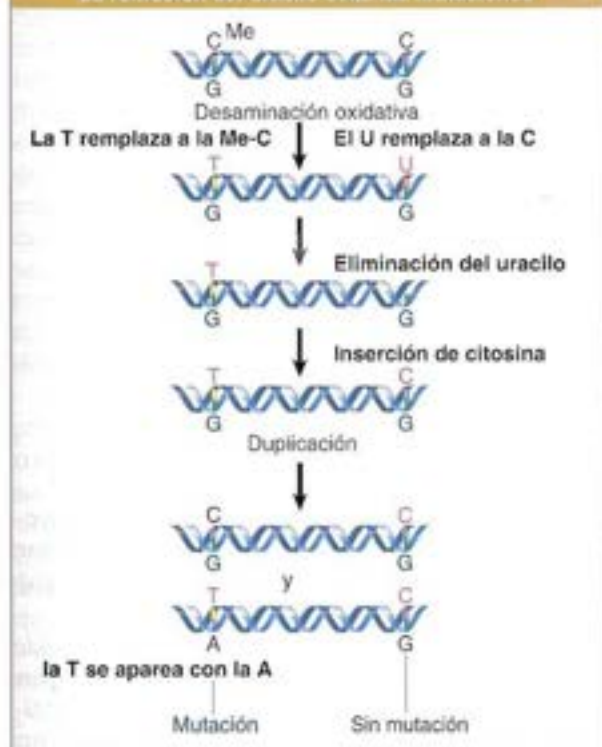


FIGURA 1.28 La desaminación de la 5-metilcitosina produce timina (por transiciones de C-G a T-A), en tanto que la desaminación de la citosina produce uracilo (que suele ser eliminado y posteriormente reemplazado por citosina).

erróneo no se corrige antes del siguiente ciclo de duplicación, resulta una mutación. En la siguiente duplicación, las bases del apareamiento erróneo entre G y T se separan y posteriormente se aparean con nuevas parejas para producir un par G-C de tipo silvestre y un par A-T mutante.

La desaminación de la 5-metilcitosina es la causa más común de la producción de pares erróneos de G-T en el DNA. Los sistemas de reparación que actúan en dichos pares tienden a reemplazar T por C (en vez de la alternativa de reemplazar G por A), lo cual ayuda a reducir la tasa de mutación (véase la sección 20.7, Control de la dirección de la reparación del apareamiento erróneo). No obstante, estos sistemas no son tan efectivos como la remoción de U de los pares erróneos G-U, de modo que la desaminación de la 5-metilcitosina provoca mutaciones con mucha mayor frecuencia que la desaminación de la citosina.

La 5-metilcitosina también crea puntos calientes en el DNA de eucariotas, fenómeno común en dinucleótidos CpG concentrados en regiones denominadas islas CpG (véase la sección 24.19, Los islotes CpG son dianas reguladoras). Si bien la 5-metilcitosina representa ~1% de las bases del DNA humano, los sitios que contienen la base modifica-

da representan ~30% de las mutaciones puntuales; esto hace del estado de la 5-metilcitosina un factor determinante de mutación muy importante en las células animales.

Los efectos de la eliminación de la enzima del ratón MBD4, glucosilasa susceptible de eliminar T (o U) de pares erróneos con G, subrayan la importancia de los sistemas de reparación para reducir la tasa de mutaciones; el resultado es una tasa de mutación tres veces mayor en los sitios CpG. (La razón de que el efecto no sea mayor es que la MBD4 constituye sólo uno de los numerosos sistemas que inciden en los pares erróneos G-T; es de imaginar que la eliminación de todos los sistemas incrementará mucho más la tasa de mutación.)

La operación de estos sistemas proyecta una luz interesante en el uso de la T en el DNA respecto de la U en el RNA que quizá se relacione con la necesidad de estabilidad de la secuencia del DNA; el uso de T significa que cualquier desaminación de C se detecta de inmediato debido a que genera una base (U) que suele estar ausente del DNA, lo cual incrementa en gran medida la eficiencia con que pueden funcionar los sistemas de reparación (comparados con la situación en que tienen que detectar apareamientos erróneos G-T, los cuales también pueden ser producidos por situaciones en que la eliminación de la T no sería la respuesta apropiada). Asimismo, el enlace fosfodiéster del esqueleto es más lábil cuando la base es U.

1.16 Algunos agentes hereditarios son extremadamente pequeños

Concepto principal

- Algunos agentes hereditarios muy pequeños no codifican proteínas, pero constan de RNA o de proteínas que tienen propiedades hereditarias.

Los **viroides** son agentes infecciosos que producen enfermedades en plantas superiores; son moléculas circulares de RNA muy pequeñas. A diferencia de los virus, en los cuales el agente infeccioso es un **virión**, genoma encapsulado en una cubierta de proteína, *el RNA viroide es el agente infeccioso*. El viroide está formado únicamente por RNA, cuyas bases están extensamente apareadas pero de forma imperfecta, de modo que forman una barra característica, como la ilustrada en la **FIGURA 1.29**. Las mutaciones que interfieren con la estructura de esta barra reducen su capacidad infecciosa.

Un RNA viroide consiste en una especie molecular individual que se replica de forma autónoma en células infectadas y cuya secuencia se perpetúa

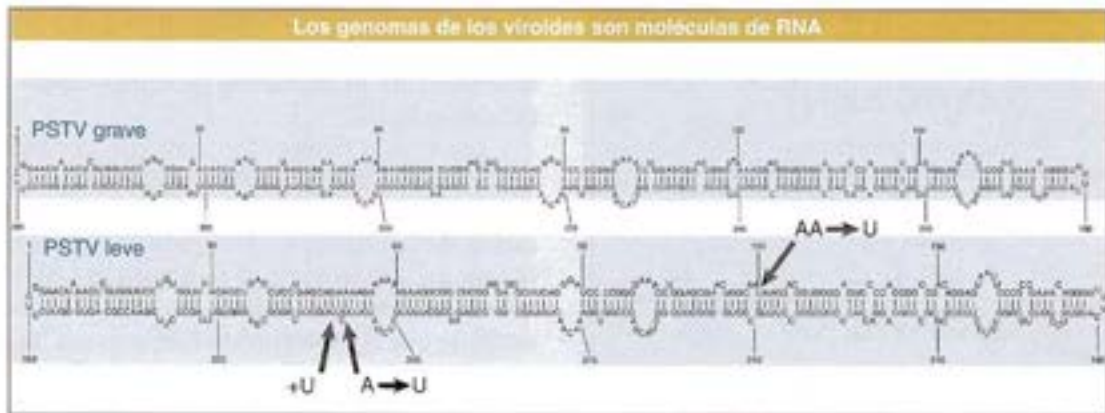


FIGURA 1.29 El RNA del PSTV es una molécula circular que forma una estructura extensa de doble cadena interrumpida por numerosos lazos interiores. La forma leve difiere de la grave en tres sitios.

fielmente en sus descendientes. Los viroides pueden clasificarse en numerosos grupos. Un viroide dado se identifica con un grupo por la similitud de su secuencia con la de otros miembros del grupo. Por ejemplo, cuatro viroides relacionados con el del tubérculo ahusado de la papa (PSTV, *potato spindle tuber viroid*), muestran similitudes de secuencia del 70 al 83 por ciento. Los diferentes aislamientos de una cepa de un viroide en particular pueden variar entre ellos, y el cambio puede afectar al fenotipo de células infectadas. Por ejemplo, las cepas *leves* y *severas* de PSTV difieren por tres sustituciones de nucleótidos.

Los viroides se asemejan a los virus en que sus genomas de ácido nucleico son heredables porque cumplen con todos los requisitos de la información genética, aunque los viroides, llamados en ocasiones **patógenos subvéricos**, difieren de los virus en estructura y función. El RNA viroide no parece ser traducido a proteínas, por lo tanto no puede codificar por sí mismo las funciones necesarias para su supervivencia, fenómeno que genera dos interrogantes: ¿Cómo se replica el RNA viroide? ¿Cómo afecta al fenotipo de la célula de la planta infectada?

La duplicación debe ser realizada por enzimas de la célula hospedadora, subvertidas de su función normal. La heredabilidad de la secuencia del viroide indica que el RNA viroide proporciona el molde.

Presumiblemente, los viroides son patógenos porque interfieren con los procesos celulares normales, quizá de forma relativamente aleatoria, por ejemplo, secuestrando una enzima esencial para su propia duplicación o interfiriendo con la producción de moléculas necesarias de RNA celular, o bien, pueden comportarse como moléculas reguladoras anormales con efectos particulares en la expresión de genes individuales.

Un agente aún menos común es la **encefalopatía espongiiforme de ovejas y cabras** (*scra-*

pie), enfermedad neurológica degenerativa de ovejas y cabras que se relaciona con kuru y síndrome de Creutzfeldt-Jakob, enfermedades humanas que afectan la función cerebral.

El agente infeccioso del *scrapie* no contiene ácido nucleico. Este agente extraordinario llamado **prión** (agente infeccioso proteínáceo) es una glicoproteína hidrófoba de 28 kD, o **PrP**, codificada por un gen celular (conservado entre los mamíferos) que se expresa en el cerebro normal. La proteína existe en dos formas, el producto que se encuentra en el cerebro normal conocido como PrP^c, que es completamente degradado por proteasas. La proteína encontrada en los cerebros infectados se conoce como PrP^{sc} y es extremadamente resistente a la degradación por proteasas. La PrP^c se convierte en PrP^{sc} por una modificación o cambio conformacional que confiere resistencia a las proteasas y que aún no ha sido descrita por completo.

Como agente infeccioso del *scrapie*, el PrP^{sc} debe modificar de alguna manera la síntesis de su contraparte celular normal, de tal forma que, de ser inocuo se torne en infeccioso (véase la sección 31.12. Los priones provocan enfermedades en los mamíferos). Los ratones que carecen del gen PrP no pueden ser infectados para que en ellos se desarrolle *scrapie*, lo cual demuestra que el PrP es esencial para el desarrollo de la enfermedad.

1.17 Resumen

Mediante dos experimentos clásicos se demostró que el DNA es el material genético. El DNA aislado de una cepa de la bacteria *Pneumococcus* puede conferir propiedades de dicha cepa a otra. Además, el DNA es el único componente heredado por la progenie de los fagos progenitores. El DNA puede utilizarse para transferir nuevas propiedades a células eucariotas.

El DNA es una hélice dúplex formada por cadenas antiparalelas en las cuales los nucleótidos están unidos por enlaces fosfodiéster 5'a -3'. El esqueleto forma la parte exterior; las bases de purina y de pirimidina están apiladas en el interior formando pares, en los cuales la A es complementaria de la T, y la G, de la C. Las cadenas se separan y recorren al apareamiento complementario de bases para ensamblar cadenas hijas siguiendo un patrón de duplicación semiconservadora. El apareamiento complementario de bases también se utiliza para transcribir un RNA que representa una sola cadena de un DNA dúplex.

Un fragmento de DNA puede codificar a una proteína. El código genético describe la relación entre la secuencia del DNA y la secuencia de la proteína. Sólo una de las dos cadenas del DNA codifica a una proteína. Un codón está formado por tres nucleótidos que representan un solo aminoácido. Una secuencia codificadora de DNA consiste en una serie de codones, los cuales son leídos desde un punto de inicio predeterminado. En general, uno de los tres marcos de lectura posibles puede ser traducido en proteínas.

Una mutación es un cambio en la secuencia de los pares de bases A-T y G-C del DNA que, en una secuencia codificadora, puede cambiar la secuencia de aminoácidos de la proteína correspondiente. Una mutación por desplazamiento del marco de lectura modifica el marco de lectura subsiguiente insertando o eliminando una base, de modo que se produce una serie completamente nueva de aminoácidos a partir del sitio de la mutación. Una mutación puntual cambia sólo al aminoácido representado por el condón en el cual sucede la mutación. Las mutaciones puntuales pueden ser revertidas por mutación inversa de la mutación original. Las inserciones pueden revertirse por la pérdida del material insertado, pero las deleciones son irreversibles. Las mutaciones también pueden ser suprimidas de manera indirecta cuando una mutación de un gen diferente antagoniza al defecto original.

La incidencia natural de las mutaciones se incrementa por medio de mutágenos. Las mutaciones pueden concentrarse en puntos calientes. Una variedad de punto caliente responsable de algunas mutaciones puntuales es provocada por la desaminación de la base modificada 5-metilcitosina.

Las mutaciones directas ocurren a una tasa de $\sim 10^{-6}$ por locus por generación; las retromutaciones (o mutaciones inversas) son menos comunes. No todas las mutaciones inciden en el fenotipo.

Aunque toda la información genética de las células es transportada por el DNA, los virus tienen genomas de doble cadena o de una sola cadena de DNA o de RNA. Los viroides son patógenos subviri-

cos formados únicamente por moléculas circulares pequeñas de RNA, sin cubierta protectora. El RNA no codifica proteínas; se desconoce su forma de perpetuación y de patogénesis. El *scrapie* es un agente proteínico infeccioso.

Referencias

1.1 Introducción

Revisiones

- Cairns, J., Stent, G., and Watson, J.D. (1996). *Phage and the Origins of Molecular Biology*. Cold Spring Harbor Symp. Quant. Biol.
- Judson, H. (1978). *The Eighth Day of Creation*. Knopf, New York.
- Olby, R. (1974). *The Path to the Double Helix*. MacMillan, London.

1.2 El DNA es el material genético de las bacterias

Artículos de investigación

- Avery, O. T., MacLeod, C. M., and McCarty, M. (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types. *J. Exp. Med.* 98, 451-460.
- Griffith, F. (1928). The significance of pneumococcal types. *J. Hyg.* 27, 113-159.

1.3 El DNA es el material genético de los virus

Artículos de investigación

- Hershey, A. D. and Chase, M. (1952). Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J. Gen. Physiol.* 36, 39-56.

1.4 El DNA es el material genético de las células animales

Artículos de investigación

- Pellicer, A., Wigler, M., Axel, R., and Silverstein, S. (1978). The transfer and stable integration of the HSV thymidine kinase gene into mouse cells. *Cell* 14, 133-141.

1.6 El DNA es una hélice dúplex

Artículos de investigación

- Watson, J. D., and Crick, F. H. C. (1953). A structure for DNA. *Nature* 171, 737-738.
- Watson, J.D., and Crick, F. H. C. (1953). Genetic implications of the structure of DNA. *Nature* 171, 964-967.
- Wilkins, M. F. H., Stokes, A. R., and Wilson, H. R. (1953). Molecular structure of DNA. *Nature* 171, 738-740.

1.7 La duplicación del DNA es semiconservadora

Revisiones

- Holmes, F. (2001). *Meselson, Stahl, and the Replication of DNA: A History of the Most Beautiful Experiment in Biology*. Yale University Press, New Haven, CT.

Artículos de investigación

Meselson, M. and Stahl, F. W. (1958). The replication of DNA in *E. coli*. *Proc. Natl. Acad. Sci. USA* 44, 671–682.

1.11 Las mutaciones cambian la secuencia del DNA

Revisiones

Drake, J. W., Charlesworth, B., Charlesworth, D., and Crow, J. F. (1998). Rates of spontaneous mutation. *Genetics* 148, 1667–1686.

Drake, J. W. and Balz, R. H. (1976). The biochemistry of mutagenesis. *Annu. Rev. Biochem.* 45, 11–37.

Artículos de investigación

Drake, J. W. (1991). A constant rate of spontaneous mutation in DNA-based microbes. *Proc. Natl. Acad. Sci. USA* 88, 7160–7164.

Grogan, D. W., Carver, G. T., and Drake, J. W. (2001). Genetic fidelity under harsh conditions: analysis of spontaneous mutation in the thermoacidophilic archaeon *Sulfolobus acidocaldarius*. *Proc. Natl. Acad. Sci. USA* 98, 7928–7933.

1.12 Las mutaciones pueden afectar a pares de bases individuales o a secuencias más largas

Revisiones

Maki, H. (2002). Origins of spontaneous mutations: specificity and directionality of base-substitution, frameshift, and sequence-substitution mutageneses. *Annu. Rev. Genet.* 36, 279–303.

1.15 Numerosos puntos calientes son resultado de bases modificadas

Artículos de investigación

Coulondre, C. et al. (1978). Molecular basis of base substitution hotspots in *E. coli*. *Nature* 274, 775–780.

Millar, C. B., Guy, J., Sansom, O. J., Selfridge, J., MacDougall, E., Hendrich, B., Keightley, P. D., Bishop, S. M., Clarke, A. R., and Bird, A. (2002). Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. *Science* 297, 403–405.

1.16 Algunos agentes hereditarios son extremadamente pequeños

Revisiones

Diener, T. O. (1986). Viroid processing: a model involving the central conserved region and hairpin. *Proc. Natl. Acad. Sci. USA* 83, 58–62.

Diener, T. O. (1999). Viroids and the nature of viroid diseases. *Arch. Virol. Suppl.* 15, 203–220.

Prusiner, S. B. (1998). Prions. *Proc. Natl. Acad. Sci. USA* 95, 13363–13383.

Artículos de investigación

Bucler, H. et al. (1993). Mice devoid of PrP are resistant to scrapie. *Cell* 73, 1339–1347.

McKinley, M. P., Bolton, D.C., and Prusiner, S. B. (1983). A protease-resistant protein is a structural component of the scrapie prion. *Cell* 35, 57–62.

Los genes codifican proteínas

ESQUEMA DEL CAPÍTULO

- 2.1** Introducción
- 2.2** Un gen codifica a un solo polipéptido
- La hipótesis de un gen : una enzima resume las bases de la genética moderna: que un gen es un fragmento de DNA que codifica a una sola cadena polipeptídica.
 - La mayoría de las mutaciones deterioran la función del gen en el cual se producen.
- 2.3** Las mutaciones que se presentan en el mismo gen no se pueden complementar
- Una mutación en un gen afecta sólo a la proteína codificada por la copia mutante del gen y no a las proteínas codificadas por algún otro alelo.
 - La incapacidad de dos mutaciones para complementarse (producir un fenotipo silvestre cuando se presentan en configuración *trans* en un heterocigoto) significa que forman parte del mismo gen.
- 2.4** Las mutaciones pueden provocar pérdida o ganancia de función
- Las mutaciones recesivas son provocadas por pérdida de función del producto proteínico.
 - Las mutaciones dominantes son resultado de una ganancia de función.
 - La evaluación de la esencialidad de un gen requiere de una mutación nula (que elimine por completo su función).
 - Las mutaciones silenciosas no tienen ningún efecto porque el cambio de base no altera la secuencia ni la cantidad de proteína, o porque el cambio de la secuencia de la proteína no tiene ningún efecto.
 - Las mutaciones rezumantes (o tripomorfas) inciden en la función del producto del gen, pero no se observan en el fenotipo debido a que prevalece la actividad suficiente.
- 2.5** Un locus puede tener numerosos alelos mutantes diferentes
- La existencia de varios alelos permite ocurrencia de heterocigotos con cualquier combinación en pares de alelos.
- 2.6** Un locus puede tener más de un alelo de tipo silvestre
- Un locus puede tener una distribución polimórfica de alelos sin un alelo individual que pueda ser considerado como el único de tipo silvestre.
- 2.7** La recombinación ocurre por el intercambio físico de DNA
- La recombinación es resultado del entrecruzamiento que ocurre en los quiasmas e involucra a dos de las cuatro cromátidas.
- La recombinación se debe a un rompimiento y una reunión que tienen lugar a través de un intermediario de DNA híbrido.
- 2.8** El código genético se lee en tripletes
- El código genético se lee en tripletes de nucleótidos denominados codones.
 - Los tripletes no están superpuestos y se leen a partir de un punto fijo de inicio.
 - Las mutaciones que insertan o eliminan bases individuales provocan un cambio en los grupos de tripletes después del sitio de la mutación.
 - Las combinaciones de mutaciones que insertan o eliminan tres bases (o múltiplos de tres), insertan o eliminan aminoácidos, pero no cambian la lectura de los tripletes más allá del último sitio de mutación.
- 2.9** Cada secuencia tiene tres marcos de lectura posibles
- Usualmente sólo un marco de lectura es traducido y los otros dos son bloqueados por señales frecuentes de terminación.
- 2.10** Los genes procarióticos son colineales con sus proteínas
- Un gen procariótico es un fragmento constante de $3N$ nucleótidos que codifica a N aminoácidos.
 - El gen, el rRNA y la proteína son todos colineales.
- 2.11** Numerosos procesos son necesarios para expresar el producto proteínico de un gen
- Un gen procariótico se expresa por transcripción en rRNA y posteriormente por traducción del rRNA en una proteína.
 - En las eucariotas, un gen puede contener regiones internas no representadas en la proteína.
 - Las regiones internas son eliminadas del transcrito de RNA por el corte y empalme de éste para dar origen a un rRNA colineal respecto del producto proteínico.
 - Cada rRNA está formado por una región líder 5' no traducida, una región codificadora y una región posterior 3' no traducida.
- 2.12** Las proteínas actúan en *trans*, pero los sitios del DNA, en *cis*
- Todos los productos de los genes (RNA o proteínas) actúan en *trans*; pueden actuar en cualquier copia de un gen en la célula.
 - Las mutaciones de actuación en *cis* identifican a las secuencias de DNA que son el objetivo de reconocimiento de los productos de actuación en *trans*. No se expresan como RNA ni como proteínas, y afectan sólo al fragmento contiguo de DNA.
- 2.13** Resumen

2.1 Introducción

El gen es la unidad funcional de la herencia que constituye una secuencia del genoma cuya función es dar origen a un producto discreto (ya sea una proteína o un RNA); su comportamiento básico fue definido por Mendel hace más de un siglo. Resumido en sus dos leyes, el gen fue reconocido como un "factor de partículas" que se transmite sin cambios del progenitor a su progenie. Un gen puede existir en formas alternas que se conocen como alelos.

En los organismos diploides, los cuales tienen dos conjuntos de cromosomas, una copia de cada uno de ellos se hereda de cada progenitor, mismo comportamiento exhibido por los genes. Una de las dos copias de cada gen es el alelo paterno (heredado del padre), el otro es el materno (heredado de la madre). La equivalencia condujo al descubrimiento de que, de hecho, los cromosomas portan a los genes.

Cada cromosoma consiste en una estructura lineal de genes. Cada gen reside en una localización particular del cromosoma. La localización se conoce de manera más formal como locus. Los alelos de un

gen son las diferentes formas que se encuentran en su locus.

La clave para comprender la organización de los genes en los cromosomas fue el descubrimiento del ligamiento genético, o tendencia de los genes de un mismo cromosoma a mantenerse juntos en la progenie, en lugar de ordenarse de manera independiente, como pronostican las leyes de Mendel. Una vez que se introdujo la unidad de recombinación (reordenamiento) como medida de vinculación, fue posible construir mapas genéticos.

La resolución del mapa de recombinación de una eucariota superior está restringida por lo reducido de la progenie que puede obtenerse de cada cruce. La recombinación es tan poco frecuente entre puntos cercanos, que se observa rara vez entre mutaciones diferentes del mismo gen. Por ello, en los mapas de ligamiento clásico de las eucariotas se puede colocar a los genes en orden, pero no es posible determinar las relaciones en un gen. Al cambiar a un sistema microbiano en el cual se puede obtener una progenie muy numerosa de cada cruce genético, los investigadores pudieron demostrar que la recombinación ocurre dentro de los genes y que sigue las mismas reglas que se dedujeron previamente para la recombinación entre genes.

En un gen, las mutaciones pueden estar organizadas de forma lineal, con lo cual se demuestra que el gen mismo posee la misma construcción lineal que el ordenamiento de genes en un cromosoma, de modo que el mapa genético es lineal en los loci y entre éstos: está formado por una secuencia continua dentro de la cual residen los genes. Esta conclusión condujo naturalmente a la perspectiva moderna que se resume en la FIGURA 2.1, en la cual se considera que el material genético de un cromosoma está formado por una molécula ininterrumpida de DNA que representa a numerosos genes.

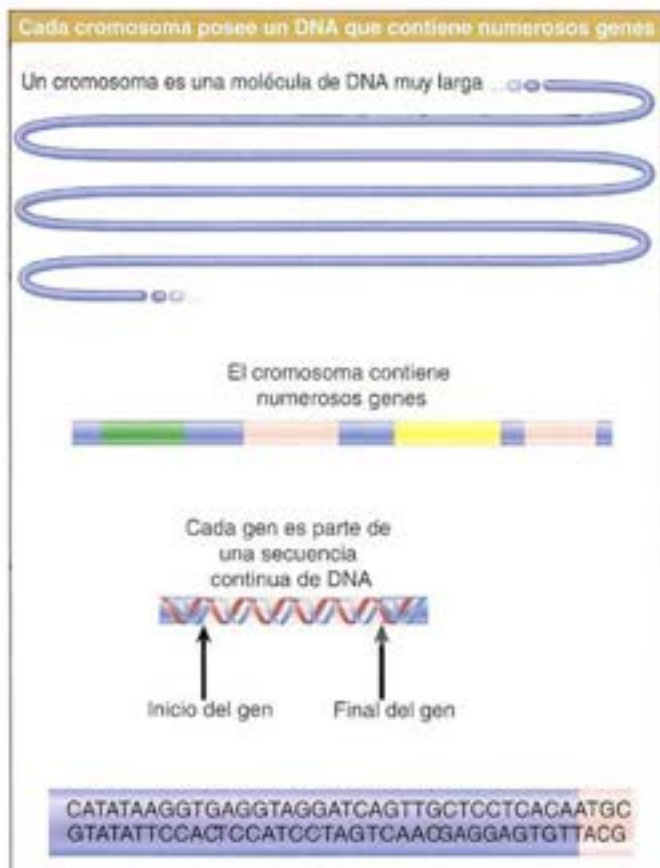


FIGURA 2.1 Cada cromosoma tiene una sola molécula larga de DNA dentro de la cual se encuentran las secuencias de genes individuales.

2.2 Un gen codifica a un solo polipéptido

Conceptos principales

- La hipótesis de un gen: una enzima resume las bases de la genética moderna; que un gen es un fragmento de DNA que codifica a una sola cadena polipeptídica.
- La mayoría de las mutaciones deterioran la función del gen en el cual se producen.

Mediante el primer intento sistemático de relacionar a los genes con las enzimas se demostró que cada etapa de la ruta metabólica es catalizada por una sola enzima y que puede ser bloqueada por mutaciones en un gen diferente. Esto condujo a la hipótesis de un gen:

una enzima. Cada paso metabólico es catalizado por una enzima específica cuya producción es responsabilidad de un solo gen. Una mutación en el gen altera la actividad de la proteína de la cual es responsable.

Para incluir a las proteínas formadas por más de una subunidad, es necesario modificar la hipótesis. Si las subunidades son todas iguales, la proteína es un **homomultímero**, y está representada por un solo gen. Si las subunidades son diferentes, la proteína es un **heteromultímero**. Expresada como una regla más general aplicable a cualquier proteína homomultimérica, la hipótesis de un gen : una enzima se manifiesta con mayor precisión como *un gen : una cadena polipeptídica*.

Identificar qué proteína representa a un gen en particular puede ser tardado ¡La mutación que dio lugar a los chícharos rugosos mutantes de Mendel no se identificó sino hasta 1990, como una alteración que inactiva al gen que codifica a una enzima ramificadora del almidón!

Es importante recordar que un gen no produce directamente una proteína. Como se mostró previamente en la Figura 1.2, un gen codifica a un RNA, que a su vez puede codificar a una proteína. La mayor parte de los genes codifica a las proteínas, pero algunos codifican moléculas de RNA que no producen proteínas. Dichas moléculas de RNA pueden ser componentes estructurales del aparato responsable de la síntesis de proteínas, o bien, regular la expresión genética. El principio básico consiste en que el gen es una secuencia de DNA que especifica la secuencia de un producto independiente. El proceso

de la expresión de los genes puede terminar en un producto que será RNA o proteína.

Una mutación es un evento aleatorio respecto de la estructura del gen en el cual es muy probable que se dañe, o incluso que se suprima, la función del gen. Casi todas las mutaciones que afectan el funcionamiento de un gen son recesivas: *representan ausencia de función, pues al gen mutante se le ha impedido producir su proteína usual*. En la **FIGURA 2.2** se ilustra la relación entre alelos de tipo recesivo y alelos de tipo silvestre. Cuando un heterocigoto contiene un alelo de tipo silvestre y un alelo de tipo mutante, este último es susceptible de regir la producción de la enzima, y por lo tanto, es dominante. (Esto implica que el único alelo de tipo silvestre produce una *cantidad* adecuada de proteína. Cuando esto es falso, la menor cantidad producida por un alelo, comparada con la de dos alelos, resulta en el fenotipo intermedio de un alelo parcialmente dominante en un heterocigoto.)

2.3 Las mutaciones que se presentan en el mismo gen no se pueden complementar

Conceptos principales

- Una mutación en un gen afecta sólo a la proteína codificada por la copia mutante del gen y no a las proteínas codificadas por algún otro alelo.
- La incapacidad de dos mutaciones para complementarse (producir un fenotipo silvestre cuando se presentan en configuración *trans* en un heterocigoto) significa que forman parte del mismo gen.

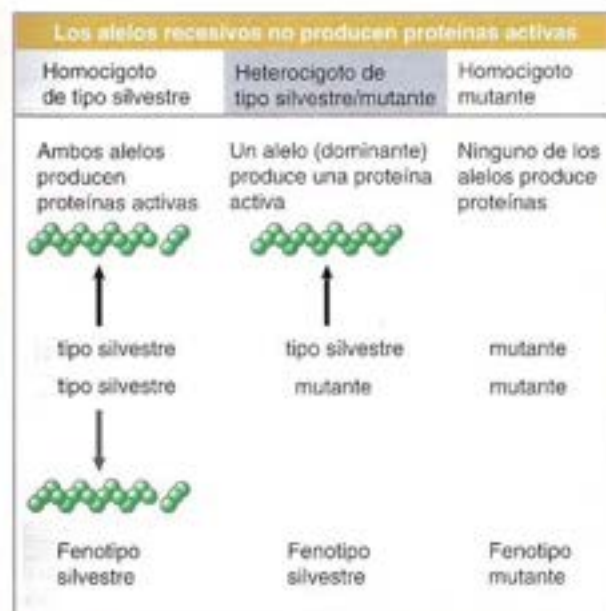


FIGURA 2.2 Los genes codifican proteínas; la dominancia se explica por las propiedades de las proteínas mutantes. Un alelo recesivo no contribuye al fenotipo debido a que no produce ninguna proteína (o produce una proteína que no es funcional).

¿Cómo determinamos si dos mutaciones que provocan un fenotipo similar se encuentran en el mismo gen? Si en el mapa están muy cerca, pueden ser alelos. Pero también podrían representar mutaciones de dos genes *diferentes* cuyas proteínas están implicadas en la misma función. La **prueba de complementación** permite determinar si dos mutaciones yacen en el mismo gen o en genes diferentes. La prueba consiste en hacer un heterocigoto para las dos mutaciones (apareando a progenitores homocigóticos para cada mutación).

Si las mutaciones se encuentran en el mismo gen, los genotipos progenitores pueden representarse como:

$$\frac{m_1}{m_1} \text{ y } \frac{m_2}{m_2}$$

El primer progenitor proporciona un alelo mutante m_1 y el segundo, un alelo m_2 , por lo tanto, el heterocigoto tendrá la constitución:

$$\frac{m_1}{m_2}$$

No se observa ningún gen de tipo silvestre, por lo tanto, el heterocigoto tiene un fenotipo mutante.

Si la mutación yace en genes diferentes, los genotipos progenitores pueden representarse como:

$$\frac{m_1+}{m_1+} \text{ y } \frac{+m_2}{+m_2}$$

Cada cromosoma tiene una copia tipo silvestre de un gen (representada por el signo más) y una copia mutante del otro, de modo que la constitución del heterocigoto será:

$$\frac{m_1+}{+m_2}$$

en la cual los dos progenitores han proporcionado una copia de tipo silvestre de cada gen. El heterocigoto tiene el fenotipo silvestre, de modo que se dice que los dos genes se **complementan**.

La prueba de la complementación se describe más ampliamente en la FIGURA 2.3. La básica consiste en la comparación que se muestra en la parte superior de la figura. Si dos mutaciones se encuentran en el mismo gen, se observa una diferencia en los fenotipos de la configuración *trans* y de la configuración *cis*. La configuración *trans* es mutante, pues cada alelo tiene una mutación (diferente). Sin embargo, la configuración *cis* es de tipo silvestre, debido a que un alelo tiene dos mutaciones y el otro, ninguna. En la parte inferior de la figura se muestra que si las dos mutaciones se encuentran en genes diferentes,

siempre se observará un fenotipo silvestre. Siempre hay un alelo de tipo silvestre y un alelo mutante de cada gen, y la configuración no tiene importancia. La incapacidad para complementar significa que dos mutaciones son parte de la *misma* unidad genética. Se dice que las mutaciones que no se complementan forman parte del mismo **grupo de complementación**. Otro término utilizado para describir a la unidad definida por la prueba de complementación es **cistrón**, que significa lo mismo que gen. Básicamente estos tres términos describen a un fragmento de DNA que funciona como una unidad que dará lugar a un RNA o un producto proteínico. Las propiedades del gen respecto de la complementación se explican por el hecho de que este producto es una molécula única que se comporta como una unidad funcional.

2.4

Las mutaciones pueden provocar pérdida o ganancia de función

Conceptos principales

- Las mutaciones recesivas son provocadas por pérdida de función del producto proteínico.
- Las mutaciones dominantes son resultado de una ganancia de función.
- La evaluación de la esencialidad de un gen requiere de una mutación nula (que elimine por completo su función).
- Las mutaciones silenciosas no tienen ningún efecto porque el cambio de base no altera la secuencia ni la cantidad de proteína, o porque el cambio de la secuencia de la proteína no tiene ningún efecto.
- Las mutaciones rezumantes (o hipomorfas) afectan a la función del producto del gen, pero no se observan en el fenotipo porque queda actividad suficiente.



FIGURA 2.3 El cistrón se define mediante la prueba de complementación. Los genes están representados por espirales; los asteriscos rojos identifican a los sitios de mutación.

Los diversos efectos posibles de las mutaciones en un gen se resumen en la FIGURA 2.4.

Cuando un gen ha sido identificado, en principio se puede llegar a conocer su función produciendo un organismo mutante que carezca completamente del gen. Una mutación que elimina por completo la función de un gen, normalmente porque el gen ha sido eliminado, se llama **mutación nula**, y si el gen es esencial, la mutación nula es letal.

Para determinar el efecto del gen en el fenotipo, es fundamental caracterizar a un mutante nulo. Cuando una mutación no afecta al fenotipo, siempre cabe la posibilidad de que se trate de una **mutación rezumante**, es decir, se produce una cantidad suficiente de producto activo para que cumpla con su función, aunque la actividad se reduce cuantitativamente o es cualitativamente diferente del tipo silvestre. Sin embargo, si una mutación nula no afecta al fenotipo, se puede concluir con toda seguridad que la función del gen no es necesaria.

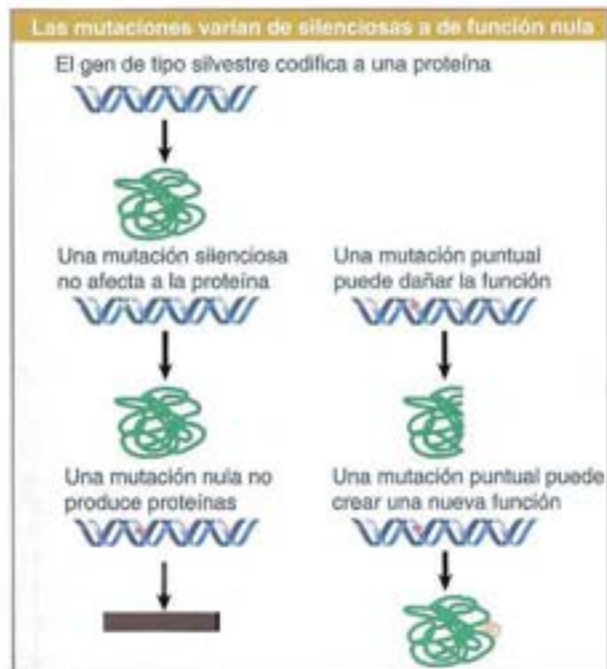


FIGURA 2.4 Las mutaciones que no afectan a la secuencia o a la función de la proteína son silenciosas, en tanto que las mutaciones que eliminan toda la actividad de la proteína son nulas. Las mutaciones puntuales que provocan pérdida de la función son recesivas, a diferencia de las que provocan ganancia de la función, que son dominantes.

Las mutaciones nulas, u otras mutaciones que impiden la función del gen (pero que no necesariamente la eliminan por completo) se denominan **mutaciones de pérdida de función** (o amorfas). Una mutación de pérdida de función es recesiva (como en el ejemplo de la Figura 2.2). En ocasiones, una mutación tiene el efecto opuesto y provoca que una proteína adquiera una nueva función; dicho cambio es conocido como **mutación de ganancia de función**, la cual es dominante.

No todas las mutaciones del DNA provocan cambios detectables en el fenotipo; aquellas sin un efecto aparente se conocen como **mutaciones silenciosas**, las cuales pueden ser de dos tipos; uno de ellos implica cambios de bases en el DNA que no provocan ningún cambio en el aminoácido de la proteína correspondiente, en tanto que el segundo, cambia al aminoácido, pero el remplazo en la proteína no afecta a su actividad; a estas sustituciones se les llama **neutrales**.

2.5 Un locus puede tener numerosos alelos mutantes diferentes

Concepto principal

- La existencia de alelos múltiples permite que los heterocigotos representen cualquier combinación en pares de alelos.

Si se produce una mutación recesiva por cada cambio que impide la producción de una proteína activa en un gen, debe haber gran número de dichas mutaciones en cualquier gen. Numerosos remplazos de aminoácidos pueden cambiar la estructura de la proteína lo suficiente como para impedir su función.

Las diferentes variantes del mismo gen se llaman **alelos múltiples**, y su existencia permite la creación de un heterocigoto entre alelos mutantes. La relación entre estos alelos múltiples asume diversas formas.

En el caso más sencillo, un gen de tipo silvestre codifica a un producto proteínico funcional. El alelo mutante, o los alelos mutantes, codifica(n) proteínas que no son funcionales.

Sin embargo, con frecuencia se presentan casos en que una serie de alelos mutantes tiene fenotipos diferentes. Por ejemplo, la función de tipo silvestre del locus *blanco* de la *Drosophila melanogaster* es necesaria para el desarrollo del color rojo normal de los ojos. El locus recibe su nombre según el efecto de mutaciones (nulas) extremas, las cuales provocan que la mosca tenga ojos blancos en homocigotos mutantes.

Para describir a los alelos mutantes y a los de tipo silvestre, el genotipo silvestre se indica mediante un superíndice "más" (+) después del nombre del locus (w^+ es el alelo silvestre para los ojos [rojos] de la *D. melanogaster*). En algunos casos se utiliza el símbolo + para describir el alelo de tipo silvestre, y sólo los alelos mutantes se indican con el nombre del locus.

Una forma completamente defectuosa del gen (o ausencia de fenotipo) suele indicarse mediante un superíndice "menos" (-). Para poder distinguir entre una variedad de alelos mutantes con efectos diferentes, pueden utilizarse otros superíndices, como w^1 o w^2 .

El alelo w^+ es dominante respecto de cualquiera otro en los heterocigotos. Hay muchos alelos mutantes diferentes; una muestra (pequeña) se ilustra en la **FIGURA 2.5**. Si bien algunos alelos carecen de color de ojos, muchos otros producen algún color, de modo que cada uno de estos alelos mutantes representará una mutación diferente del gen, la cual no elimina por completo su función, sino que deja una actividad residual que produce un fenotipo característico. En un homocigoto, estos alelos son denominados por el color de los ojos. (La mayoría de los alelos w afectan la cantidad de pigmento del ojo. Los ejemplos de la figura están organizados [más o menos] en orden descendente de cantidad de color, pero otros, como el w^e , afectan el patrón en el cual es depositado.)

Cuando existen muchos alelos, un animal puede ser un heterocigoto que porte dos alelos mutantes diferentes, y su fenotipo dependerá de la naturaleza de la actividad residual de cada alelo. En principio, la relación entre dos alelos mutantes no difiere de la que se observa entre los alelos mutantes y los de

Cada alelo tiene un fenotipo diferente	
Alelo	Fenotipo del homocigoto
w^r	ojos rojos (fenotipo silvestre)
w^{bl}	sangre
w^{ch}	cereza
w^{st}	gamuza
w^h	miel
w^a	albaricoque
w^e	eosina
w^l	marfil
w^2	cáscara de limón (amarillo limón)
w^{sp}	moteado, el color varía
w^1	blanco (sin color)

FIGURA 2.5 El locus w tiene una serie extensa de alelos cuyos fenotipos van del color de tipo silvestre (rojo) a la ausencia total de pigmento.

tipo silvestre: un alelo puede ser dominante, puede haber dominancia parcial, o codominancia.

2.6 Un locus puede tener más de un alelo de tipo silvestre

Concepto principal

- Un locus puede tener una distribución polimórfica de alelos sin un alelo individual que pueda ser considerado como el único de tipo silvestre.

No necesariamente hay un solo tipo de alelo silvestre en un locus específico, por ejemplo, el control del sistema del grupo sanguíneo. La falta de función es representada por el tipo nulo, o grupo O . No obstante, los alelos funcionales A y B proporcionan actividades codominantes entre sí y dominantes respecto del grupo O . Las bases de esta relación se ilustran en la **FIGURA 2.6**.

El antígeno O (o H) se genera en todos los individuos, y consiste de un grupo carbohidrato particular que se agrega a las proteínas. El locus ABO codifica a una enzima galactosil-transferasa que agrega otro grupo de azúcar al antígeno O . La especificidad de esta enzima determina el grupo sanguíneo. El alelo A da lugar a una enzima que utiliza el cofactor UDP-N-acetilgalactosa y crea el antígeno A . El alelo B produce una enzima que utiliza el cofactor UDP-galactosa y crea el antígeno B . Las versiones A y B de la proteína transferasa difieren en cuatro aminoácidos que presumiblemente afectan su reconocimiento del tipo de cofactor. El alelo O tiene una mutación (delección pequeña) que elimina la actividad, de modo que en el antígeno O no hay modificaciones.

Esto explica porqué los alelos A y B son dominantes en los heterocigotos AO y BO : la actividad

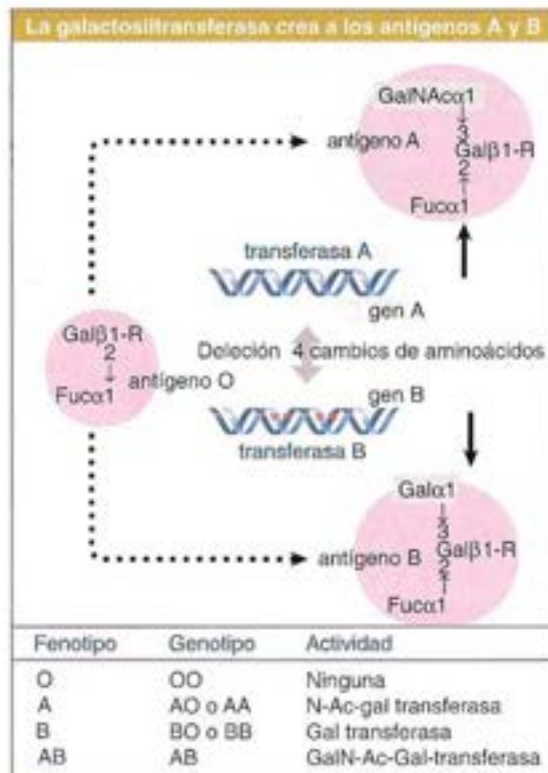


FIGURA 2.6 El locus del grupo sanguíneo ABO codifica a la galactosiltransferasa cuya especificidad determina el grupo sanguíneo.

transferasa correspondiente crea el antígeno A o el B . Los alelos A y B son codominantes en los heterocigotos AB debido a que ambas actividades transferasa están expresadas. El homocigoto OO es nulo y no tiene ninguna actividad, por lo tanto, carece de ambos antígenos.

Ni A ni B pueden ser considerados únicamente como de tipo silvestre porque representan actividades alternativas, más que pérdida o ganancia de función. Una situación como ésta, en la que hay múltiples alelos funcionales en una población, se describe como **polimorfismo** (véase la sección 4.3, Los genomas individuales muestran una variación extensa).

2.7 La recombinación ocurre por intercambio físico de DNA

Conceptos principales

- La recombinación es resultado del entrecruzamiento en los quiasmas e involucra a dos de las cuatro cromátidas.
- La recombinación ocurre por rompimiento y reunión, que tienen lugar a través de un intermediario de DNA híbrido.

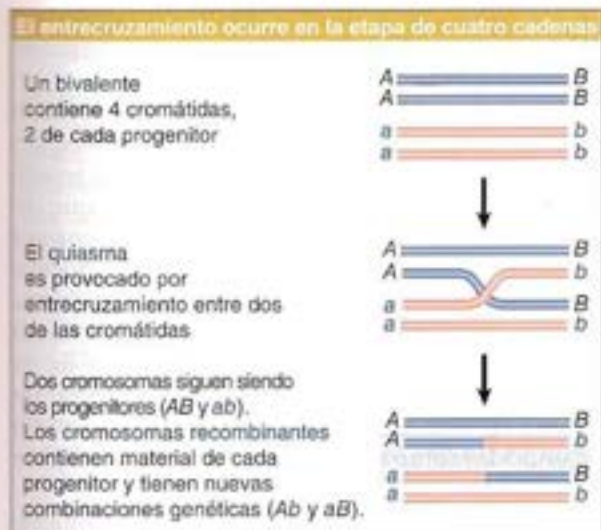


FIGURA 2.7 La formación del quiasma es responsable de la producción de recombinantes.

La recombinación genética describe la generación de combinaciones nuevas de alelos en cada generación de organismos diploides. Las dos copias de cada cromosoma pueden tener diferentes alelos en algunos loci. Intercambiando las partes correspondientes entre los cromosomas, se pueden generar cromosomas recombinantes diferentes de los cromosomas progenitores.

La recombinación resulta de un intercambio físico de material cromosómico, fenómeno visible en el **entrecruzamiento** que ocurre durante la meiosis (división especializada por la que se producen las células germinales haploides). La meiosis empieza con una célula que ha duplicado sus cromosomas, de tal forma que posee cuatro copias de cada uno. En las primeras etapas de la meiosis, esas cuatro copias están íntimamente relacionadas (en sinapsis) en una estructura llamada **bivalente**. En esta etapa, cada unidad cromosómica se denomina **cromátida**. Los intercambios en pares de material ocurren entre cromátidas.

El resultado visible de un evento de entrecruzamiento se denomina **quiasma**, y se ilustra a manera de diagrama en la **FIGURA 2.7**. Un quiasma es un sitio en el que dos de las cromátidas de un bivalente se han roto en los puntos correspondientes. Los extremos rotos se han reunido de forma entrecruzada, generando nuevas cromátidas. Cada cromátida nueva está formada por material proveniente de una cromátida de un lado del punto de unión y por material proveniente de la otra cromátida en el lado opuesto. Las dos cromátidas recombinantes poseen estructuras recíprocas. El evento se define como **rotura y reunión**. Su naturaleza explica porqué un solo evento de recombinación puede producir sólo el 50% de recombinantes: cada evento de recombinación involucra sólo dos de las cuatro cromátidas asociadas.

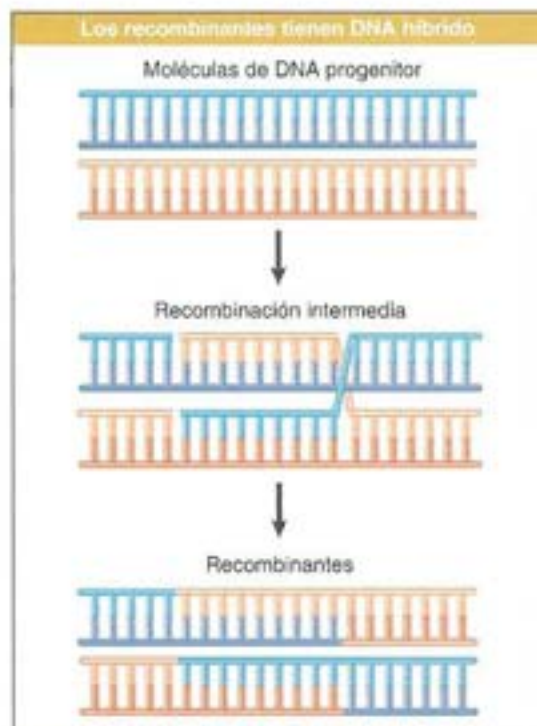


FIGURA 2.8 La recombinación implica el apareamiento entre las cadenas complementarias de dos dúplex de DNA progenitores.

La complementariedad de las dos cadenas de DNA es esencial para el proceso de recombinación. Cada una de las cromátidas que se muestran en la **FIGURA 2.7** consiste en un larguísimo dúplex de DNA. Para que éstas se rompan y reconecten sin pérdida de material, se requiere de un mecanismo que reconozca exactamente las posiciones correspondientes, lo cual sucede gracias al apareamiento complementario de bases.

La recombinación implica un proceso en el cual las cadenas individuales de la región de entrecruzamiento intercambian parejas. En la **FIGURA 2.8** se muestra que este evento da lugar a un fragmento de **DNA híbrido**, en el cual la cadena individual de un dúplex se aparea con su complemento proveniente del otro dúplex. Por supuesto, el mecanismo involucra otras etapas (las cadenas deben romperse y resellarse) que se analizarán en detalle en el Capítulo 20, Sistemas de reparación, pero la característica crucial que hace posible la recombinación exacta es la complementariedad de las cadenas de DNA. La figura muestra sólo algunas etapas de la reacción, pero podemos observar que en la recombinación intermedia se forma un fragmento de DNA híbrido cuando una sola cadena cruza de un dúplex a otro. Cada recombinante está formado por un dúplex de DNA progenitor del lado izquierdo, el cual está conectado por un fragmento de DNA híbrido al otro dúplex progenitor del lado derecho. Cada dúplex de

DNA corresponde a una de las cromátidas implicadas en el proceso de recombinación de la Figura 2.7.

La formación del DNA híbrido exige que las secuencias de los dúplex recombinantes se encuentren lo suficientemente cerca como para lograr el apareamiento entre las cadenas complementarias. Si no hay diferencias entre los dos genomas progenitores en esta región, la formación de DNA híbrido será perfecta. Sin embargo, la reacción puede ser tolerada aun cuando haya diferencias diminutas, en cuyo caso, el DNA híbrido tiene puntos de apareamiento erróneo, en los cuales una base de una cadena confronta a una base de la otra cadena que no es complementaria. La corrección de dichos errores de apareamiento es otra característica de la recombinación genética (véase Capítulo 20, Sistemas de reparación).

2.8 El código genético se lee en tripletes

Conceptos principales

- El código genético se lee en tripletes de nucleótidos denominados codones.
- Los tripletes no están superpuestos y se leen a partir de un punto fijo de inicio.
- Las mutaciones que insertan o eliminan bases individuales provocan un cambio en los grupos de tripletes después del sitio de la mutación.
- Las combinaciones de mutaciones que insertan o eliminan a tres bases en conjunto (o en múltiplos de tres) insertan o eliminan aminoácidos, pero no cambian la lectura de los tripletes más allá del último sitio de mutación.

Cada gen representa una cadena proteínica específica. El concepto de que cada proteína está formada por una serie de aminoácidos en particular, data de la caracterización de Sanger de la insulina, de la década de 1950. El descubrimiento de que un gen está formado por DNA dio lugar a la interrogante de cómo una secuencia de nucleótidos del DNA representa una secuencia de aminoácidos en las proteínas.

Una característica clave de la estructura general del DNA es que *es independiente de la secuencia particular de los nucleótidos que la componen*. La secuencia de nucleótidos del DNA es importante no por su estructura *per se*, sino porque *codifica* a la secuencia de aminoácidos que constituye al polipéptido correspondiente. La relación entre una secuencia de DNA y la secuencia de la proteína correspondiente se denomina **código genético**.

La estructura y la actividad enzimática de las proteínas se deben a su secuencia primaria de aminoácidos, que al ser determinada en cada una, el gen puede transportar toda la información necesaria para especificar una cadena polipeptídica acti-

va. De esta manera, un sólo tipo de estructura —el gen— es susceptible de representarse a sí mismo en innumerables formas de polipéptidos.

En conjunto, los diversos productos proteínicos de una célula asumen las actividades catalíticas y estructurales responsables de establecer su fenotipo. Obviamente, además de las secuencias que codifican a las proteínas, el DNA también contiene ciertas secuencias cuya función es ser reconocidas por moléculas reguladoras, usualmente proteínas. En este caso, la función del DNA es determinada directamente por su secuencia, no a través de un código intermediario. Ambos tipos de región, los genes expresados como proteínas y las secuencias reconocidas como tales, constituyen la información genética.

El código genético es descifrado por un mecanismo complejo que interpreta las secuencias de los ácidos nucleicos, el cual es esencial si la información transportada en el DNA tiene que tener algún significado. En una región dada, sólo una de las dos cadenas de DNA codifica a una proteína, por lo tanto, el código genético se representa como una secuencia de bases (más que como pares de bases).

El código genético se lee en grupos de tres nucleótidos, cada uno de los cuales representa a un aminoácido; cada secuencia de tres nucleótidos se denomina **codón**. Un gen incluye una serie de codones que se lee de forma secuencial desde un punto de partida de un extremo hasta un punto final, en el otro extremo. Escrita en la dirección convencional 5' a 3', la secuencia de nucleótidos de la cadena de DNA que codifica a la proteína corresponde a la secuencia de aminoácidos de la proteína escrita de terminal N hasta terminal C.

El código genético se lee en tripletes no superpuestos a partir de un punto fijo de inicio:

- No superpuestos implica que cada codón está formado por tres nucleótidos y que los codones sucesivos están representados por trinucleótidos sucesivos.
- El uso de un *punto fijo de iniciación* significa que el ensamblaje de una proteína debe comenzar en un extremo y avanzar hacia el otro, de manera que las diferentes partes de la secuencia codificadora no puedan ser leídas de manera independiente.

La naturaleza del código pronostica que dos tipos de mutaciones tendrán efectos diferentes. Si una secuencia en particular se lee de forma secuencial como:

UUU AAA GGG CCC (codones)
aa1 aa2 aa3 aa4 (aminoácidos)

entonces una mutación puntual afectará sólo a un aminoácido. Por ejemplo, la sustitución de una A por cualquier otra base (X) provoca que aa2 sea remplazado por aa5:

UUU AAX GGG CCC
aa1 aa5 aa3 aa4

porque sólo el segundo codón ha sufrido cambios.

Sin embargo, una mutación que inserta o elimina una sola base cambiará a los grupos de tripletes de toda la secuencia subsiguiente. Un cambio de este tipo se llama **desplazamiento del marco de lectura**. Una inserción puede asumir la siguiente forma:

UUU AAX AGG GCC C
aa1 aa5 aa6 aa7

Debido a que la nueva secuencia de tripletes es completamente diferente a la antigua, la secuencia completa de aminoácidos de la proteína se modifica a partir del sitio de la mutación, de modo que es probable que la función de la proteína se pierda por completo. Las mutaciones de cambio del marco de lectura son inducidas por las **acridinas**, compuestos que se unen al DNA y que distorsionan la estructura de la doble hélice, provocando la incorporación de bases adicionales o la omisión de bases durante el proceso de replicación. Cada evento mutagénico provocado por una acridina resulta en la adición o eliminación de un solo par de bases.

Si se produce un mutante de acridina por, digamos, la adición de un nucleótido, éste debe regresar al tipo silvestre por la delección de dicho nucleótido. Sin embargo, la reversión también puede ser provocada por delección de una base diferente en un sitio cercano al primero. Las combinaciones de dichas

mutaciones proporcionan evidencias reveladoras referentes a la naturaleza del código genético.

En la **FIGURA 2.9** se ilustran las propiedades de las mutaciones de cambio del marco de lectura. Una inserción o una delección cambia la secuencia completa de la proteína después del sitio de la mutación, pero la combinación de una inserción y una delección hace que el código se lea de forma incorrecta sólo entre los dos sitios de mutación; la lectura correcta se restablece después del segundo sitio.

En 1961, mediante el análisis genético de las mutaciones por acridina en la región *rII* del fago T6 se demostró que todas las mutaciones podrían ser clasificadas en uno de dos grupos descritos como (+) o (-). Cada tipo de mutación provoca por sí misma un cambio en el marco de lectura: el tipo (+) por adición de una base y el tipo (-) por delección de una base. Las combinaciones de mutantes dobles de los tipos (++) y (--) siguen mostrando un comportamiento mutante, pero las combinaciones de los tipos (+-) o (-+) se suprimen una a la otra. En este caso, la segunda mutación se describe como **supresora del cambio en el marco de lectura** de la primera mutación. (En el contexto de este trabajo, la palabra "supresor" se utiliza en un sentido inusual porque la segunda mutación está en el mismo gen que la primera.)

Estos resultados muestran que el código genético debe ser leído como una secuencia establecida por el punto de iniciación, de modo que las adiciones o las delecciones se compensan mutuamente, mientras que las adiciones o las delecciones dobles siguen siendo mutantes. De cualquier forma, estos hallazgos no revelan cuántos nucleótidos constituyen cada codón.

Cuando se construyen mutantes triples, sólo las combinaciones (+++) y (---) muestran el fenotipo silvestre, mientras que las otras siguen siendo mutantes. Si se considera que tres adiciones o tres delecciones corresponden respectivamente a la adición o la omisión global de un solo aminoácido, esto implica que el código se lee en tripletes. Entre los dos sitios de mutación hay una secuencia incorrecta de aminoácidos, mientras que a cada lado de los sitios de mutación las secuencias siguen siendo de tipo silvestre, como se indica en la **Figura 2.9**.



FIGURA 2.9 Las mutaciones de cambio del marco de lectura muestran que el código genético se lee en tripletes desde un punto fijo de inicio.

2.9 Cada secuencia tiene tres marcos de lectura posibles

Concepto principal

- En general, sólo un marco de lectura es traducido, los otros dos son bloqueados por señales frecuentes de terminación.

Si el código genético se lee en tripletes no superpuestos, son tres las formas posibles de traducir cualquier

secuencia de nucleótidos en proteínas, dependiendo del punto de iniciación; a dichas secuencias se les llama **marcos de lectura**. Para la secuencia

ACGACGACGACGACGACG

los tres marcos de lectura posibles son

ACG ACG ACG ACG ACG ACG ACG
CGA CGA CGA CGA CGA CGA CGA
GAC GAC GAC GAC GAC GAC GAC

Un marco de lectura que consta exclusivamente de tripletes que representan a aminoácidos, se llama **marco de lectura abierto** u **ORF** (*open reading frame*). Una secuencia que se traduce en una proteína tiene un marco de lectura que empieza con un **codón de iniciación** especial (AUG) y que posteriormente se extiende a través de una serie de tripletes que representan a aminoácidos, hasta terminar en uno de los tres tipos de **codones de terminación** (véase el Capítulo 7, El RNA mensajero).

Cuando un marco de lectura no puede ser traducido en proteína porque con frecuencia se presentan codones de terminación, se dice que está **bloqueado**. Si una secuencia está bloqueada en los tres marcos de lectura, no puede tener la función de codificar a una proteína.

Cuando se obtiene la secuencia de una región de DNA de función desconocida, se analiza cada marco de lectura posible para determinar si está abierto o bloqueado. Normalmente, no más de uno de los tres marcos de lectura posibles está abierto en un solo fragmento de DNA. En la **FIGURA 2.10** se muestra el ejemplo de una secuencia que puede leerse en solo un marco de lectura debido a que los marcos de lectura alternativos están bloqueados por codones de terminación frecuentes. Es poco probable que por casualidad exista un marco de lectura largo y abierto; si no se tradujera en una proteína, no habría presión selectiva para evitar la acumulación de codones de terminación, de modo que la identificación de un marco de lectura largo y abierto se considera como evidencia en *primera instancia* de que la secuencia se traduce en proteína en dicho marco. Un ORF para el cual no se ha identificado un producto proteínico suele denominarse marco de lectura no identificado (URF, *unidentified reading frame*).



FIGURA 2.10 Un marco de lectura abierto empieza con AUG y continúa en tripletes hasta un codón de terminación. Los marcos de lectura bloqueados pueden ser interrumpidos con frecuencia por codones de terminación.

2.10 Los genes procarióticos son colineales con sus proteínas

Conceptos principales

- Un gen procariótico es un fragmento constante de $3N$ nucleótidos que codifica a N aminoácidos.
- El gen, el RNA y la proteína son todos colineales.

Comparando la secuencia de nucleótidos de un gen con la secuencia de aminoácidos de una proteína, podemos determinar directamente si el gen y la proteína son **colineales**; esto es, si la secuencia de nucleótidos del gen corresponde exactamente a la secuencia de aminoácidos de la proteína. En las bacterias y sus virus hay una equivalencia exacta, cada gen contiene un fragmento continuo de DNA cuya longitud es directamente proporcional al número de aminoácidos de la proteína que representa. Es necesario un gen de $3N$ pares de bases (bp, *base pairs*) para codificar una proteína de N aminoácidos, de acuerdo con el código genético.

La equivalencia del gen bacteriano y su producto significa que un mapa físico de DNA corresponderá exactamente a un mapa de aminoácidos de la proteína. ¿Qué tan bien se acoplan estos mapas con el mapa de recombinación?

El paralelismo del gen y la proteína se investigó originalmente en el gen de la triptófano sintetasa de *E. coli*. La distancia genética se valoró a través del porcentaje de recombinación entre mutaciones, en tanto que la distancia de la proteína se midió por medio del número de aminoácidos que separaban a los sitios de remplazo. En la **FIGURA 2.11** se comparan ambos mapas. El orden de siete sitios de mutación es el mismo que el orden de los sitios correspondientes a los remplazos de aminoácidos, y las distancias de recombinación son relativamente similares a las distancias reales de la proteína. El mapa de recombinación amplía las distancias entre algunas mutaciones, pero por lo demás, la distorsión del mapa de recombinación es escasa, respecto del mapa físico.

El mapa de recombinación hace ver dos puntos generales acerca de la organización del gen. Mutaciones diferentes pueden provocar que un aminoácido de tipo silvestre sea remplazado por sustitutos diferentes. Si dos de estas mutaciones no pueden recombinarse, deben involucrar mutaciones puntuales diferentes en la misma posición del DNA. Si las mutaciones pueden separarse en el mapa genético, pero afectan al mismo aminoácido en el mapa superior (las líneas comunicantes convergen en la figura), deben implicar a mutaciones puntuales en posiciones diferentes que afectan al mismo aminoácido debido a que la unidad de recombinación ge-

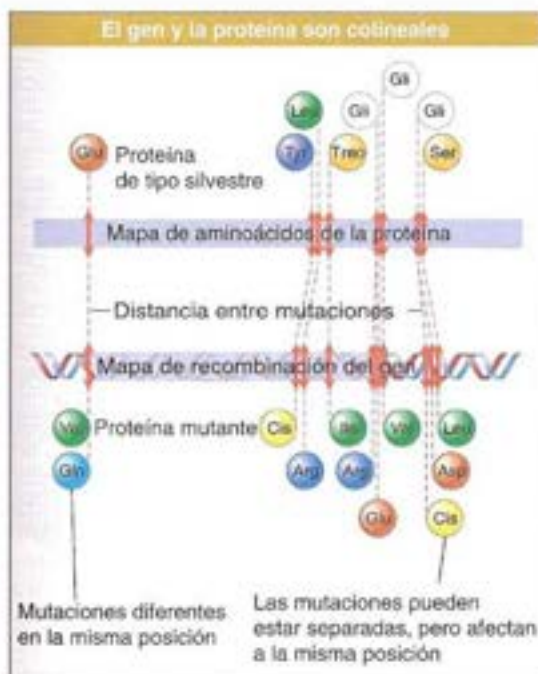


FIGURA 2.11 El mapa de recombinación del gen de la triptófano sintetasa corresponde a la secuencia de aminoácidos de la proteína.

nética (en realidad 1 bp) es más pequeña que la unidad que codifica al aminoácido (en realidad 3 bp).

2.11 Numerosos procesos son necesarios para expresar el producto proteínico de un gen

Conceptos principales

- Un gen procariontico se expresa por transcripción en RNAm y posteriormente por traducción del RNAm en una proteína.
- En las eucariotas, un gen puede contener regiones internas no representadas en la proteína.
- Las regiones internas son eliminadas del transcrito de RNA por el corte y empalme de éste para dar origen a un RNAm colineal respecto del producto proteínico.
- Cada RNAm está formado por una región líder 5' no traducida, una región codificadora y una región posterior 3' no traducida.

Al comparar un gen y una proteína se trata sólo con la secuencia de DNA ubicada entre los puntos correspondientes a los extremos de la proteína. Sin embargo, un gen no se traduce directamente a una proteína, se expresa mediante la producción de un **RNA mensajero (RNAm, messenger RNA)**, que es un ácido nucleico intermedio utilizado, efectiva-

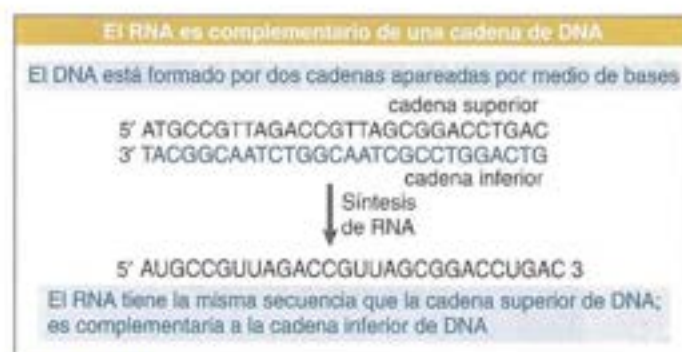


FIGURA 2.12 El RNA se sintetiza utilizando una cadena de DNA como molde para el apareamiento complementario de las bases.

mente, para sintetizar proteína (como se observa en detalle en el Capítulo 7, El RNA mensajero).

El RNA mensajero se sintetiza por el mismo proceso de apareamiento complementario de bases utilizado para replicar el DNA, con la importante diferencia de que corresponde sólo a una cadena del DNA de doble hélice. En la **FIGURA 2.12** se muestra que la secuencia del RNAm es complementaria a la secuencia de una cadena de DNA y que es idéntica (excepto por el remplazo de T por U) a la otra cadena de DNA. La forma convencional de escribir las secuencias del DNA implica que la cadena de arriba vaya en dirección 5' → 3', con la secuencia igual a la del RNA.

El proceso por el cual un gen da lugar a una proteína se llama expresión génica, que en las bacterias consta de dos etapas. La primera etapa es la **transcripción**, durante la cual se produce una copia de RNAm de una cadena del DNA, en tanto que la segunda es la **traducción** del RNAm en una proteína. Mediante este proceso se lee en tripletes la secuencia de un RNAm para originar la serie de aminoácidos que crea a la proteína correspondiente.

Un RNAm incluye a una secuencia de nucleótidos que corresponde a la secuencia de aminoácidos de la proteína; esta parte del ácido nucleico se conoce como **región codificadora**. Sin embargo, el RNAm incluye secuencias adicionales en ambos extremos, las cuales no representan directamente a una proteína. La región 5' no traducida se llama **líder**, en tanto que a la región 3' no traducida se le conoce como **remolque**.

El gen incluye la secuencia completa representada en el RNA mensajero. Ocasionalmente, las mutaciones que impiden la función del gen se encuentran en las regiones adicionales no codificadas, con lo cual se confirma la perspectiva de que dichas regiones comprenden una parte legítima de la unidad genética.

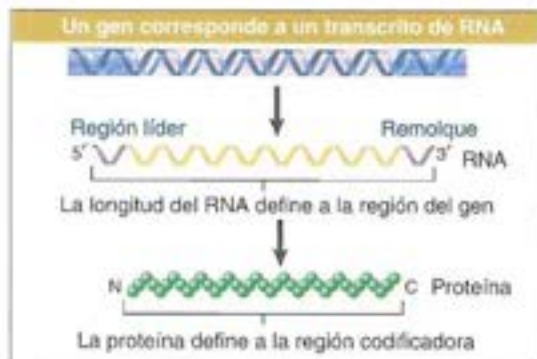


FIGURA 2.13 El gen puede ser más largo que la secuencia que codifica a la proteína.

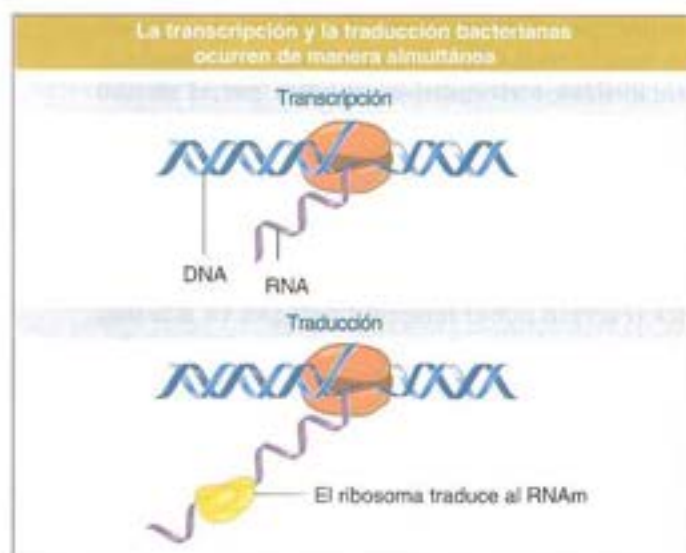


FIGURA 2.14 En las bacterias, la transcripción y la traducción ocurren en el mismo compartimiento.

En la FIGURA 2.13 se ilustra esta situación, en la cual se considera que el gen comprende un fragmento continuo de DNA necesario para producir una proteína en particular; incluye la secuencia codificadora para dicha proteína, pero también secuencias a ambos lados de la región codificadora.

Una bacteria está formada por un solo compartimiento, de modo que la transcripción y la traducción ocurren en el mismo lugar, como se ilustra en la FIGURA 2.14. En las eucariotas, la transcripción tiene lugar en el núcleo, no obstante, el producto de RNA debe ser transportado al citoplasma para ser traducido. En los genes más simples de las eucariotas (igual que en las bacterias), el RNA transcrito es de hecho el RNAm. Sin embargo, para los genes más complejos, el transcrito inmediato del gen es el pre-RNAm, que requiere de cierto procesamiento para generar al RNAm maduro. Las etapas básicas de la expresión génica de una eucariota se bosquejan en la FIGURA 2.15. Este proceso resulta en

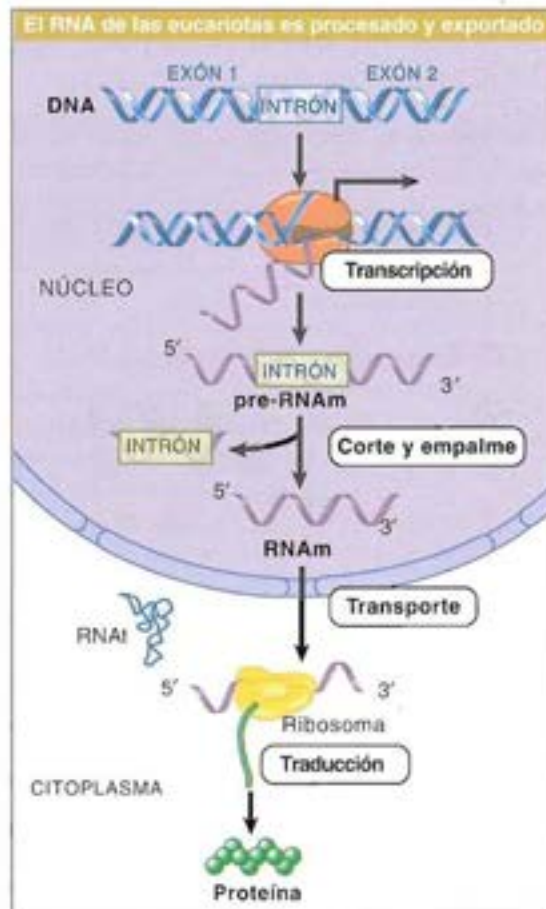


FIGURA 2.15 La expresión de un gen es un proceso que sucede en múltiples etapas.

una separación espacial entre la transcripción (en el núcleo) y la traducción (en el citoplasma).

La etapa más importante del procesamiento es el **corte y empalme del RNA**. Numerosos genes de las eucariotas (y casi todos los de las eucariotas superiores) contienen regiones internas que no codifican proteínas. En el proceso de corte y empalme se eliminan estas regiones del pre-RNAm para generar un RNA con un marco de lectura abierto y continuo (véase la Figura 3.1). Otros eventos procesadores que ocurren en esta etapa implican la modificación de los extremos 5' y 3' del pre-RNAm (véase la Figura 7.16).

La traducción tiene lugar mediante un complejo mecanismo que incluye tanto proteínas como componentes de RNA. La "máquina" que se encarga del proceso es el **ribosoma**, un gran complejo que incluye algunas moléculas extensas de RNA (RNA ribosómico, o rRNA) y numerosas proteínas pequeñas. El proceso por el cual se reconoce qué aminoácido corresponde a un triplete de nucleótidos en particular implica un RNA de transferencia intermedio (o tRNA); hay cuando menos una especie de tRNA para cada aminoácido. Numerosas proteínas accesorias están implicadas. La traducción se describe en el Capítulo 7, El RNA

mensajero, pero considérese por ahora que los ribosomas son las grandes estructuras que se mueven a lo largo del RNAm en la Figura 2.14.

El punto importante de esta etapa es que el proceso de la expresión génica implica al RNA no sólo como sustrato esencial, también como proveedor de componentes para el mecanismo. Los componentes RNAr y RNAt son codificados por genes y generados por el proceso de transcripción (como el RNAm, excepto que no hay una etapa de traducción subsiguiente).

2.12 Las proteínas actúan en *trans*, pero los sitios del DNA, en *cis*

Conceptos principales

- Todos los productos de los genes (RNA o proteínas) actúan en *trans*; pueden actuar en cualquier copia de un gen de la célula.
- Las mutaciones de actuación en *cis* identifican a las secuencias de DNA que son el objetivo de reconocimiento de los productos de actuación en *trans*. No se expresan como RNA ni como proteínas, y afectan sólo al fragmento contiguo de DNA.

Una etapa clave para la definición del gen fue el descubrimiento de que todas sus partes deben estar presentes en un fragmento contiguo de DNA. En la terminología genética, se dice que la configuración de los sitios localizados en el mismo DNA está en *cis*, mientras que de la configuración de los localizados en dos moléculas diferentes de DNA, se dice que está en *trans*, de modo que dos mutaciones pueden presentarse en *cis* (en el mismo DNA) o en *trans* (en moléculas diferentes de DNA). La prueba de complementación recurre a este concepto para determinar si dos mutaciones están en el mismo gen (véase la Figura 2.3 de la Sección 2.3). Las mutaciones que se presentan en el mismo gen no se pueden complementar, y ahora podemos ampliar el concepto de la diferencia entre los efectos *cis* y *trans* de la definición de la región codificadora de un gen a la descripción de la interacción entre los elementos reguladores y un gen.

Supongamos que la capacidad de un gen para ser expresado depende de una proteína que se une al DNA cerca de la región codificadora. En el ejemplo representado en la FIGURA 2.16, el RNAm puede ser sintetizado sólo cuando la proteína está unida al DNA, pero supóngase que en la secuencia del DNA en la cual se une esta proteína ocurre una mutación y la proteína ya no reconoce al DNA: el DNA ya no podría expresarse.

Por lo tanto, un gen puede ser desactivado tanto por una mutación en un sitio de control, como por una mutación en una región codificadora. Las mutaciones no pueden

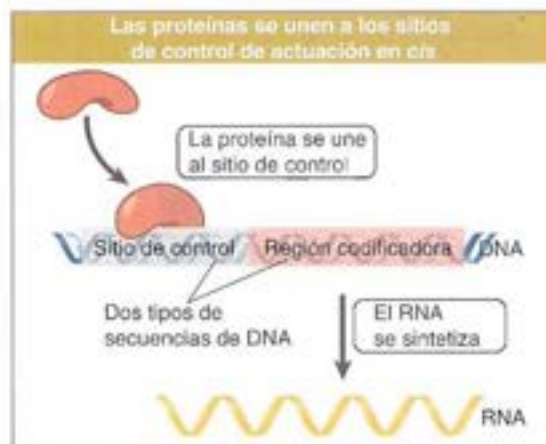


FIGURA 2.16 Los sitios de control del DNA proporcionan sitios de unión para proteínas; las regiones codificadoras se expresan a través de la síntesis de RNA.

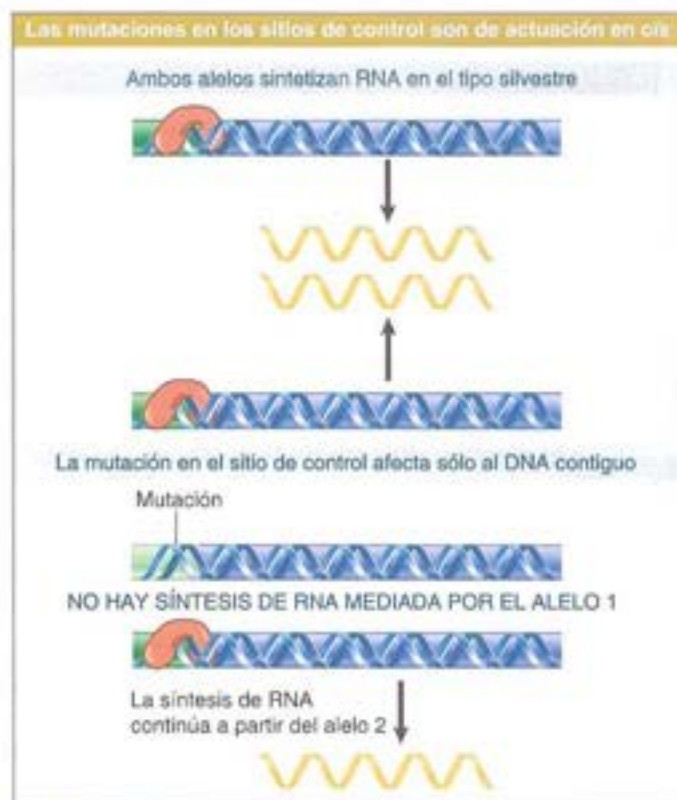


FIGURA 2.17 Un sitio de actuación en *cis* controla al DNA adyacente, pero no influye en el otro alelo.

distinguirse genéticamente porque ambos tienen la propiedad de actuar sólo en la secuencia del DNA del alelo individual en el cual ocurren, sus propiedades son idénticas en la prueba de complementación, de modo que una mutación en una región de control se define como parte integrante del gen, de la misma forma que una mutación en la región codificadora.

La FIGURA 2.17 muestra que una deficiencia en el sitio de control afecta sólo a la región codificadora a la cual está conectada; no afecta la capacidad del otro alelo

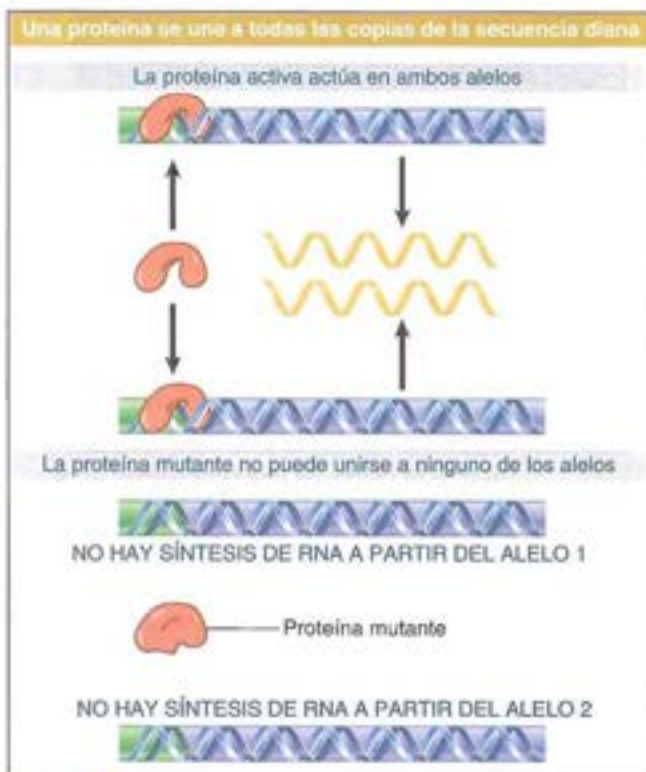


FIGURA 2.18 Una mutación de actuación en *trans* de una proteína afecta a los dos alelos del gen que controla.

para expresarse. Una mutación que actúa exclusivamente afectando las propiedades de la secuencia contigua del DNA se denomina de **actuación en *cis***.

El comportamiento de la mutación de actuación en *cis* que se muestra en la Figura 2.17 se puede comparar con el resultado de una mutación en el gen que codifica a la proteína reguladora. En la **FIGURA 2.18** se muestra que la ausencia de proteína reguladora evitará que *ambos* alelos se expresen; se dice que una mutación de este tipo es de actuación en *trans*.

Revirtiendo el argumento, se sabe que si una mutación es de actuación en *trans*, su efecto deberá ser ejercido a través de algún producto difusible (típicamente una proteína) que actúe en múltiples dianas en una célula. Sin embargo, si una mutación es de actuación en *cis*, su función afectará directamente a las propiedades del DNA contiguo, lo cual significa que *no se expresa en la forma de RNA o de una proteína*.

2.13 Resumen

Un cromosoma es un fragmento ininterrumpido de DNA dúplex que contiene numerosos genes. Cada

gen (o cistrón) se transcribe en un producto de RNA, el cual a su vez es traducido en una secuencia polipeptídica si el gen codifica a una proteína. Se dice que un RNA o un producto proteínico de un gen son de actuación en *trans*. Un gen se define mediante la prueba de complementación como una unidad de un fragmento individual de DNA. Se dice que un sitio del DNA que regula la actividad de un gen adyacente es de actuación en *cis*.

Cuando un gen codifica a una proteína, la relación entre la secuencia de DNA y la secuencia de la proteína depende del código genético. Sólo una de las dos cadenas de DNA codifica a una proteína. Un codón está formado por tres nucleótidos que representan un solo aminoácido. Una secuencia codificadora de DNA consiste en una serie de codones, los cuales son leídos desde un punto fijo de inicio. En general, sólo uno de los tres marcos de lectura posibles puede ser traducido en proteínas.

Un gen puede tener múltiples alelos; los recesivos son provocados por mutaciones de pérdida de función que interfieren con la función de la proteína. Un alelo nulo tiene pérdida total de la función. Los alelos dominantes son provocados por mutaciones de ganancia de función que crean una nueva propiedad en la proteína.

Referencias

2.8 El código genético se lee en tripletes

Revisiones

Roth, J. R. (1974). Frameshift mutations. *Annu. Rev. Genet.* 8, 319-346.

Artículos de investigación

Benzer, S. and Champe, S.P. (1961). Ambivalent rII mutants of phage T4. *Proc. Natl. Acad. Sci. USA* 47, 403-416.

Crick, F.H.C., Barnett, L., Brenner, S., and Watts-Tobin, R. J. (1961). General nature of the genetic code for proteins. *Nature* 192, 1227-1232.

2.10 Los genes procarióticos son colineales con sus proteínas

Artículos de investigación

Yanofsky, C., Drapeau, G.R., Guest, J.R., and Carlton, B.C. (1967). The complete amino acid sequence of the tryptophan synthetase A protein (subunit) and its colinear relationship with the genetic map of the A gene. *Proc. Natl. Acad. Sci. USA* 57, 2966-2968.

Yanofsky, C. et al. (1964). On the colinearity of gene structure and protein structure. *Proc. Natl. Acad. Sci. USA*, 51, 266-272.

El gen interrumpido

ESQUEMA DEL CAPÍTULO

3.1 Introducción

3.2 Un gen interrumpido está formado por intrones y exones

- Los intrones se eliminan mediante el proceso de corte y empalme de RNA, el cual ocurre sólo en configuración *cis* en una molécula individual de RNA.
- Sólo las mutaciones que ocurren en los exones pueden influir en la secuencia de las proteínas; sin embargo, las mutaciones de los intrones pueden afectar el procesamiento del RNA y, por lo tanto, evitar la producción de proteínas.

3.3 Las endonucleasas de restricción son una herramienta esencial en el mapeo del DNA

- Las endonucleasas de restricción pueden utilizarse para dividir al DNA en fragmentos definidos.
- Un mapa se genera utilizando las superposiciones de los fragmentos generados por diferentes enzimas de restricción.

3.4 La organización de los genes interrumpidos puede conservarse

- Los intrones suelen detectarse por la presencia de regiones adicionales al comparar los genes con sus productos de RNA mediante mapeo de restricción o por microscopía electrónica, sin embargo, la definición definitiva se basa en la comparación de secuencias.
- Los intrones normalmente conservan su posición respecto de genes homólogos cuando se comparan organismos diferentes. Las longitudes de los intrones correspondientes pueden variar mucho.
- Normalmente los intrones no codifican proteínas.

3.5 Las secuencias de los exones se mantienen, no así las de los intrones

- La comparación entre genes relacionados de diferentes especies muestra que las secuencias de los exones correspondientes usualmente se conservan, pero la relación entre secuencias de intrones no es tan buena.
- Los intrones evolucionan mucho más rápidamente que los exones debido a la falta de presión selectiva para producir una proteína con una secuencia útil.

3.6 La distribución de tamaños de los genes es amplia

- La mayoría de los genes de las levaduras son continuos, no así los de las eucariotas superiores.
- Los exones suelen ser cortos y codificar menos de 100 aminoácidos.
- Los intrones son cortos en las eucariotas inferiores, pero en las superiores pueden llegar a tener varias decenas de kb (kilobases) de longitud.
- La longitud total de un gen se determina principalmente por sus intrones.

3.7 Algunas secuencias de DNA codifican a más de una proteína

- El uso de codones alternativos de iniciación o de terminación permite que se produzcan dos proteínas cuando una es equivalente a un fragmento de la otra.
- Se pueden producir secuencias no homólogas de proteínas a partir de la misma secuencia de DNA cuando ésta es leída en marcos de lectura diferentes por dos genes (superpuestos).
- Las proteínas homólogas que difieren por la presencia o ausencia de ciertas regiones se deben a un corte y empalme diferencial (alternativo) en el cual se incluyen o excluyen ciertos exones, ya sea por la inclusión o exclusión de exones específicos o al optar por exones alternativos.

3.8 ¿Cómo evolucionaron los genes interrumpidos?

- La interrogante principal al respecto es si los genes se originaron como secuencias interrumpidas por intrones o si originalmente eran moléculas ininterrumpidas.
- Probablemente la forma original de la mayoría de los genes codificadores de proteínas fue la interrumpida, si bien en un principio los genes interrumpidos que codifican al RNA pudieron haber sido moléculas ininterrumpidas.
- Una clase especial de intrones es móvil y puede insertarse en los genes.

3.9 Algunos exones pueden equipararse con funciones proteínicas

- Lo que sugiere que los exones fueron las unidades básicas de la evolución y que los primeros genes fueron interrumpidos en lo siguiente:
 - La estructura génica es igual entre genes de especies muy distantes.
 - Muchos exones pueden ser equiparados con secuencias codificadoras de proteínas que tienen funciones específicas.
 - En genes diferentes se encuentran exones relacionados.

3.10 Los miembros de una familia de genes tienen una organización común

- Se supone que una característica común de una agrupación de genes identifica una propiedad que antecede a la separación durante la evolución.
- La forma común de organización de todos los genes de globina, tres exones y dos intrones, sugiere que descienden de un solo gen ancestral.

3.11 ¿Se encuentra toda la información genética contenida en el DNA?

- La definición del gen se ha invertido, de "un gen : una proteína" a "una proteína : un gen".
- La información sobre la posición también es importante en el desarrollo.

3.12 Resumen

3.1 Introducción

La forma más sencilla de un gen es un fragmento de DNA colineal con su producto proteínico. Los genes bacterianos casi siempre son de este tipo, una secuencia codificadora continua de $3N$ pares de bases representa una proteína de N aminoácidos. Sin embargo, en las eucariotas, un gen puede incluir secuencias adicionales en la región codificadora, las cuales interrumpen a la secuencia que representa a la proteína. Estas secuencias se eliminan del producto de RNA durante la expresión del gen, con lo cual se genera un RNAm que incluye una secuencia de nucleótidos que corresponde exactamente al producto proteínico, de acuerdo con las reglas del código genético.

Las secuencias de DNA que forman un gen interrumpido se dividen en las dos categorías descritas en la FIGURA 3.1:

- Los **exones** son las secuencias representadas en el RNA maduro. Por definición, un gen comienza y termina con exones que corresponden a los extremos 5' y 3' del RNA.
- Los **intrones** son secuencias interpuestas que se eliminan cuando el transcrito primario es procesado para dar lugar al RNA maduro.

Las secuencias de exones siguen el mismo orden en el gen y el RNA, sin embargo, un gen interrumpido es más largo que su producto final de RNA debido a la presencia de los intrones.

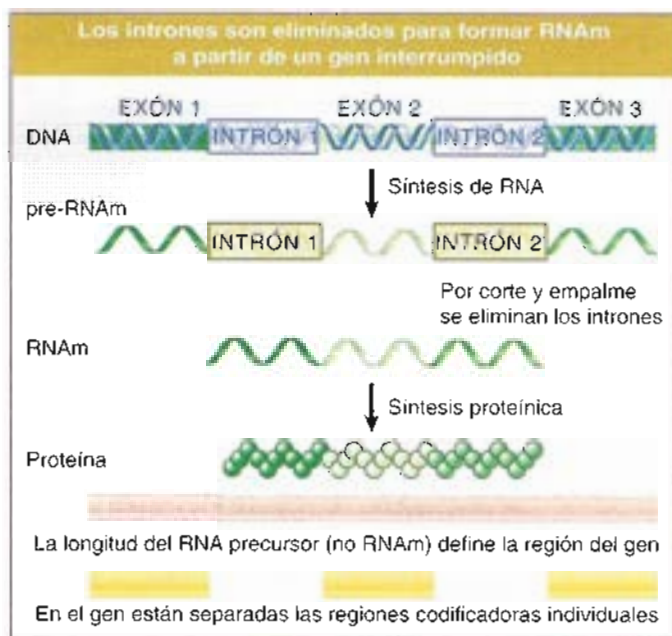


FIGURA 3.1 Los genes interrumpidos se expresan a través de un precursor del RNA. Los intrones se eliminan cuando se cortan y empalman los exones. El RNAm tiene sólo las secuencias de los exones.

La expresión de los genes interrumpidos requiere de un paso adicional que no es necesario en los continuos. El DNA de un gen interrumpido origina una copia de RNA (un **transcrito**) que representa exactamente a la secuencia del genoma. Sin embargo, este RNA sólo es un precursor, no puede utilizarse para producir proteínas; deben eliminarse primero los intrones del RNA para dar lugar a un RNA mensajero formado solamente por una serie de exones. Este proceso se conoce como corte y empalme (véase la sección 2.11. Para expresar el producto proteínico de un gen son necesarios numerosos procesos) e implica precisamente la delección de un intrón del transcrito primario y el empalme de los extremos del RNA en ambos lados para formar una molécula con enlaces covalentes intactos (véase Capítulo 26, Corte, empalme y procesamiento del RNA).

El gen comprende la región del genoma localizada entre los puntos correspondientes a las bases terminales 5' y 3' del RNAm maduro. Se sabe que la transcripción se inicia en el extremo 5' del RNAm y que suele ir más allá del extremo 3', formado por división del RNA (véase la sección 26.19, Los extremos 3' de los RNAm son formados por división y poliadenilación). Se considera que el gen incluye las regiones reguladoras de ambos lados, las cuales son necesarias para iniciar y (en ocasiones) terminar la expresión génica.

3.2 Un gen interrumpido está formado por intrones y exones

Conceptos principales

- Los intrones son eliminados por el proceso de corte y empalme del RNA, el cual ocurre sólo en configuración *cis* en una molécula individual de RNA.
- Sólo las mutaciones ocurridas en los exones pueden afectar la secuencia de las proteínas, pero las que se presentan en los intrones suelen incidir en el procesamiento del RNA y, por lo tanto, evitar la producción de proteínas.

¿Cómo cambia la existencia de los intrones nuestra visión del gen? Después del corte y empalme, los exones siempre se unen en el orden que ocupan en el DNA, de modo de conservar la colinealidad del gen y la proteína entre cada exón y las partes correspondientes de la cadena proteínica. En la FIGURA 3.2 se muestra que, en el gen, el *orden* de las mutaciones sigue siendo el mismo que el de los remplazos de aminoácidos en la proteína. Sin embargo, las *distancias* en el gen no corresponden en absoluto a las distancias en la proteína. Como se observa en un mapa de recombinación, las distancias genéticas no

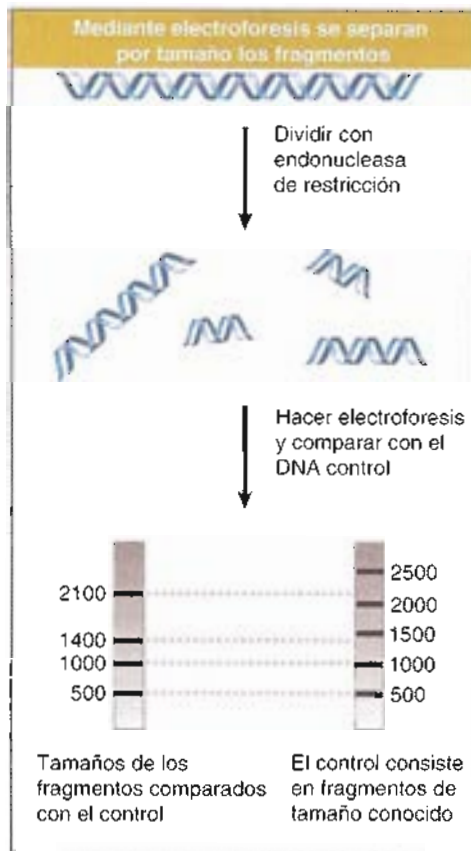


FIGURA 3.3 Los fragmentos generados al dividir al DNA con una endonucleasa de restricción pueden ser separados en función del tamaño.

Un **mapa de restricción** representa una secuencia lineal de los sitios en los cuales enzimas de restricción específicas encuentran a sus secuencias diana. En dichos mapas, las distancias cortas se miden directamente en **pares de bases (bp, base pairs)**, en tanto que las más largas se expresan en **kilobases (kb)**, que corresponde a pares de kilobases (10^3) en el DNA o a kilobases en el RNA. En el ámbito de los cromosomas, un mapa se describe en **pares de megabases (1 Mb = 10^6 bp)**.

Cuando se corta una molécula de DNA con una enzima de restricción adecuada, se divide en distintos fragmentos, que pueden ser separados en función de su tamaño por electroforesis en gel, como se muestra en la **FIGURA 3.3**. El DNA dividido se coloca sobre un gel fabricado con agarosa o poliacrilamida. Cuando se pasa corriente eléctrica a través del gel, cada fragmento se mueve hacia abajo a una velocidad inversamente proporcional al logaritmo de su peso molecular, movimiento que produce una serie de bandas, cada una de las cuales corresponde a un fragmento de un tamaño específico que disminuye el gel.

Analizando los fragmentos de restricción del DNA es posible crear un mapa de la molécula ori-

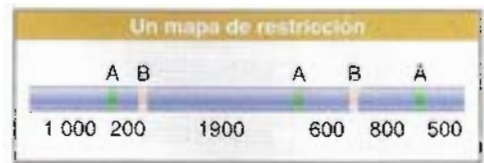


FIGURA 3.4 Un mapa de restricción es una secuencia lineal de sitios separados por distancias definidas en el DNA. En el mapa se identifican los sitios divididos por las enzimas A y B, según se definen los fragmentos individuales producidos por las digestiones dobles e individuales.

ginal tal como se muestra en la **FIGURA 3.4**. En el mapa aparecen las posiciones en que las enzimas de restricción específicas cortan el DNA; las distancias entre los sitios de corte se miden en pares de bases, de modo que el DNA se divide en una serie de regiones de longitudes definidas que yacen entre sitios reconocidos por las enzimas de restricción. Una característica importante es que se puede obtener un mapa de restricción de cualquier secuencia de DNA, a pesar de que se hayan identificado mutaciones en ella, o de hecho, que se conozca, o no, algo de su función.

3.4 La organización de los genes interrumpidos puede conservarse

Conceptos principales

- Los intrones pueden ser detectados por la presencia de regiones adicionales cuando los genes son comparados con sus productos de RNA por mapeo de restricción o por microscopía electrónica, sin embargo, la última definición se basa en la comparación de secuencias.
- Los intrones normalmente conservan su posición respecto de genes homólogos cuando se comparan organismos diferentes. Las longitudes de los intrones correspondientes pueden variar mucho.
- Los intrones usualmente no codifican proteínas.

Cuando un gen es continuo, el mapa de restricción de su DNA corresponde exactamente al mapa de su RNAm.

Cuando un gen posee un intrón, el mapa de cada extremo corresponde al mapa de cada extremo de la secuencia mensajera. Sin embargo, en el gen, los mapas difieren porque las regiones adicionales no están representadas en el mensaje. Cada una de dichas regiones corresponde a un intrón. En el ejemplo de la **FIGURA 3.5** se comparan los mapas de restricción del gen de la β -globina y del RNAm. Hay dos intrones, cada uno de los cuales contiene una serie de sitios de restricción que están ausentes en

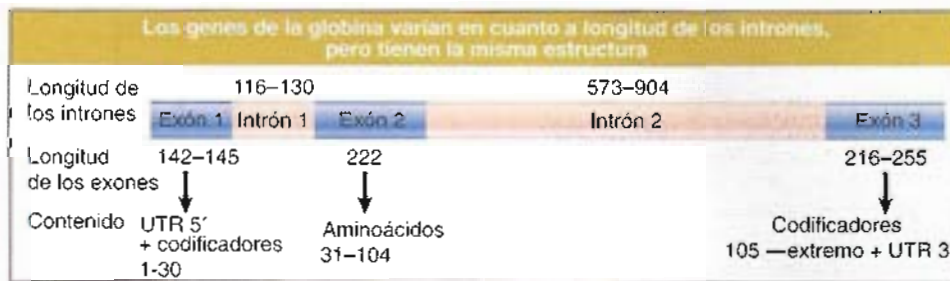


FIGURA 3.7 Todos los genes funcionales de globina tienen una estructura interrumpida con tres exones. Las longitudes indicadas en la figura aplican a los genes de la globina- β de los mamíferos.

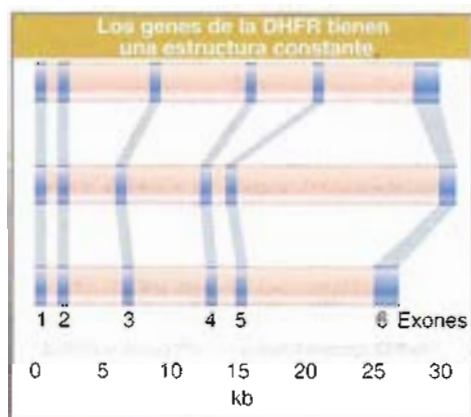


FIGURA 3.8 Los genes de la DHFR de los mamíferos tienen la misma organización relativa de exones más bien cortos e intrones muy largos, pero varía ampliamente en las longitudes de los intrones.

150 bp, por lo tanto, la longitud total del gen es de 850 bp, respecto del gen mayor de la β -globina, en el cual, los 585 bp del intrón resultan en un gen cuya longitud total es de 1382 bp. La variación en la longitud de los genes es mucho mayor que el rango de longitudes de los RNAm (el de la α -globina tiene 585 bases y el de la β -globina, 620 bases).

El ejemplo de la DHFR, un gen un tanto más grande, se muestra en la **FIGURA 3.8**. Dicho gen (dihidrofolato reductasa) de los mamíferos está organizado en seis exones que corresponden al RNAm de 2 000 bases con una longitud mucho mayor de DNA debido a que los intrones son muy largos. En tres mamíferos, los exones son esencialmente iguales y las posiciones relativas de los intrones no cambian, a diferencia de la longitud de cada intrón; no obstante, el resultado de la variación en la longitud del gen es de 25 a 31 kb.

Los genes de la globina y de la DHFR son ejemplos de un fenómeno general: *los genes relacionados por evolución tienen organizaciones relacionadas con el mantenimiento de la posición de los intrones (al menos de algunos de ellos). Las variaciones en la longitud de los genes son determinadas principalmente por la longitud de los intrones.*

3.5 Las secuencias de los exones se mantienen, no así las de los intrones

Conceptos principales

- La comparación de genes relacionados de diferentes especies muestra que las secuencias de los exones correspondientes usualmente se conservan, pero las secuencias de los intrones están mucho menos relacionadas.
- Los intrones evolucionan mucho más rápidamente que los exones por la falta de presión selectiva para producir una proteína con una secuencia útil.

¿Un gen estructural es único en su genoma? La respuesta puede ser ambigua. La longitud completa del gen es única, pero con frecuencia sus exones están relacionados con los exones de otros genes. Como regla general, cuando dos genes están relacionados, la relación entre sus exones es más cercana que la relación entre sus intrones. En un caso extremo, los exones de dos genes pueden codificar a la misma secuencia proteínica, mientras que los intrones pueden ser diferentes, lo cual implica que ambos genes fueron originados por una duplicación de algún gen ancestral común. Posteriormente, se acumularon diferencias entre una y otra copia, pero en los exones se vieron restringidas por la necesidad de codificar funciones proteínicas.

Como se verá más adelante, al analizar la evolución del gen, los exones pueden ser considerados como las unidades básicas ensambladas en varias combinaciones. Un gen puede tener algunos exones relacionados con los exones de otro gen, pero otros pueden no tener relación. En general, los intrones no tienen ninguna relación en dichos casos, pueden surgir por duplicación y translocación de exones individuales.

La relación entre dos genes puede representarse como una comparación de matriz de puntos, como en la **FIGURA 3.9**. Un punto indica la posición de la misma

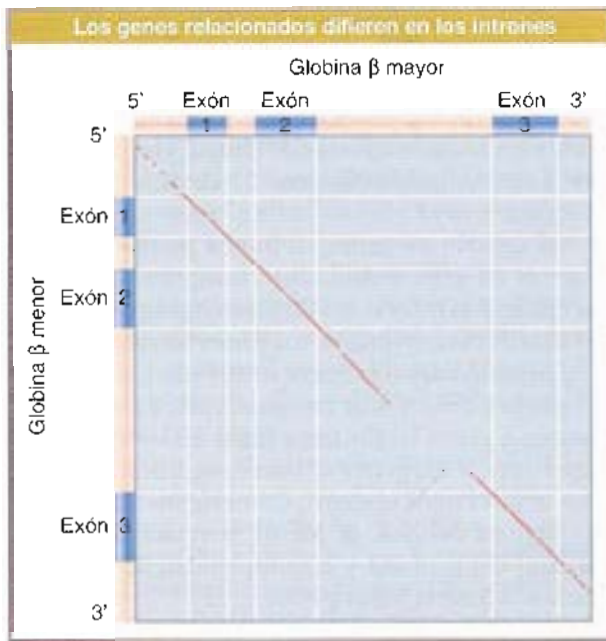


FIGURA 3.9 Las secuencias de los genes de la globina β^{mayor} y de la globina β^{menor} del ratón están íntimamente relacionadas en las regiones codificadoras, pero difieren en las regiones flanqueantes y en el intrón largo. Datos proporcionados por Philip Leder, Harvard Medical School.

secuencia en cada gen. Si dos secuencias son idénticas, los puntos forman una línea en ángulo de 45°. La línea es interrumpida por regiones no similares y se desplaza de forma lateral o vertical por deleciones o inserciones en una secuencia respecto de la otra.

Cuando se comparan los dos genes de la globina- β del ratón, dicha línea abarca los tres exones y el intrón pequeño para luego desvanecerse en las regiones flanqueantes y el intrón largo; es un patrón típico en el cual las secuencias codificadoras están bien relacionadas y la relación puede ir más allá de las fronteras de los exones; sin embargo, el patrón se pierde en los intrones más largos y en las regiones de ambos lados del gen.

El grado total de divergencia entre dos exones depende de las diferencias entre las proteínas, provocadas principalmente por sustituciones de bases. En las regiones traducidas, los exones se ven restringidos por la necesidad de codificar secuencias de aminoácidos, de modo que su potencial de cambio de secuencia es limitado. Muchos de los cambios no afectan el significado de los codones porque se cambia un codón por otro que representa al mismo aminoácido. Los cambios ocurren con mayor libertad en las regiones no traducidas (las cuales corresponden al líder 5' y al remolque 3' del RNA_m).

En los intrones correspondientes, el patrón de divergencia implica cambios tanto de tamaño (por deleciones e inserciones) como sustituciones de bases.

Los intrones evolucionan mucho más rápidamente que los exones. Cuando se compara un gen en diferentes especies, en ocasiones, los exones son homólogos, pero los intrones han cambiado tanto, que no se reconocen las secuencias correspondientes.

Las mutaciones ocurren a la misma velocidad en exones e intrones, pero son eliminadas más eficientemente de los exones por selección adversa. Sin embargo, en ausencia de restricciones impuestas por una función codificadora, un intrón es susceptible de acumular libremente sustituciones puntuales y otros cambios, los cuales implican que el intrón no tenga una función regida por una secuencia específica; aún no se sabe si su presencia es necesaria para la función del gen.

3.6 La distribución de tamaños de los genes es amplia

Conceptos principales

- La mayoría de los genes de las levaduras son continuos, no así los de las eucariotas superiores.
- Los exones suelen ser cortos y codificar menos de 100 aminoácidos.
- Los intrones son cortos en las eucariotas inferiores, pero en las superiores pueden llegar a tener varias decenas de kb (kilobases) de longitud.
- La longitud total de un gen se determina principalmente por sus intrones.

En la **FIGURA 3.10** se muestra la organización global de los genes en las levaduras, los insectos y los mamíferos. En *Saccharomyces cerevisiae*, la gran mayoría de los genes (>96%) son continuos, y los que tienen exones suelen mantenerse razonablemente compactos; virtualmente ninguno de sus genes tiene más de cuatro exones.

En insectos y mamíferos sucede lo contrario, sólo algunos genes tienen secuencias codificadoras ininterrumpidas (6% en los mamíferos). Los genes de los insectos tienden a presentar un número bastante reducido de exones, en general menos de 10. Los genes de los mamíferos están divididos en más fracciones y algunas tienen varias decenas de exones. Aproximadamente el 50% de los genes de los mamíferos tienen >10 intrones.

Al examinar las consecuencias de este tipo de organización para el tamaño total del gen, en la **FIGURA 3.11** se observa que entre las levaduras y las eucariotas superiores hay una diferencia notable. La longitud del gen promedio de una levadura es de 1.4 kb, y en muy pocos es superior a los 5 kb. Sin embargo, la predominancia de genes interrumpidos

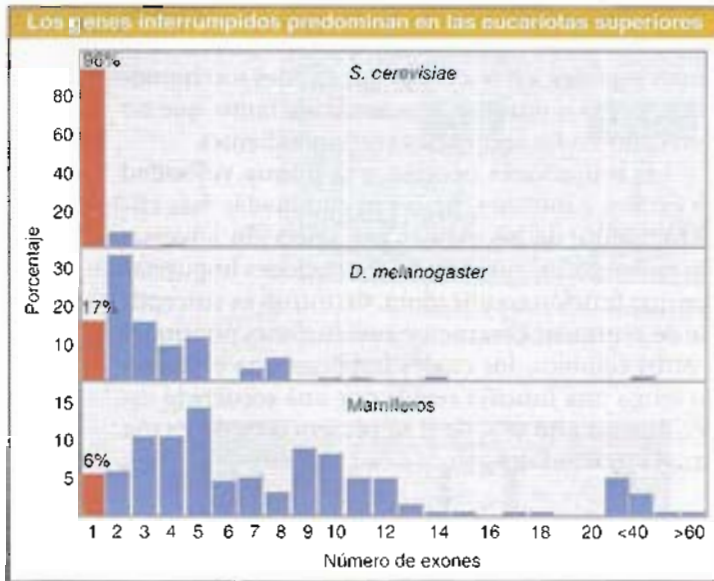


FIGURA 3.10 La mayoría de los genes de las levaduras son continuos, pero en las moscas y los mamíferos, casi todos son interrumpidos. (Los genes continuos tienen un solo exón y se suman en la columna de la extrema izquierda.)

en las eucariotas superiores significa que el gen puede ser mucho más largo que la unidad que codifica una proteína. Relativamente pocos genes de mosca o de mamífero presentan una longitud inferior a 2 kb, y en muchos casos es de entre 5 kb y 100 kb. El gen humano promedio tiene 27 kb de largo (véase la Figura 5.11).

El cambio de genes en buena parte continuos a genes en gran medida interrumpidos ocurre en las eucariotas inferiores. En los hongos (excepto las levaduras), la mayoría de los genes son interrumpidos, pero tienen un número relativamente pequeño de exones (<6) y son bastante cortos (<5 kb). El cambio a genes largos tiene lugar en las eucariotas superiores, y los genes se hacen significativamente más largos en los insectos. Con este incremento en la longitud del gen, se pierde la relación entre el tamaño del genoma y la complejidad de los organismos (véase la Figura 4.5).

Conforme se incrementa el tamaño del genoma, la tendencia es a intrones más largos, mientras que los exones siguen siendo bastante pequeños.

En la **FIGURA 3.12** se observa que los exones que codifican a los fragmentos de una proteína tienden a ser bastante pequeños, de modo que en las eucariotas superiores, el exón promedio codifica a ~50 aminoácidos y la distribución general encaja bien

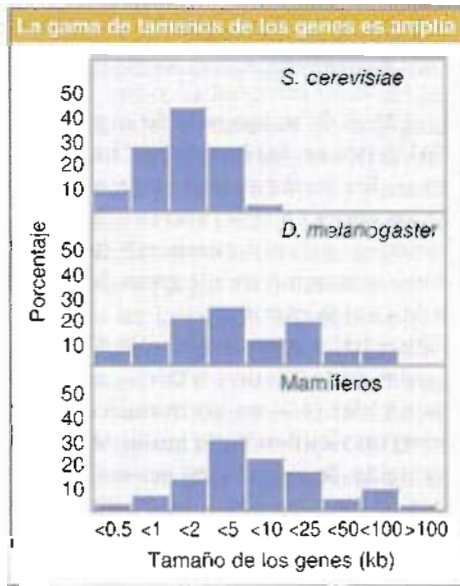


FIGURA 3.11 Los genes de las levaduras son cortos, pero los de las moscas y los mamíferos tiene una distribución dispersa que se extienden a longitudes muy amplias.

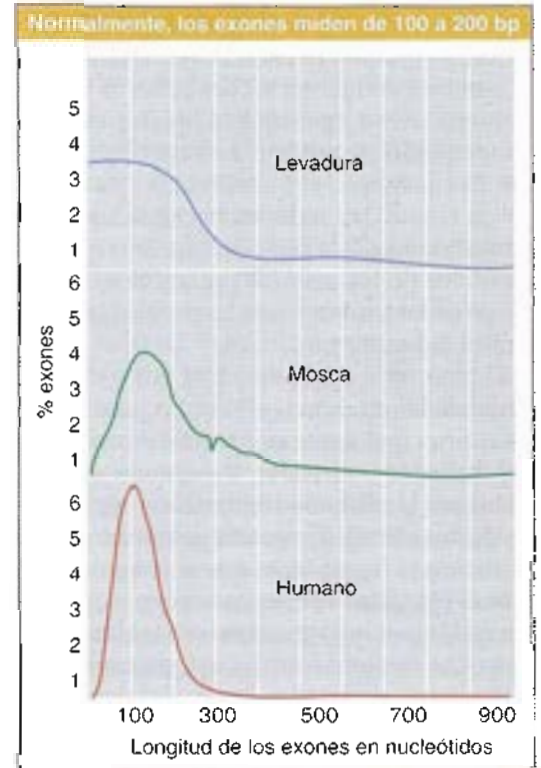


FIGURA 3.12 Los exones que codifican proteínas usualmente son cortos.

con la idea de que los genes han evolucionado por la adición lenta de unidades que codifican a dominios pequeños e individuales de proteínas (véase la sección 3.8, ¿Cómo evolucionaron los genes interrumpidos?). No hay una diferencia significativa en el tamaño de los exones de tipos diferentes de eucariotas superiores, aunque la distribución es más compacta en los vertebrados en que hay pocos exones cuya longitud rebase los 200 bp. En las levaduras, hay algunos exones más largos que representan a genes continuos en los cuales la secuencia codificadora está intacta. Los exones que codifican a regiones 5' y 3' no traducidas tienden a ser más largos que los que codifican proteínas.

La FIGURA 3.13 muestra que el tamaño de los intrones varía ampliamente. En los gusanos y las moscas, el intrón promedio no es mucho más largo que los exones. En el primer caso no hay intrones muy largos, sin embargo las moscas contienen una proporción significativa de ellos. En los vertebrados, la distribución de tamaño es mucho más amplia, desde aproximadamente la misma longitud que los exones (<200 bp) hasta longitudes medidas en decenas de kb, incluso 50 a 60 kb en casos extremos.

Los genes muy largos son resultado de intrones muy largos, no de secuencias codificadoras de productos de gran longitud. No hay una correlación

entre el tamaño del gen y el tamaño del RNAm en las eucariotas superiores, tampoco una buena correlación entre el tamaño del gen y el número de exones, de modo que el tamaño de un gen depende principalmente de las longitudes de cada uno de sus intrones. En los mamíferos, los insectos y las aves, la longitud del gen "promedio" es de aproximadamente cinco veces la de su RNAm.

3.7 Algunas secuencias de DNA codifican a más de una proteína

Conceptos principales

- El uso de codones alternativos de iniciación o de terminación permite que se produzcan dos proteínas cuando una es equivalente a un fragmento de la otra.
- Se pueden producir secuencias no homólogas de proteínas a partir de la misma secuencia de DNA cuando ésta es leída en marcos de lectura diferentes por dos genes (superpuestos).
- Las proteínas homólogas que difieren por la presencia o ausencia de ciertas regiones se deben a un corte y empalme diferencial (alternativo) en el cual se incluyen o excluyen ciertos exones, ya sea por la inclusión o exclusión de exones específicos o al optar por exones alternativos.

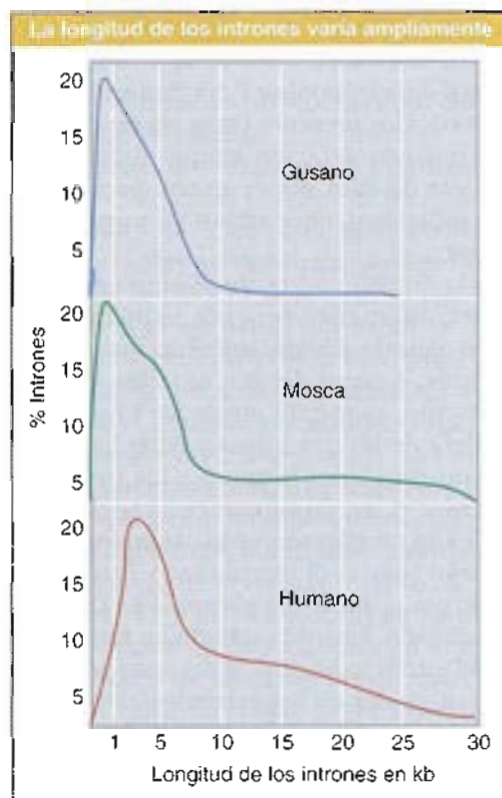


FIGURA 3.13 Los intrones varían de muy cortos a muy largos.

La mayoría de los genes consta de una secuencia de DNA dedicada únicamente a codificar a una proteína (aunque el gen puede incluir regiones no codificadoras en cualquier extremo e intrones en la región codificadora), aunque en algunos casos una sola secuencia de DNA codifica a más de una proteína.

Los genes *superpuestos* se presentan en una situación relativamente sencilla en la cual un gen es parte del otro. La primera mitad de un gen (o la segunda) es utilizada de manera independiente para especificar una proteína que representa a la primera mitad (o a la segunda) de la proteína especificada por el gen completo. Esta relación se ilustra en la FIGURA 3.14. El resultado final es muy similar a una división parcial del producto proteínico para generar formas tanto de tamaño parcial como de tamaño total.

Dos genes pueden superponerse de manera más sutil cuando la misma secuencia de DNA es compartida por dos proteínas *no homólogas*, situación que se presenta cuando la misma secuencia de DNA se traduce en más de un marco de lectura. En los genes celulares, una secuencia de DNA suele leerse sólo en uno de los tres marcos de lectura potenciales. Sin embargo, en algunos genes víricos y mitocondriales hay superposición entre dos genes adyacentes que

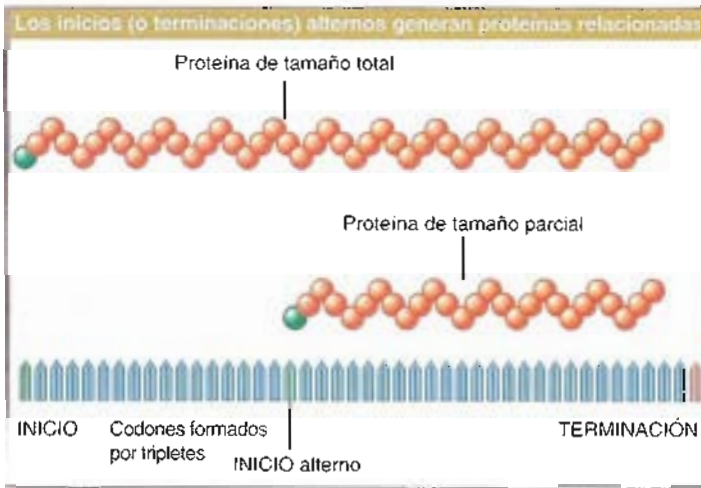


FIGURA 3.14 A partir de un sólo gen se pueden formar dos proteínas que inician (o terminan) su expresión en puntos diferentes.

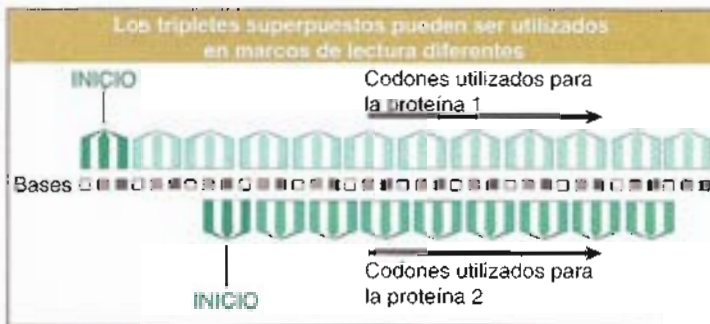


FIGURA 3.15 Dos genes pueden compartir la misma secuencia leyendo el DNA en diferentes marcos de lectura.

se leen en marcos de lectura diferentes, situación ilustrada en la **FIGURA 3.15**. La distancia de la superposición es por lo general relativamente corta, de tal forma que la mayor parte de la secuencia que representa a la proteína conserva una función codificadora única.

En algunos genes, los patrones *alternativos* de la expresión génica crean cambios en la ruta utilizada para conectar los exones. Un sólo gen puede generar diversos productos de RNAm diferentes en cuanto a contenido de exones, y la diferencia puede ser que ciertos exones son opcionales, es decir, pueden ser incluidos o desempalmados. Asimismo, puede haber exones que se excluyan mutuamente, se incluye uno u otro, pero no ambos. Las formas alternas producen proteínas en las cuales una parte es común y la otra es diferente.

En algunos casos, los medios alternos de expresión no afectan a la secuencia de la proteína, por ejemplo, los que inciden en el líder 5' no traduci-

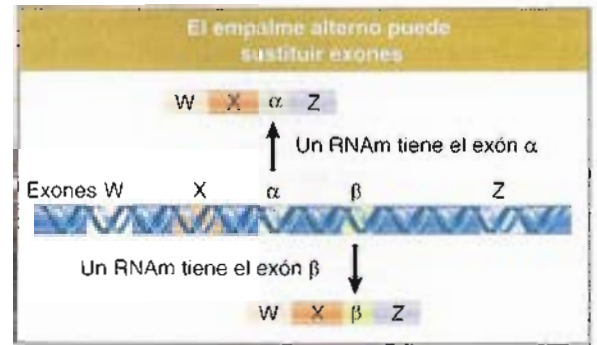


FIGURA 3.16 El corte y empalme alternativo genera las variantes α y β de la troponina T.

do o en el remolque 3' no traducido pueden tener consecuencias reguladoras, pero se crea la misma proteína. En otros casos, un exón es sustituido por otro, como se indica en la **FIGURA 3.16**.

En este ejemplo, las proteínas producidas por los dos RNAm contienen secuencias que se superponen extensamente pero que son diferentes en la región alterna de corte y empalme. La mitad 3' del gen de la troponina T del músculo de la rata contiene cinco exones, pero sólo cuatro se utilizan para constituir un RNAm individual y tres, WXZ, son los mismos en ambos patrones de expresión. No obstante, en uno de éstos, el exón α es empalmado entre X y Z y en el otro se utiliza el exón β , de modo que las formas α y β de la troponina T difieren en la secuencia de aminoácidos presente entre las secuencias W y Z, dependiendo del exón alternativo utilizado, α o β ; cualquiera de ellos puede usarse para formar un RNAm individual, pero ambos no pueden estar en el mismo.

En la **FIGURA 3.17** es un ejemplo en el cual el empalme alternativo provoca la inclusión de un exón en algunos RNAm, mientras que lo deja fuera de otros. A partir del gen se forma un solo tipo de transcrito, pero éste puede ser empalmado en cualquiera de las dos formas. En la primera ruta, se desempalman dos intrones y los tres exones se unen entre sí, en tanto que en la segunda, el segundo exón no es reconocido, de tal forma que un solo intrón largo es desempalmado. Este intrón está formado por el intrón 1 + el exón 2 + el intrón 2. En efecto, el exón 2 ha sido tratado en esta ruta como parte del intrón individual. Estas rutas producen dos proteínas iguales en los extremos, pero una tiene una secuencia adicional en la mitad, por lo tanto, la región del DNA codifica a más de una proteína. (Otros tipos de combinaciones producidas por empalme alternativo se analizan en la sección 26.12,

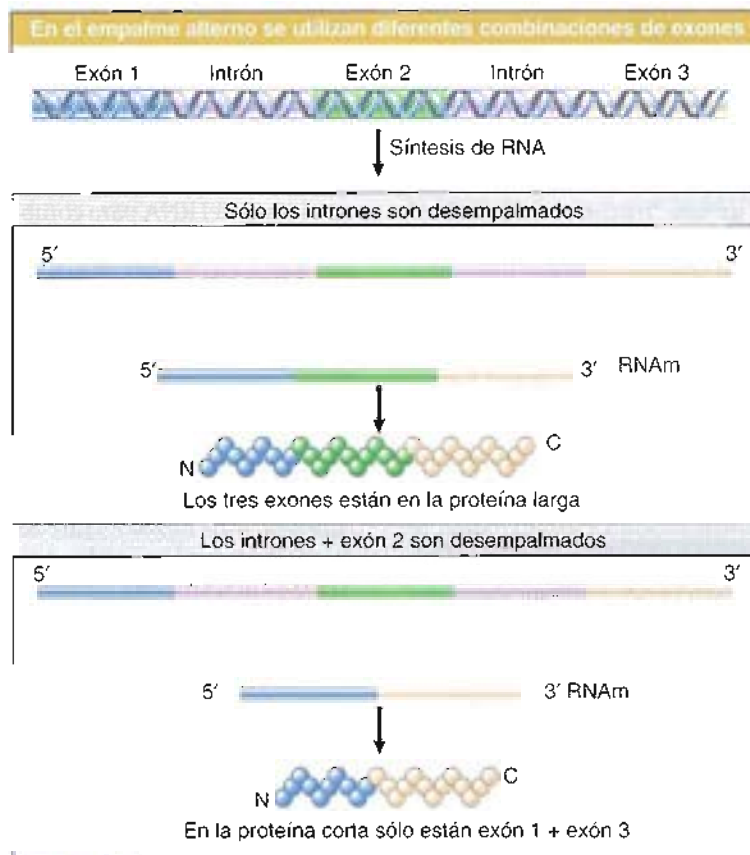


FIGURA 3.17 El corte y empalme alternativo utiliza el mismo pre-RNA para generar moléculas de RNA con combinaciones diferentes de exones.

El empalme alternativo involucra el uso diferencial de uniones de empalme.)

En ocasiones dos rutas operan simultáneamente, en cuyo caso, cierta porción de RNA es empalmada en cada ruta, y en otras, las rutas son alternativas expresadas en condiciones diferentes, una en un tipo de célula y otra en otro.

Así pues, el empalme alternativo o (diferencial) puede generar proteínas con secuencias superpuestas de un solo fragmento de DNA. Es curioso que el genoma eucariótico superior sea extremadamente espacioso y tenga genes grandes que con frecuencia están muy dispersos y que al mismo tiempo forme múltiples productos a partir de un locus individual. El empalme alternativo expande el número de proteínas relativo al número de genes por ~15% en las moscas y en los gusanos, pero tiene efectos mucho mayores en el hombre, en el cual, ~60% de los genes puede tener formas alternas de expresión (véase la Sección 5.5, El genoma humano tiene menos genes que los esperados). Aproximadamente el 80% de los eventos de empalme alternativo resultan en cambios de la secuencia protéica.

3.8

¿Cómo evolucionaron los genes interrumpidos?

Conceptos principales:

- La interrogante principal al respecto es si los genes se originaron como secuencias interrumpidas por intrones o si originalmente eran moléculas ininterrumpidas.
- Probablemente la forma original de la mayoría de los genes codificadores de proteínas fue la interrumpida, si bien en un principio los genes interrumpidos que codifican al RNA pudieron haber sido moléculas ininterrumpidas.
- Una clase especial de intrones es móvil y puede insertarse en los genes.

La estructura ampliamente interrumpida de los genes de las eucariotas sugiere una imagen similar a un mar de intrones (casi todos únicos en secuencia, pero no exclusivamente), en el que las islas de exones (en ocasiones muy cortos) están esparcidas en archipiélagos individuales que representan a los genes.

¿Cuál era la forma original de los genes que actualmente están interrumpidos?

zona de unión 3' introducida por el exón nuevo, y no con su compañero original. De forma similar, la zona de unión 5' del nuevo exón interactuará con la zona de unión 3' del intrón original. El resultado es la inserción del nuevo exón en el producto de RNA entre los dos exones originales. Siempre y cuando el nuevo exón se encuentre en el mismo marco de lectura que los exones originales, se producirá una nueva secuencia proteínica. Este tipo de evento pudo haber generado nuevas combinaciones de exones durante la evolución. Nótese que su principio es imitado por la técnica de captura de exones que se utiliza para identificar a los exones funcionales (véase la Figura 4.11).

En ocasiones se encuentran formas alternas de genes de RNAr y RNAL, con y sin intrones. En el caso de los RNAL, en los cuales todas las moléculas están conformadas por la misma estructura general, parece poco probable que la evolución haya reunido a las dos regiones del gen, después de todo, las diferentes regiones participan en el apareamiento de bases que da importancia a la estructura, de modo que, en este caso, los intrones debieron haber sido insertados en genes continuos.

Los genomas de los organelos proporcionan algunas conexiones notables entre el mundo de las procariontes y el de las eucariontes. Hay muchas similitudes generales entre las mitocondrias o los cloroplastos y las bacterias, por ello, parece probable que los organelos se originaron en una *endosimbiosis* en la cual un prototipo bacteriano ancestral fue insertado en el citoplasma eucariótico. Sin embargo, a diferencia de las similitudes con la bacteria, por ejemplo, cómo se conserva en la síntesis proteínica o del RNA, algunos genes de organelos poseen intrones y por lo tanto se asemejan a los genes nucleares eucarióticos.

En muchos genes de los cloroplastos se encuentran intrones, incluidos algunos homólogos de los genes de la *E. coli*, lo cual sugiere que el evento endosimbiótico ocurrió antes de que la línea procarionte perdiera los intrones. Si se encuentra un gen adecuado, sería posible rastrear el linaje génico hasta al periodo en que ocurrió la endosimbiosis.

El genoma mitocondrial es un caso particularmente sorprendente. Los genes de las mitocondrias de las levaduras y de los mamíferos codifican proteínas mitocondriales virtualmente idénticas, pese a una diferencia considerable en la organización génica. Los genomas mitocondriales de los vertebrados son muy pequeños, con una organización de genes continuos extremadamente compacta, mientras que los genomas mitocondriales de las levaduras son más grandes y tienen algunos genes complejos interrumpidos ¿Cuál es la forma ancestral? Los in-

trones mitocondriales de las levaduras (y algunos otros intrones) pueden tener como propiedad la movilidad, son secuencias autónomas que pueden desempalmarse del RNA e insertar copias de DNA en cualquier otra parte, lo cual sugiere que podrían haber surgido por inserciones en el genoma (véase la sección 27.5. Algunos intrones del grupo I codifican a endonucleasas que promueven la movilidad y la 27.6, Los intrones del grupo II pueden codificar proteínas multifuncionales).

3.9

Algunos exones pueden equipararse con funciones proteínicas

Conceptos principales

- Lo que sugiere que los exones fueron las unidades básicas de la evolución y que los primeros genes fueron interrumpidos en lo siguiente:
 - La estructura génica es igual entre genes de especies muy distantes.
 - Muchos exones pueden ser equiparados con secuencias codificadoras de proteínas que tienen funciones específicas.
 - En genes diferentes se encuentran exones relacionados.

Si las proteínas actuales evolucionaron al combinar proteínas ancestrales que originalmente fueron independientes, es probable que la acreción de unidades haya ocurrido de manera secuencial en algún periodo, agregándose un exón a la vez ¿Es posible ver las diferentes funciones por las que estos genes fueron incrustados en sus estructuras presentes? En otras palabras, ¿podemos equiparar funciones específicas de proteínas actuales con determinados exones?

En algunos casos hay una clara relación entre las estructuras del gen y de la proteína; el ejemplo *por excelencia* es el de las proteínas inmunoglobulinas, las cuales son codificadas por genes en los cuales cada exón corresponde exactamente a un dominio funcional conocido de la proteína. En la **FIGURA 3.19** se compara la estructura de una inmunoglobulina con su gen.

Una inmunoglobulina es un tetrámero formado por dos cadenas ligeras y dos cadenas pesadas unidas para formar una proteína con numerosos dominios distintos. Las cadenas ligeras y las pesadas difieren en cuanto a estructura, y hay numerosos tipos de cadenas pesadas, y cada tipo es expresado desde un gen cuya serie de exones corresponde a los dominios estructurales de la proteína.

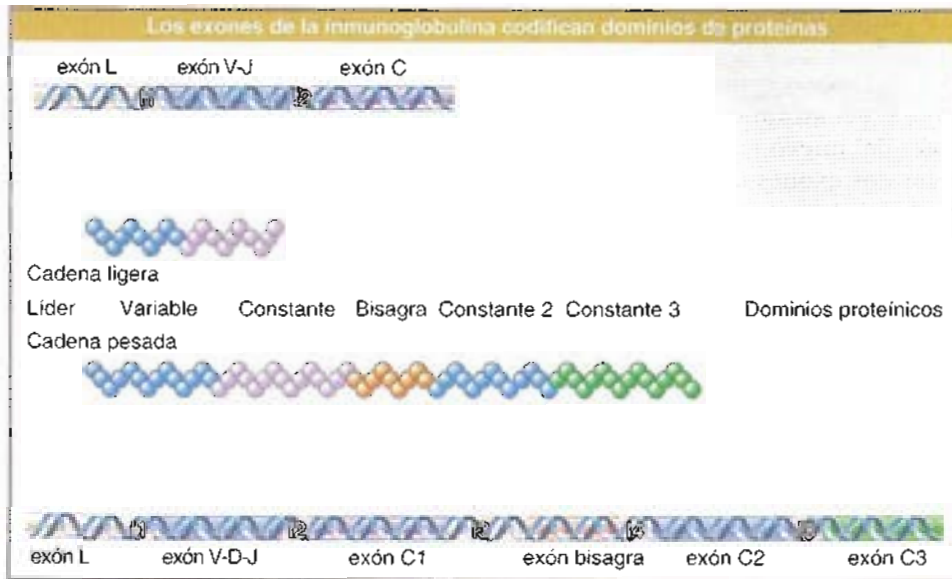


FIGURA 3.19 Las cadenas ligeras y las cadenas pesadas de la inmunoglobulina son codificadas por genes cuyas estructuras (en sus formas expresadas) corresponden a los distintos dominios de la proteína. Cada dominio proteínico corresponde a un exón; los intrones van numerados del 1 al 5.

En muchos casos, algunos de los exones de un gen pueden identificarse con funciones específicas. En proteínas secretoras, el primer exón, que codifica a la región terminal N del polipéptido, con frecuencia específica a la secuencia señal implicada en la secreción membranosa, por ejemplo en la insulina.

La perspectiva de que los exones son las unidades básicas funcionales de los genes está respaldada por casos en que dos genes pueden tener exones relacionados entre sí, mientras que otros se encuentran sólo en uno de los genes. En la **FIGURA 3.20** se resume la relación entre el receptor de la lipoproteína de baja densidad (LDL, *plasma low density lipoprotein*) y otras proteínas. En el centro del gen receptor de la LDL se encuentra una serie de exones relacionados con los exones del gen para el precursor del factor de crecimiento epidérmico (EGF, *epidermal growth factor*). En la región terminal N de la proteína, una serie de exones codifica a una secuencia relacionada con la proteína sanguínea del factor del complemento C9, de modo que el gen del receptor de la LDL se creó ensamblando *módulos* para sus diversas funciones. Estos módulos también son utilizados en combinaciones diferentes en otras proteínas.

Los exones tienden a ser bastante pequeños (véase la Figura 4.11), aproximadamente del tamaño del polipéptido más pequeño susceptible de asumir una estructura plegada estable (~20 a 40 residuos). Quizá las proteínas fueron ensambladas originalmente a partir de módulos más bien pequeños. Cada módulo no necesariamente debe corresponder a una función actual, varios podrían haberse combinado para generar una función. El número de exones de

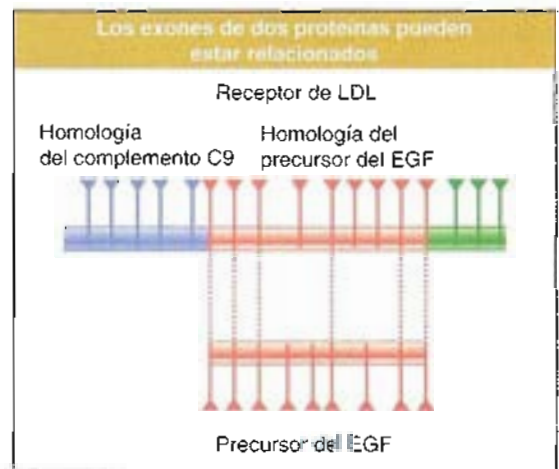


FIGURA 3.20 El gen del receptor de la LDL está formado por 18 exones, algunos de los cuales están relacionados con los exones del precursor del EGF y otros con el gen del complemento sanguíneo C9. Los triángulos marcan las posiciones de los intrones.

un gen tiende a incrementarse con la longitud de su proteína, lo cual concuerda con la perspectiva de que las proteínas adquieren múltiples funciones agregando sucesivamente los módulos apropiados.

Esta idea explicaría otra característica de la estructura de las proteínas. Aparentemente los sitios representados en las fronteras exón-intrón con frecuencia se localizan en la superficie de una proteína. Conforme se agregan módulos a una proteína, las conexiones, cuando menos las de los módulos agregados más recientemente, suelen tener la tendencia a localizarse en la superficie.

3.10 Los miembros de una familia de genes tienen una organización común

Conceptos principales

- Se supone que una característica común de una agrupación de genes identifica una propiedad que antecede a la separación durante la evolución.
- La forma común de organización de todos los genes de globina, tres exones y dos intrones, sugiere que descienden de un solo gen ancestral.

Muchos de los genes de un genoma eucariótico superior están relacionados con otros del mismo genoma. Una **familia de genes** puede definirse como un grupo de genes que codifican a proteínas idénticas o relacionadas; la familia se origina al duplicarse un gen. Inicialmente, las dos copias son idénticas, pero después divergen, conforme se acumulan mutaciones en ellas. Las duplicaciones posteriores y la divergencia hacen aún más extensa la familia. Los genes de la globina son un ejemplo de una familia que puede dividirse en dos subfamilias (globina α y globina β), aunque todos sus miembros tienen la misma estructura básica y la misma función. El concepto puede ampliarse más cuando se encuentran genes relacionados de forma más distante pero cuyo ancestro común aún puede reconocerse; en este caso, se considera que un grupo de familias de genes forma una **superfamilia**.

Las globinas α y β , y otras dos proteínas relacionadas con ellas, constituyen un caso fascinante de conservación evolutiva. La mioglobina es una proteína monomérica que se liga al oxígeno, está presente en los animales y su secuencia de aminoácidos sugiere un origen común (aunque antiguo) con las subunidades de la globina. Las leghemoglobinas son proteínas que se unen al oxígeno y que se encuentran en las legumbres; igual que la mioglobina, son monoméricas. Además, también comparten un origen común con las otras proteínas de unión al grupo hemo. Juntas, las globinas, la mioglobina y la leghemoglobina constituyen la superfamilia de la globina, grupo de familias de genes descendientes de algún ancestro (distante) común.

Los genes de la globina α y de la globina β tienen tres exones (véase la Figura 3.7). Los dos intrones están en posiciones constantes respecto de la secuencia codificadora. El exón central representa al dominio de unión al grupo hemo de la cadena de globina.

En el genoma humano, la mioglobina está representada por un solo gen cuya estructura es esencialmente la misma que la de los genes de la globina, de modo que la estructura de tres exones antecede

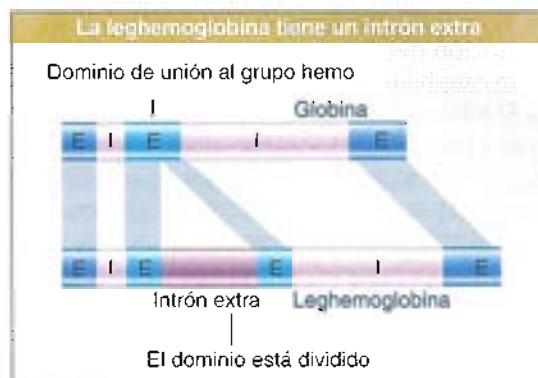


FIGURA 3.21 La estructura de los exones de los genes de globina corresponde a la función proteínica, pero la leghemoglobina tiene un intrón extra en el dominio central.

a la evolución de las funciones separadas de la mioglobina y de la globina.

Los genes de la leghemoglobina contienen tres intrones, de los cuales, el primero y el último se presentan en puntos de la secuencia codificadora homólogos a las localizaciones de los dos intrones de los genes de globina. Esta notable similitud sugiere un origen extremadamente antiguo de las proteínas de unión al grupo hemo en forma de un gen dividido, como se ilustra en la **FIGURA 3.21**.

El intrón central de la leghemoglobina separa dos exones que juntos codifican a la secuencia correspondiente al exón central individual de la globina. ¿Pudo el exón central del gen de la globina haber sido derivado de una fusión de dos exones centrales del gen ancestral? ¿O es el exón individual central la forma ancestral, en este caso, un intrón que debió haber sido insertado en él al inicio de la evolución de la planta?

Los casos en que la estructura de genes homólogos difiere, podrían proporcionar información acerca de su evolución, por ejemplo, la insulina. Los mamíferos y las aves tienen sólo un gen para codificar insulina, excepto los roedores, que tienen dos. En la **FIGURA 3.22** se ilustran las estructuras de estos genes.

El principio que se utiliza para comparar la organización de genes relacionados en especies diferentes consiste en que *una característica común identifica a una estructura que precedió a la separación evolutiva de las dos especies*. En los pollos, el gen único de la insulina tiene dos intrones; uno de los dos genes de la rata tiene la misma estructura. La estructura común implica que el gen ancestral de la insulina tenía dos intrones, sin embargo, el segundo gen de las ratas tiene sólo uno, de modo que debe haber evolucionado por una duplicación génica en los roedores seguida de la eliminación precisa de un intrón de una de las copias.

La organización de algunos genes muestra discrepancias considerables entre especies, en cuyo

caso, la eliminación o inserción de intrones durante la evolución debió haber sido intensa.

Un caso bien caracterizado son los genes de la actina. El típico gen de la actina tiene un líder no traducido de <100 bases, una región codificadora de ~1200 bases y un remolque de ~200 bases. La mayoría de los genes de la actina son interrumpidos; las posiciones de los intrones pueden alinearse de acuerdo con la secuencia codificadora (excepto por un único intrón que en ocasiones se encuentra en el líder).

En la FIGURA 3.23 se observa que el patrón de interrupciones de casi todos los genes de la actina es diferente, y agrupando a todos los genes, los intrones ocurren en 19 sitios diferentes. Sin embargo, ningún gen tiene más de seis intrones, algunos sólo tienen uno y uno de ellos es completamente continuo ¿Cómo surgió esta situación? Si suponemos que el gen primordial de la actina era interrumpido,

y que todos los genes de la actina están relacionados con él por pérdida de intrones, en cada rama evolutiva se han perdido intrones diferentes. Probablemente algunos se han perdido por completo, de modo que el gen primordial pudo bien haber tenido 20 o más. La alternativa es suponer que un proceso de inserción de intrones continuó de manera independiente en las diferentes líneas de evolución. En última instancia, la relación entre las localizaciones de los intrones en diferentes especies puede utilizarse para construir un árbol de la evolución del gen.

La relación entre los exones y los dominios de proteínas es un tanto errática; en ciertos casos es claramente de 1:1 y, en otros, no es posible discernir un patrón. Una posibilidad es que por la eliminación de intrones se han fusionado los exones adyacentes, lo cual significa que el intrón debió haber sido eliminado de forma precisa, sin modificación de la integridad de la región codificadora. Una alternativa es que algunos intrones surgieron por inserción en un dominio coherente. Además de las variaciones observadas en la colocación de los exones en casos como el de los genes de la actina, esto sugiere que las posiciones de los intrones pueden ser ajustadas en el curso de la evolución.

La ecuación de cuando menos algunos exones con los dominios de proteínas y la aparición de exones relacionados en proteínas diferentes, no deja duda de que la duplicación y la yuxtaposición de exones ha desempeñado un papel importante en la evolución. Es posible que el número de exones ancestrales, de los cuales se han derivado todas las proteínas por duplicación, variación y recombinación, sea relativamente pequeño (algunos miles o decenas de miles). Considerando al exón como unidad fundamental de la evolución, desde esta perspectiva se acepta implí-

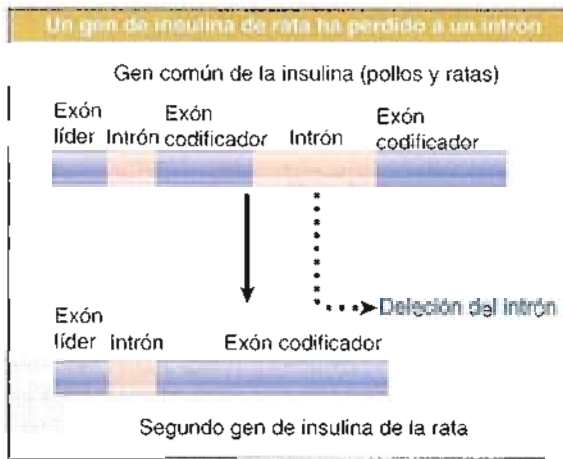


FIGURA 3.22 El gen de la insulina de la rata, el cual cuenta con un intrón, evolucionó por la pérdida de un intrón de un ancestro con dos intrones.

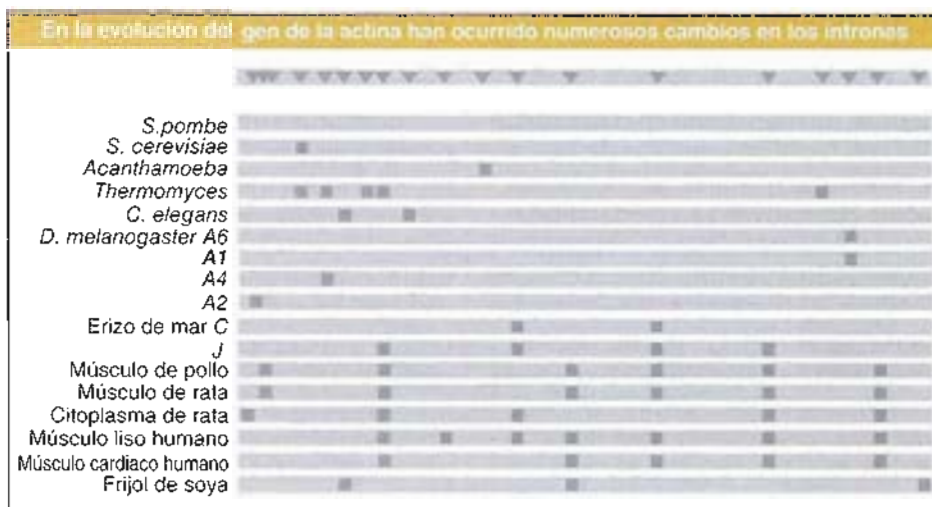


FIGURA 3.23 Los genes de la actina varían ampliamente en cuanto a organización. Los sitios de los intrones se indican en tono más oscuro.

citamente el modelo de los intrones ancestrales como origen de los genes que codifican proteínas.

3.11 ¿Se encuentra toda la información genética contenida en el DNA?

Conceptos principales

- La definición del gen se ha invertido, de "un gen : una proteína" a "una proteína : un gen".
- La información sobre la posición también es importante en el desarrollo.

El concepto del gen ha evolucionado significativamente en los últimos años. La interrogante respecto de qué hay en un nombre es especialmente apropiada para el gen. Ya no podemos aseverar que éste sea una secuencia de DNA que codifica continua y únicamente a una proteína en particular. En situaciones en las que un fragmento de DNA es responsable de la producción de una proteína específica, hoy día se considera que la secuencia entera de DNA —desde el primer punto representado en el RNA mensajero hasta el último correspondiente a su extremo— comprende al "gen", los exones, los intrones y todo lo demás.

Cuando las secuencias que representan a proteínas se superponen o tienen formas alternas de expresión, es posible revertir la descripción común del gen, y en vez de decir "un gen-un polipéptido", podemos describir a la relación como "un polipéptido-un gen", de modo que se considera a la secuencia responsable de hecho de la producción del polipéptido (incluidos intrones y exones) como productora del gen, al mismo tiempo que se acepta que desde la perspectiva de otra proteína, parte de esta misma secuencia también pertenece a *su* gen. Esto permite el uso de descripciones como genes "superpuestos" o "alternativos".

Ahora podemos ver qué tan lejos se ha llegado desde la hipótesis original de un gen:una enzima. Hasta ese momento, la pregunta fundamental se relacionaba con la naturaleza del gen, pero una vez que se descubrió que los genes representan a proteínas, el paradigma se estableció como el concepto de que cada unidad genética funciona a través de la síntesis de una proteína específica.

Este punto de vista sigue siendo el paradigma central de la biología molecular: una secuencia de DNA funciona codificando directamente a una proteína en particular o porque es necesaria para el uso de un segmento adyacente que codifica realmente a la proteína. ¿Qué tan lejos nos lleva este paradigma

aparte de explicar la relación básica entre los genes y las proteínas?

El desarrollo de organismos multicelulares se sustenta en el uso de genes diferentes para generar los diversos fenotipos celulares de cada tejido. La expresión de los genes depende de una red reguladora que asume la forma de una cascada. La expresión del primer conjunto de genes al principio del desarrollo embrionario conduce a la expresión de los genes implicados en la siguiente etapa de desarrollo, y así sucesivamente, hasta que todos los tejidos del adulto son funcionales. La naturaleza molecular de esta red reguladora es en gran medida desconocida, pero suponemos que consiste en genes que codifican a productos (probablemente proteínas, quizá en ocasiones RNA) que actúan sobre otros genes.

Si bien es casi seguro que dicha serie de interacciones es el medio por el cual se ejecuta el programa de desarrollo, es posible preguntarse si es suficiente. Una pregunta específica concierne a la naturaleza y la función de la **información posicional**. Sabemos que no todas las partes de un óvulo fertilizado son iguales: una de las características responsables del desarrollo de diferentes partes de tejidos de distintas regiones del óvulo es la localización de la información (presumiblemente macromoléculas específicas) dentro de la célula.

Desconocemos cómo se forman esas regiones particulares, pero se puede especular que la existencia de la información posicional en el óvulo conduce a la expresión diferencial de los genes en las células formadas después en estas regiones. Esto conduce al desarrollo del organismo adulto, lo cual a su vez da lugar al desarrollo de un óvulo con la información posicional apropiada.

Esta posibilidad incita a preguntar si información necesaria para el desarrollo de organismos está contenida en una forma que no pueda ser atribuida directamente a una secuencia de DNA (aunque la expresión de secuencias particulares puede ser necesaria para perpetuar la información posicional). De una manera más general, es posible preguntar lo siguiente: cuando leemos la secuencia completa de DNA que comprende al genoma de algún organismo y la interpretamos en función de proteínas y regiones reguladoras, ¿se podría, en principio, construir un organismo (o incluso una sola célula viviente) por medio de la expresión controlada de los genes adecuados?

3.12 Resumen

Todos los tipos de genomas eucarióticos contienen genes interrumpidos. La proporción de éstos es baja

en las levaduras y se incrementa en las eucariotas inferiores; en las eucariotas superiores son pocos los genes continuos.

Los intrones se encuentran en todas las clases de genes eucarióticos. La estructura del gen interrumpido es la misma en todos los tejidos: en el RNA, los exones se unen en el mismo orden en que están organizados en el DNA y los intrones usualmente carecen de función codificadora. Los intrones se eliminan del RNA por corte y empalme. Algunos genes se expresan mediante patrones alternos de corte y empalme, en los cuales una secuencia específica es eliminada como intrón en algunas situaciones, pero retenida como exón en otras.

Las posiciones de los intrones suelen conservarse cuando se compara la organización de genes homólogos entre especies. Las secuencias de intrones varían, incluso pueden no estar relacionadas, aunque las secuencias de exones se mantienen bien relacionadas. La conservación de los exones puede utilizarse para aislar genes relacionados en especies diferentes.

El tamaño de un gen depende sobre todo de las longitudes de sus intrones, que en las eucariotas superiores se alargaron en las primeras etapas, cuando, por lo tanto, las dimensiones de los genes se incrementaron significativamente. La gama de dimensiones de los genes de los mamíferos suele fluctuar entre 1 y 100 kb, pero es posible que sean aún más largos: el caso conocido es el de la distrofia, que mide 2000 kb.

Algunos genes comparten sólo algunos de sus exones con otros genes, lo cual sugiere que han sido ensamblados por adición de exones que representan a módulos individuales de la proteína. Dichos módulos podrían haber sido incorporados en una gran variedad de proteínas diferentes. La idea de que los genes se han ensamblado por acreción de exones implica que en los genes de organismos primitivos hubo intrones. Algunas de las relaciones entre genes homólogos pueden explicarse por la pérdida de intrones de los genes primordiales, con pérdida de intrones diferentes en líneas de ascendencia diferentes.

Referencias

3.2 Un gen interrumpido está formado por intrones y exones

Revisiones

Breathnach, R. and Chambon, P. (1981). Organization and expression of eukaryotic split genes coding for proteins. *Annu. Rev. Biochem.* 50, 349-383.

Faustino, N. A. and Cooper, T. A. (2003). Pre-mRNA splicing and human disease. *Genes Dev.* 17, 419-437.

3.3 Las endonucleasas de restricción son una herramienta esencial en el mapeo del DNA

Revisiones

Nathans, D. and Smith, H. O. (1975). Restriction endonucleases in the analysis and restructuring of DNA molecules. *Annu. Rev. Biochem.* 44, 273-293.

Wu, R. (1978). DNA sequence analysis. *Annu. Rev. Biochem.* 47, 607-734.

Artículos de investigación

Danna, K. J., Sack, G. H., and Nathans, D. (1973). Studies of SV40 DNA VII A cleavage map of the SV40 genome. *J. Mol. Biol.* 78, 363-376.

3.4 La organización de los genes interrumpidos puede ser conservada

Artículos de investigación

Berget, S. M., Moore, C., and Sharp, P. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci. USA* 74, 3171-3175.

Chow, L. T., Gelinas, R. E., Broker, T. R., and Roberts, R. J. (1977). An amazing sequence arrangement at the 5' ends of adenovirus 2 mRNA. *Cell* 12, 1-8.

Glover, D. M. and Hogness, D. S. (1977). A novel arrangement of the 8S and 28S sequences in a repeating unit of *D. melanogaster* rDNA. *Cell* 10, 167-176.

Jeffreys, A. J. and Flavell, R. A. (1977). The rabbit β -globin gene contains a large insert in the coding sequence. *Cell* 12, 1097-1108.

Wensink, P. et al. (1974). A system for mapping DNA sequences in the chromosomes of *D. melanogaster*. *Cell* 3, 315-325.

3.9 Algunos exones pueden ser equiparados con funciones proteínicas

Revisiones

Blake, C.C. (1985). Exons and the evolution of proteins. *Int. Rev. Cytol.* 93, 149-185.

El contenido del genoma

ESQUEMA DEL CAPÍTULO

4.1 Introducción

4.2 Pueden trazarse mapas de los genomas por ligamiento, restricción, división o secuencia de DNA

4.3 Los genomas individuales son muy variables

- El polimorfismo puede ser detectado en el nivel fenotípico cuando una secuencia afecta a la función génica, en el de los fragmentos de restricción cuando afecta a un sitio diana de una enzima de restricción y en el secuencial por análisis directo del DNA.
- Los alelos de un gen muestran un amplio polimorfismo en el nivel secuencial, pero muchos cambios de secuencia no inciden en la función.

4.4 Los RFLP y los SNP pueden ser utilizados para el mapeo genético

- Los RFLP y los SNP pueden constituir la base de los mapas de ligamiento; permiten establecer relaciones entre progenitores y su progenie.

4.5 ¿Por qué los genomas son tan grandes?

- No hay una buena correlación entre el tamaño del genoma y la complejidad genética.
- Hay un incremento en el tamaño mínimo del genoma necesario para formar organismos de mayor complejidad.
- En numerosos filos, las dimensiones de los genomas de los organismos varían ampliamente.

4.6 Los genomas eucarióticos contienen secuencias de DNA repetitivas y no repetitivas

- La cinética de la reasociación del DNA después de que un genoma ha sido desnaturalizado distingue a las secuencias por la frecuencia con se repiten en el genoma.
- En el DNA no repetitivo, los genes generalmente son codificados por secuencias localizadas.
- Los genomas más grandes de un filo no contienen más genes, pero tienen cantidades mayores de DNA repetitivo.
- Gran parte del DNA repetitivo puede estar formada por transposones.

4.7 Los genes pueden ser aislados por la conservación de los exones

- La conservación de los exones puede ser utilizada como base para localizar regiones codificadoras al identificar fragmentos cuyas secuencias están presentes en múltiples organismos.

4.8 La conservación de la organización del genoma ayuda a identificar genes

- Los algoritmos para identificar genes no son perfectos, de modo que son necesarias numerosas correcciones al conjunto inicial de datos.
- Los pseudogenes deben distinguirse de los genes activos.
- Hay amplias relaciones sinténicas entre los genomas del ratón y los del humano, y la mayoría de los genes activos se encuentran en una región sinténica.

4.9 Los organelos contienen DNA

- Las mitocondrias y los cloroplastos tienen genomas cuya herencia es no mendeliana; típicamente son heredados de la madre.
- Los genomas de los organelos pueden ser sometidos a segregación somática en las plantas.
- Las comparaciones del DNA mitocondrial sugieren que los humanos descienden de una sola hembra, la cual vivió hace 200 000 años en África.

4.10 Los genomas de los organelos son moléculas circulares de DNA que codifican proteínas de los organelos

- Los genomas de los organelos suelen ser moléculas circulares de DNA (pero no siempre).
- Los genomas de los organelos codifican algunas de las proteínas que se encuentran en el organelo, pero no a todas.

4.11 La organización del DNA mitocondrial es variable

- El DNA mitocondrial de las células animales es extremadamente compacto y suele codificar a 13 proteínas, dos moléculas de RNAr y 22 de RNAt.
- El DNA mitocondrial de las levaduras es cinco veces más largo que el DNAm de las células animales por la presencia de intrones largos.

4.12 El genoma de los cloroplastos codifica a numerosas proteínas y moléculas de RNA

- Los genomas de los cloroplastos varían en cuanto a tamaño, pero son lo suficientemente grandes como para codificar de 50 a 100 proteínas, así como a los RNAr y a los RNAt.

4.13 Las mitocondrias evolucionaron por endosimbiosis

4.14 Resumen

4.1 Introducción

La pregunta crucial sobre el genoma es cuántos genes contiene. Se puede pensar acerca del número total de genes en cuatro niveles, los cuales corresponden a etapas sucesivas de la expresión génica:

- El **genoma** es el conjunto completo de los genes de un organismo. En última instancia se define por la secuencia completa de DNA, aunque por una cuestión práctica podría no ser posible identificar cada gen inequívocamente sólo con base en la secuencia.
- El **transcriptoma** es el conjunto completo de los genes expresados en determinadas condiciones. Se define en función del conjunto de moléculas de RNA presente y puede referirse a un solo tipo de célula, a cualquier ensamble de células más complejo o al organismo completo. Debido a que algunos genes generan múltiples moléculas de RNAm, es probable que el transcriptoma sea mayor que el número de genes definidos directamente en el genoma. El transcriptoma incluye a las moléculas de RNA no codificadoras, así como a las moléculas de RNAm.
- El **proteoma** es el conjunto completo de proteínas. Debe corresponder a las moléculas de RNAm del transcriptoma, aunque puede haber diferencias de detalle que reflejen cambios de la abundancia relativa o la estabilidad de las moléculas de RNAm y de las proteínas. Puede ser utilizado para referirse al conjunto de proteínas codificadas por todo el genoma o producidas en cualquier célula o tejido específicos.
- Las proteínas pueden funcionar de manera independiente o como parte de ensamblajes de múltiples proteínas. Si fuera posible identificar todas las interacciones entre proteína y proteína, se podría definir el número total de ensamblajes independientes de proteínas.

El número de genes del genoma puede ser identificado directamente al definir marcos de lectura abiertos. El mapeo de esta naturaleza a gran escala es complicado porque los genes interrumpidos pueden estar formados por numerosos marcos de lectura abiertos separados. No necesariamente tenemos información acerca de las funciones de los productos proteínicos, o de hecho, una prueba de que lleguen a expresarse, de modo que este abordaje se restringe a la definición del *potencial* del genoma. Sin embargo, se presume con relativa certeza que cualquier marco de lectura abierto conservado probablemente llegue a expresarse.

Otro abordaje consiste en definir el número de genes directamente en función del transcriptoma (identificando directamente todas las moléculas de RNAm) o del proteoma (identificando directamente

todas las proteínas). Esto proporciona la seguridad de que se trata con genes *auténticos* que se expresan en circunstancias conocidas y permite investigar cuántos genes son expresados en un tipo de célula o un tejido particular, qué variación existe en los niveles relativos de expresión y cuántos de los genes expresados en una célula particular son únicos de esa célula o también son expresados en alguna otra parte.

Respecto de los tipos de genes, podemos preguntar si alguno en particular es esencial: ¿qué le sucede a un mutante nulo? Si una mutación nula es letal, o el organismo presenta un defecto visible, podemos concluir que el gen es esencial o que al menos transfiere una ventaja selectiva, pero algunos genes pueden ser suprimidos sin un efecto aparente en el fenotipo ¿Son estos genes realmente prescindibles, o resulta una desventaja selectiva por la ausencia del gen, quizás en otras circunstancias, o en periodos más largos?

4.2 Pueden trazarse mapas de los genomas por ligamiento, por restricción, por división o por secuencia de DNA

Definir esencialmente el contenido de un genoma significa trazar un mapa de genes y genomas en muchos niveles de resolución:

- Mediante un mapa genético (o de ligamiento) se identifica la distancia entre mutaciones en cuanto a frecuencias de recombinación. Lo limita su dependencia en la ocurrencia de mutaciones que afectan al fenotipo. Debido a que las frecuencias de recombinación pueden distorsionarse respecto de la distancia física entre los sitios, no representa con precisión distancias físicas a lo largo del material genético.
- Un mapa de ligamiento puede construirse midiendo la recombinación entre sitios en el DNA genómico. Estos sitios tienen variaciones secuenciales que generan diferencias en la susceptibilidad de división por ciertas enzimas (de restricción); como dichas variaciones son comunes, es posible hacer un mapa de ligamiento para cualquier organismo, a pesar de la ocurrencia de mutantes. Su desventaja es la misma para cualquier mapa de ligamiento, es decir, que las distancias relativas se basan en la recombinación.
- Un mapa de restricción se construye dividiendo el DNA en fragmentos con enzimas de restricción y midiendo las distancias entre los sitios de división. Este mapa representa distancias en cuanto a longitud del DNA, por lo tanto, proporciona un mapa físico del

material genético. Con un mapa de restricción no se identifican intrínsecamente los sitios de interés genético, para relacionarlo con el mapa genético, las mutaciones deben ser caracterizadas en función de sus efectos en los sitios de restricción. Pueden reconocerse grandes cambios en el genoma porque afectan a dimensiones o el número de fragmentos de restricción. Las mutaciones puntuales son más difíciles de detectar.

- El último mapa permite determinar la secuencia del DNA. A partir de la secuencia se identifican los genes y las distancias entre ellos. Al analizar el potencial de codificación de proteínas de una secuencia de DNA se puede deducir si representa a una proteína, en cuyo caso, la conjetura básica es que la selección natural impide la acumulación de mutaciones dañinas en las secuencias que codifican proteínas. Invirtiendo el argumento, se puede suponer que es probable que una secuencia codificadora intacta sea utilizada para formar proteína.

Al comparar la secuencia de un DNA de tipo silvestre con la de un alelo mutante, es factible determinar la naturaleza de una mutación y el sitio exacto en que aparecerá. Esto define la relación entre el mapa genético (el cual se basa completamente en los sitios de mutación) y el mapa físico (basado en la secuencia del DNA, incluso puede estar formado por ésta).

Técnicas similares son utilizadas para identificar y secuenciar genes y para trazar mapas del genoma, aunque por supuesto hay una diferencia de escala. En cada caso, el principio consiste en obtener una serie de fragmentos superpuestos de DNA que puedan ser conectados en un mapa continuo. La característica clave consiste en que cada segmento está relacionado con el siguiente segmento en el mapa caracterizando la superposición entre ellos, de tal forma que podemos estar seguros de que no faltan segmentos. Este principio es aplicado para ordenar grandes fragmentos en un mapa y para conectar las secuencias que constituyen a los fragmentos.

4.3 Los genomas individuales son muy variables

Conceptos principales

- El polimorfismo puede ser detectado en el nivel fenotípico cuando una secuencia afecta a la función génica, en el de los fragmentos de restricción cuando afecta a un sitio diana de una enzima de restricción y en el secuencial por análisis directo del DNA.
- Los alelos de un gen muestran un amplio polimorfismo en el nivel secuencial, pero muchos cambios de secuencia no inciden en la función.

En la perspectiva mendeliana original del genoma, los alelos se clasificaban como silvestres o mutantes; posteriormente se aceptó la existencia de alelos múltiples, cada uno con un efecto diferente en el fenotipo. En algunos casos, incluso podría no ser apropiado definir algún alelo como de "tipo silvestre".

La coexistencia de alelos múltiples en un locus se denomina **polimorfismo genético**. Por definición, cualquier sitio en el cual existen alelos múltiples como componentes estables de la población es **polimórfico**. Un alelo suele ser definido como **polimórfico** si su frecuencia en la población es de >1 por ciento.

¿Cuál es la base del polimorfismo entre los alelos mutantes? Poseen diferentes mutaciones que modifican la función proteínica, produciendo cambios en el fenotipo. Si se comparan los mapas de restricción de las secuencias de DNA de estos alelos, éstos también serán **polimórficos**, en el sentido de que cada mapa o secuencia será diferente de los otros.

Aunque no es evidente a partir del fenotipo, el tipo silvestre mismo puede ser **polimórfico**. Las versiones múltiples del alelo de tipo silvestre pueden distinguirse por diferencias secuenciales que no afectan a su función y que, por lo tanto, no producen variantes fenotípicas. Una población puede tener gran polimorfismo en el nivel del genotipo. En un locus determinado pueden existir muchas variantes de secuencia diferentes; algunas son evidentes porque afectan al fenotipo, pero otras están ocultas porque su efecto no es visible.

De manera que puede haber una sucesión de cambios en un locus, incluidos los que cambian la secuencia del DNA pero que no alteran la secuencia de la proteína, los que transforman la secuencia de la proteína sin afectar su función, los que crean proteínas con actividades diferentes y los que producen proteínas mutantes no funcionales.

Cuando se comparan los alelos, un cambio en un solo nucleótido se denomina **polimorfismo de un solo nucleótido (SNP, single nucleotide polymorphism)**. Uno de estos cambios se presenta cada ~ 1300 bases en el genoma humano. Definido por sus SNP, cada ser humano es único. Los SNP pueden ser detectados por diversos medios, desde comparaciones directas de secuencia hasta métodos bioquímicos o de espectroscopia de masas que producen diferencias basadas en variaciones de secuencia en una región definida.

Un objetivo del mapeo genético es obtener un catálogo de variantes comunes. La frecuencia observada de SNP por genoma pronostica que, respecto de la población humana en general (considerando la suma de todos los genomas humanos de todos los individuos vivientes), debe haber >10 millones de SNP que ocurren a una frecuencia de >1 por ciento; ya se ha identificado >1 millón.

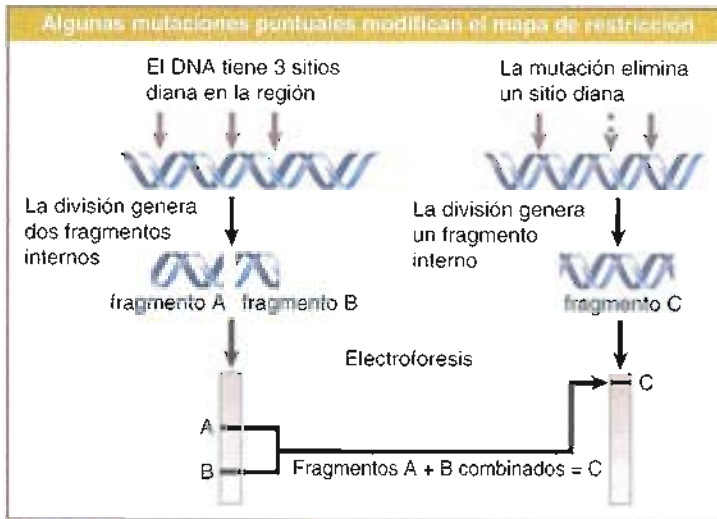


FIGURA 4.1 Una mutación puntual que afecta a un sitio de restricción se detecta por una diferencia en los fragmentos de restricción.

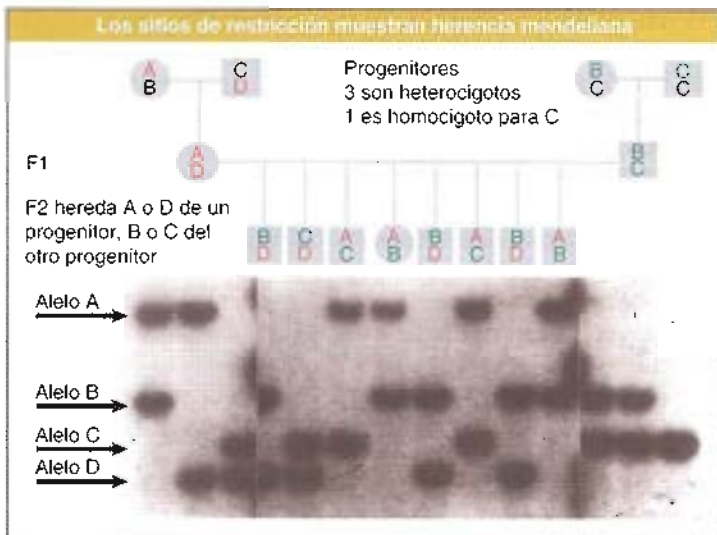


FIGURA 4.2 Los polimorfismos de los sitios de restricción se heredan según las leyes de Mendel. Hay cuatro alelos para un marcador de restricción en todas las combinaciones de pares posibles; se segregan de forma independiente en cada generación. La fotografía es cortesía de Ray White, Ernest Gallo Clinic and Research Center, University of California, San Francisco.

Algunos polimorfismos del genoma pueden detectarse al comparar los mapas de restricción de individuos diferentes. El criterio es un cambio en el patrón de fragmentos producidos por división con una enzima de restricción. En la **FIGURA 4.1** se muestra que cuando en el genoma de un individuo hay un sitio diana y en el de otro no, la división extra en el primer genoma generará dos fragmentos que corresponden al fragmento individual del segundo genoma.

El mapa de restricción es independiente de la función génica, de modo que en este nivel, un po-

limorfismo puede ser detectado a pesar de que el cambio de secuencia afecte al fenotipo. Probablemente muy pocos de los polimorfismos de los sitios de restricción de un genoma realmente afecten al fenotipo, la mayoría implican cambios de secuencia que pueden no tener efecto en la producción de proteínas (por ejemplo, debido a que se encuentran entre genes).

Una diferencia en los mapas de restricción de dos individuos es denominado **polimorfismo de la longitud de los fragmentos de restricción (RFLP, restriction fragment length polymorphism)**. Básicamente, un RFLP es un SNP localizado en el sitio diana de una enzima de restricción que puede utilizarse como marcador genético exactamente de la misma manera que cualquiera otro marcador. En vez de examinar alguna característica del fenotipo, se evalúa directamente el genotipo, como lo revela el mapa de restricción. En la **FIGURA 4.2** se observa la genealogía de un polimorfismo de restricción a través de tres generaciones; exhibe segregación mendeliana en los fragmentos de marcadores de DNA.

4.4 Los RFLP y los SNP pueden ser utilizados para el mapeo genético

Concepto principal

- Los RFLP y los SNP pueden constituir la base de los mapas de ligamiento; permiten establecer relaciones entre progenitores y su progenie.

La frecuencia de recombinación puede medirse entre un marcador de restricción y un marcador fenotípico visible, como se ilustra en la **FIGURA 4.3**, de tal manera que un mapa genético puede incluir marcadores genotípicos y fenotípicos.

Los marcadores de restricción no se limitan a los cambios del genoma que afectan al fenotipo, de modo que proporcionan las bases para una técnica extremadamente poderosa de identificación de loci genéticos en el ámbito molecular. Un problema típico concierne a una mutación con efectos conocidos en el fenotipo, en el cual el locus genético pertinente puede ser colocado en un mapa genético, pero de la cual se desconoce el gen o la proteína correspondiente. Muchas enfermedades humanas dañinas o fatales pertenecen a esta categoría. Por ejemplo, la fibrosis quística muestra herencia mendeliana, pero se desconocía la naturaleza molecular de la función mutante hasta que se pudo identificar como resultado de la caracterización del gen.

Si los polimorfismos de restricción son aleatorios en el genoma, algunos deben ocurrir cerca de un gen diana específico. Dichos marcadores de restricción se

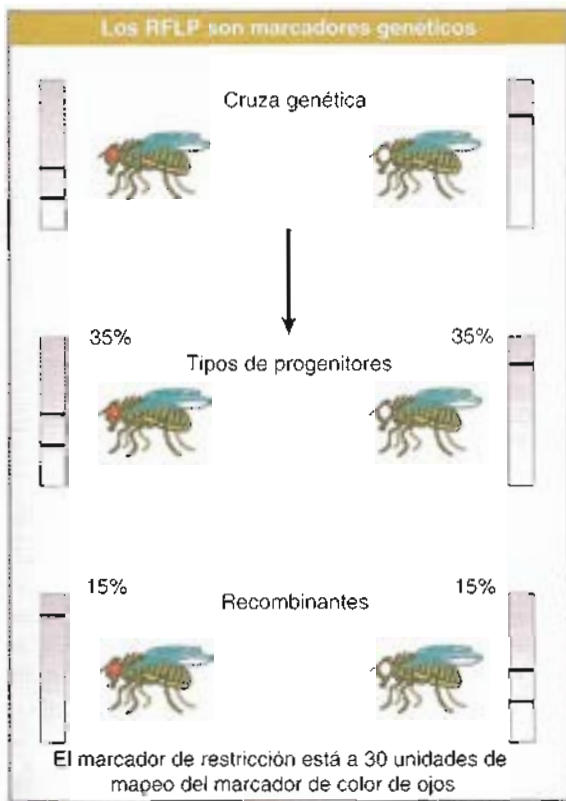


FIGURA 4.3 Un polimorfismo de restricción puede ser utilizado como marcador genético para medir la distancia de recombinación a partir de un marcador fenotípico (como el color de los ojos). En la figura se simplifica la situación mostrando sólo las bandas de DNA que corresponden al alelo de un genoma en un diploide.

identifican gracias a su estrecha vinculación con el fenotipo mutante. Si se compara el mapa de restricción del DNA de pacientes que padecen una enfermedad con el DNA de personas sanas, podría encontrarse que, en los pacientes, siempre hay un sitio de restricción en particular (o siempre está ausente).

En la **FIGURA 4.4** se muestra un ejemplo hipotético que corresponde al descubrimiento de una vinculación de 100% entre el marcador de restricción y el fenotipo, lo cual implicaría que el marcador de restricción está tan cerca del gen mutante, que nunca se separa de él por recombinación.

La identificación de un marcador de tales características tiene dos consecuencias importantes:

- Puede ofrecer un procedimiento diagnóstico para detectar la enfermedad. Algunas de las enfermedades humanas bien caracterizadas genéticamente pero mal definidas en función de las moléculas, no son fáciles de diagnosticar. Si un marcador de restricción está ligado confiablemente al fenotipo, puede servir para diagnosticar la enfermedad.
- Puede conducir al aislamiento del gen. El marcador de restricción debe encontrarse

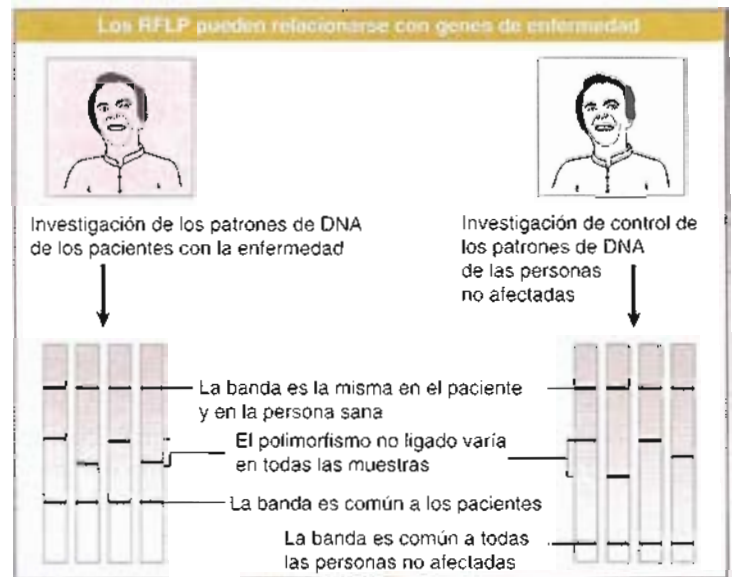


FIGURA 4.4 Si un marcador de restricción es asociado con una característica fenotípica, el sitio de restricción debe estar localizado cerca del gen responsable del fenotipo. La mutación que cambia la banda común a las personas sanas en la banda común a los pacientes, está muy íntimamente relacionada con el gen de la enfermedad.

relativamente cerca del gen en el mapa genético si ambos loci rara vez se recombinan, o nunca. Lo que se considera "relativamente cerca" en genética puede ser una distancia sustancial entre pares de bases de DNA; no obstante, el marcador de restricción proporciona un punto de inicio del cual partir en el DNA para localizar el gen.

La frecuente presencia de SNP en el genoma humano los hace útiles para el mapeo genético. De los 1.5×10^6 SNP que hasta ahora han sido identificados, en promedio hay uno cada 1 a 2 kb, lo cual debe permitir la localización rápida de genes de enfermedades nuevas al ubicarlos entre los SNP más cercanos.

Según el mismo principio, el mapeo por RFLP ha sido utilizado durante algún tiempo. Una vez que se asigna uno de ellos a un grupo de ligamiento, puede colocarse en el mapa genético. En el hombre y en el ratón, este mapeo ha conducido a la construcción de mapas de ligamiento para ambos genomas. El ligamiento de un sitio desconocido se prueba con estos sitios y puede ser colocado rápidamente en el mapa. Hay menos RFLP que SNP, así que, en principio, la resolución del mapa de RFLP es más limitada.

La frecuencia del polimorfismo significa que cada individuo tiene una constelación única de SNP o de RFLP. La combinación específica de los sitios encontrados en una región específica se denomina **haplotipo**, un genotipo en miniatura. Si bien el concepto de haplotipo fue introducido originalmente

te para describir la constitución genética del locus mayor de histocompatibilidad, región que especifica proteínas considerablemente importantes del sistema inmunológico (véase Cap. 23, Diversidad inmunitaria), ahora se ha extendido a la descripción de combinaciones particulares de alelos o sitios de restricción (o cualquier otro marcador genético) de algún área definida del genoma. Mediante los SNP se creó un detallado mapa de haplotipos del genoma humano que facilita el trazado de los mapas de los genes provocadores de enfermedades.

La existencia de RFLP constituye la base de una técnica que permite establecer relaciones inequívocas entre los progenitores y su prole. Cuando se duda de la paternidad, la comparación del mapa RFLP de una región cromosómica adecuada de los progenitores potenciales y del hijo permite asignar de manera absoluta el parentesco. El uso del análisis de restricción del DNA para la identificación de individuos se ha denominado **huella digital del DNA**. El análisis de secuencias "minisatélite" especialmente variables se utiliza en el mapeo del genoma humano (véase la sección 6.14, Los minisatélites facilitan el mapeo genético).

4.5 ¿Por qué los genomas son tan grandes?

Conceptos principales

- No hay una buena correlación entre el tamaño del genoma y la complejidad genética.
- Hay un incremento en el tamaño mínimo del genoma necesario para formar organismos de mayor complejidad.
- En numerosos filos, las dimensiones de los genomas de los organismos varían ampliamente.

La cantidad total de DNA del genoma (haploide) es una característica de cada especie viviente conocida como su **valor C**. Hay una variación enorme en el rango de los valores C, desde $<10^6$ bp en un micoplasma hasta $>10^{11}$ bp en algunas plantas y algunos anfibios.

En la **FIGURA 4.5** se resume el rango de valores C encontrados en diferentes filos evolutivos. Conforme se incrementa la complejidad, aumenta el tamaño mínimo del genoma encontrado en cada grupo. Sin embargo, al incrementarse las cantidades absolutas de DNA en las eucariotas superiores, se observan amplias variaciones en las dimensiones de los genomas de algunos filos.

En la **FIGURA 4.6**, el diagrama de la cantidad mínima de DNA requerida por un miembro de cada

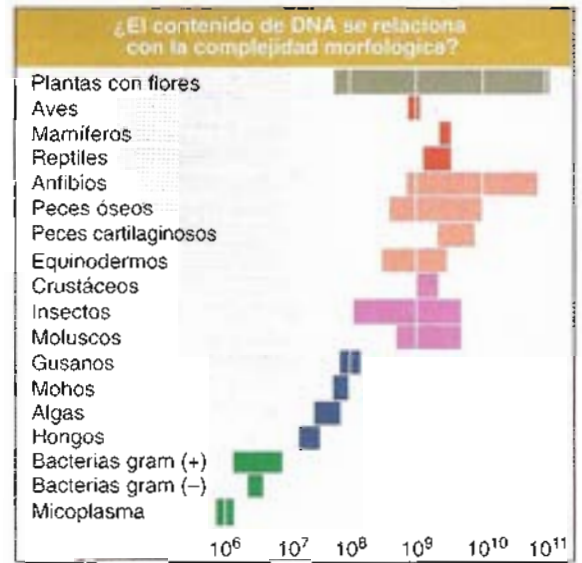


FIGURA 4.5 El contenido de DNA del genoma haploide se incrementa con la complejidad morfológica de las eucariotas inferiores, pero varía ampliamente en algunos grupos de eucariotas superiores. El rango de los valores de DNA de cada grupo se indica mediante áreas sombreadas.

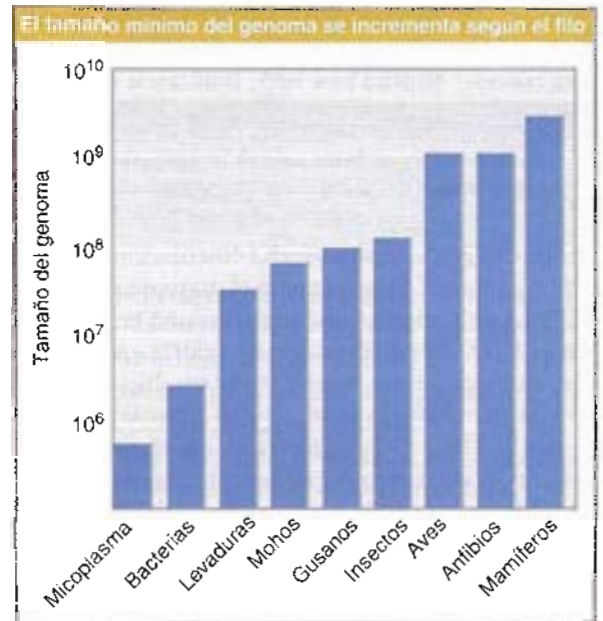


FIGURA 4.6 El genoma mínimo encontrado en cada filo se incrementa de las procariotas a los mamíferos.

grupo sugiere que para formar procariotas más complejas y eucariotas inferiores, es necesario que el genoma aumente de tamaño.

Los micoplasmas son las procariotas más pequeñas, y sus genomas son sólo ~3 veces el tamaño de un bacteriófago grande. El tamaño de las bacterias parte de $\sim 2 \times 10^6$ bp. Las eucariotas unicelulares (cuyos estilos de vida pueden parecerse a los de las

Tamaños de genomas útiles		
Filo	Especies	Genoma (bp)
Alga	<i>Pyrenomas salina</i>	6.6×10^5
Micoplasma	<i>M. pneumoniae</i>	1.0×10^6
Bacteria	<i>E. coli</i>	4.2×10^6
Levadura	<i>S. cerevisiae</i>	1.3×10^7
Moho limoso	<i>D. discoideum</i>	5.4×10^7
Nematodo	<i>C. elegans</i>	8.0×10^7
Insecto	<i>D. melanogaster</i>	1.4×10^8
Ave	<i>G. domesticus</i>	1.2×10^9
Anfibio	<i>X. laevis</i>	3.1×10^9
Mamífero	<i>H. sapiens</i>	3.3×10^9

FIGURA 4.7 Dimensiones de los genomas de algunos organismos experimentales comunes.

procariotas) también subsisten con genomas pequeños, aunque son más grandes que los de las bacterias. El hecho de que un organismo sea eucariótico en sí, no implica un gran incremento en cuanto al tamaño del genoma; una levadura puede tener un genoma de $\sim 1.3 \times 10^7$ bp, el cual es apenas dos veces el tamaño de un genoma bacteriano promedio.

Otro doble incremento en el tamaño del genoma es adecuado para respaldar al moho limoso *Dictyostelium discoideum*, que puede vivir tanto en la modalidad unicelular como en la multicelular. Es necesario otro incremento en la complejidad para dar lugar a los primeros organismos completamente multicelulares; el nematodo *Caenorhabditis elegans* tiene un DNA de 8×10^7 bp.

Asimismo, en el listado de la **FIGURA 4.7** se observa el incremento constante del tamaño del genoma de acuerdo con la complejidad de algunos de los organismos más comúnmente analizados. Es necesario que aumente el tamaño del genoma para que se formen insectos, aves, anfibios y mamíferos. Sin embargo, después de este punto no hay una buena relación entre las dimensiones del genoma y la complejidad morfológica del organismo.

Se sabe que los genes son mucho más grandes que las secuencias necesarias para codificar proteínas porque los exones (regiones codificadoras) pueden incluir sólo una pequeña parte de la longitud total de un gen, lo cual explica la razón de que haya mucho más DNA del necesario para proporcionar marcos de lectura para todas las proteínas del organismo, pues grandes segmentos de un gen interrumpido pueden no estar relacionados con la codificación de proteínas. Adicionalmente puede haber fragmentos significativos de DNA entre los genes, de modo que no es posible hacer deducciones respecto del número de genes a partir del tamaño total del genoma.

La **paradoja del valor C** se refiere a la falta de correlación entre el tamaño del genoma y la complejidad genética. Hay algunas variaciones muy curiosas

respecto del tamaño del genoma. En el sapo *Xenopus* y el hombre, los genomas son esencialmente del mismo tamaño, sin embargo, ¿se supone que el hombre es más complejo en cuanto a desarrollo genético! En algunos filos, las variaciones en contenido de DNA entre organismos que no difieren mucho en cuanto a complejidad extremadamente grandes (véase la Fig. 4.5). (Esto es especialmente notable en insectos, anfibios y plantas, no así en aves, reptiles ni mamíferos, los cuales muestran pocas variaciones dentro del grupo, con un rango de dimensiones de genoma de ~ 2 veces). Un grillo tiene un genoma 11 veces mayor que el de una mosca de la fruta. En los anfibios, los genomas más pequeños son de $<10^9$ bp, mientras que los más grandes son de $\sim 10^{11}$ bp. Es poco probable que se necesite una gran diferencia en el número de genes para especificar a estos anfibios. No se sabe por qué la selección natural permite esta variación, ni si tiene consecuencias evolutivas.

4.6 Los genomas eucarióticos contienen secuencias de DNA repetitivas y no repetitivas

Conceptos principales

- La cinética de la reasociación del DNA después de que un genoma ha sido desnaturalizado distingue a las secuencias por la frecuencia con se repiten en el genoma.
- En el DNA no repetitivo, los genes generalmente son codificados por secuencias localizadas.
- Los genomas más grandes de un filo no contienen más genes, pero tienen cantidades mayores de DNA repetitivo.
- Gran parte del DNA repetitivo puede estar formada por transposones.

Las características generales del genoma eucariótico pueden evaluarse por la cinética de reasociación del DNA desnaturalizado, técnica ampliamente utilizada antes de que fuera posible la secuenciación del DNA a gran escala.

Con la cinética de reasociación se identificaron dos tipos generales de secuencias genómicas:

- El **DNA no repetitivo** consta de secuencias únicas: sólo hay una copia en un genoma haploide.
- El **DNA repetitivo** consta de secuencias presentes en más de una copia en cada genoma.

El DNA repetitivo con frecuencia se divide en dos tipos generales:

- El DNA moderadamente repetitivo consiste en secuencias relativamente cortas que se repiten típicamente de 10-1 000 veces en el genoma, dispersas a lo largo de éste; son

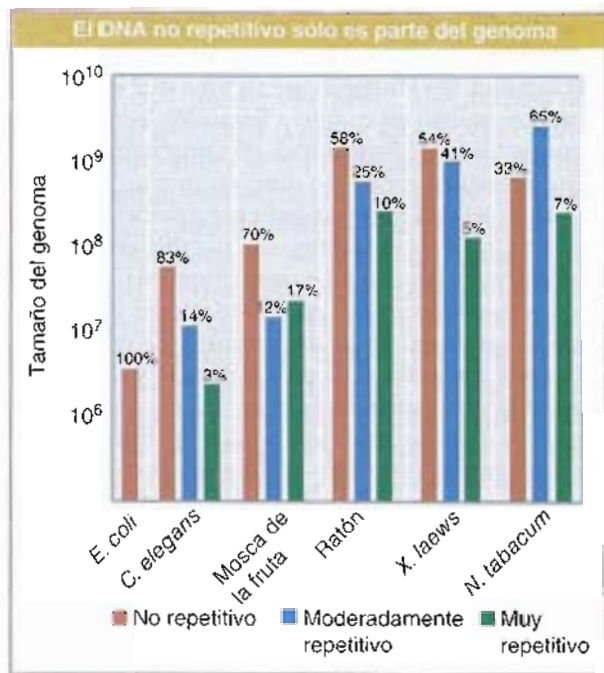


FIGURA 4.3 Las proporciones de los diferentes componentes secuenciales varían en los genomas eucarióticos. El contenido absoluto de DNA no repetitivo se incrementa con el tamaño del genoma, pero alcanza una meseta a $\sim 2 \times 10^9$ bp.

responsables del alto grado de formación de la estructura secundaria en el pre-RNAm, cuando repeticiones (invertidas) en los intrones se aparean para formar regiones dúplex.

- El DNA altamente repetitivo está formado por secuencias muy cortas (típicamente <100 bp) presentes muchos miles de veces en el genoma, frecuentemente organizadas en forma de repeticiones en tándem (véase la sección 6.11, Los DNA satélite frecuentemente se encuentran en la heterocromatina). Ninguna clase representa a proteínas.

La proporción del genoma ocupado por DNA no repetitivo varía mucho. En la **FIGURA 4.3** se resume la organización genómica de algunos organismos representativos. Las procariontas contienen sólo DNA no repetitivo. En las eucariotas inferiores, la mayor parte del DNA es no repetitivo; $<20\%$ forma parte de uno o más componentes moderadamente repetitivos. En las células animales, a menudo hasta la mitad del DNA es ocupado por componentes moderada y altamente repetitivos. En las plantas y en los anfibios, los componentes moderada y altamente repetitivos pueden representar hasta el 80% del genoma, de tal manera que el DNA no repetitivo se reduce a un componente minoritario.

Una parte significativa del DNA moderadamente repetitivo consiste en **transposones**, secuencias cortas de DNA (de ~ 1 kb) susceptibles de moverse a localizaciones nuevas en el genoma y de crear copias adicionales de sí mismos (véanse Cap. 21, Transposones, y Cap. 22, Retrovirus y retroposones). En algunos genomas eucarióticos superiores pueden ocupar incluso más de la mitad del genoma (véase la sección 5.5, El genoma humano tiene menos genes de los esperados).

En ocasiones se considera que los transposones se adaptan al concepto de **DNA redundante**, el cual se define como secuencias que se propagan por sí mismas en un genoma sin contribuir al desarrollo del organismo. Los transposones suelen fomentar las reestructuraciones genómicas, y éstas podrían conferir ventajas selectivas. Sin embargo, es justo decir que en realidad no se sabe por qué las fuerzas selectivas no contrarrestan el que los transposones se conviertan en una proporción tan grande del genoma. Otro término que se utiliza para describir el exceso aparente de DNA es "*DNA basura*", que define a secuencias genómicas sin una función aparente. Obviamente, es probable que en el genoma haya un equilibrio entre la generación de secuencias nuevas y la eliminación de secuencias indeseables, y que cierta proporción del DNA que aparentemente carece de función, esté en proceso de ser eliminada.

La longitud del componente no repetitivo de DNA tiende a incrementarse con el tamaño total del genoma conforme se procede a un tamaño total del genoma de $\sim 3 \times 10^9$ (característico de los mamíferos). Sin embargo, incrementos posteriores generalmente reflejan un aumento de la cantidad y la proporción de componentes repetitivos, de tal forma que es raro que un organismo tenga un componente de DNA no repetitivo $>2 \times 10^9$. Por lo tanto, el contenido de DNA no repetitivo de los genomas concuerda con nuestra percepción de la complejidad relativa del organismo. *E. coli* tiene 4.2×10^6 bp; *C. elegans* se incrementa en un orden de magnitud, a 6.6×10^7 bp; *D. melanogaster* se incrementa aún más, a $\sim 10^8$ bp, y en los mamíferos, el incremento es de otro orden de magnitud, a $\sim 2 \times 10^9$ bp.

¿Qué tipo de DNA corresponde a los genes codificadores de proteínas? La cinética de reasociación suele mostrar que el RNAm es derivado del DNA no repetitivo, de modo que la cantidad de éste es un mejor indicador del potencial codificador que el DNA total. (No obstante, un análisis más detallado basado en las secuencias genómicas muestra que numerosos exones tienen secuencias relacionadas en otros exones [véase la sección 3.5, Las secuencias de los exones se conservan, pero las de los intrones varían]. Dichos exones evolucionan por duplicación

para formar copias que inicialmente son idénticas, pero que después difieren en secuencia durante la evolución.)

4.7 Los genes pueden ser aislados por la conservación de los exones

Concepto principal

- La conservación de los exones puede ser utilizada como base para localizar regiones codificadoras al identificar fragmentos cuyas secuencias están presentes en múltiples organismos.

Algunas estrategias importantes para la identificación de genes se basan en el contraste entre la conservación de los exones y la variación de los intrones. En una región que contiene un gen cuya función se ha conservado en un rango de especies, la secuencia que representa a la proteína debe tener dos propiedades distintivas:

- Un marco de lectura abierto,
- y es probable que tenga una secuencia relacionada en otras especies.

Estas características pueden utilizarse para aislar genes.

Supóngase que por datos genéticos se sabe que un rasgo genético específico se localiza en determinada región cromosómica. Si se desconoce la naturaleza del producto del gen, ¿cómo se identifica al gen en una región que puede ser, por ejemplo, >1 Mb?

Una estrategia colosal cuya eficacia ha sido demostrada con algunos genes de importancia médica, consiste en detectar en fragmentos relativamente cortos de la región las dos propiedades esperadas de un gen conservado. Primero se intenta identificar fragmentos sometidos a hibridación cruzada con los genomas de otras especies y después se examinan dichos fragmentos para detectar marcos de lectura abiertos.

El primer criterio se aplica al realizar una **zootransferencia**, para lo cual se utilizan fragmentos cortos de la región como sondas (radioactivas) para evaluar la presencia de DNA relacionado en una variedad de especies a través de la hibridación de Southern. Si se encuentran fragmentos de hibridación relacionados con el de la sonda en numerosas especies —la sonda suele ser humana—, ésta se convierte en candidato para un exón del gen.

Los candidatos van en secuencia y, si contienen marcos de lectura abiertos, se utilizan para aislar a las regiones genómicas circundantes. Si parecen ser parte de un exón, posteriormente podrán ser utilizados para identificar al gen completo, aislar al

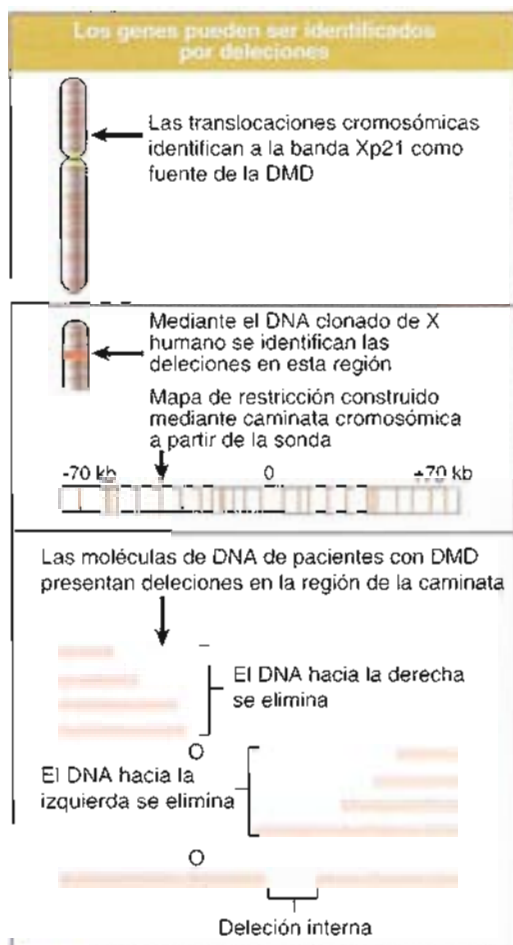


FIGURA 4.9 El gen implicado en la distrofia muscular de Duchenne fue detectado mediante un mapeo cromosómico y “caminando” hacia una región en la cual se pueden identificar deleciones cuando se presenta la enfermedad.

DNAc o al RNAm correspondiente y, por último, para identificar a la proteína.

Esta estrategia es particularmente importante cuando el gen diana está disperso porque posee numerosos intrones prominentes, como en el caso de la distrofia muscular de Duchenne (DMD, *Duchenne muscular dystrophy*), trastorno degenerativo de los músculos ligado al cromosoma X, presente en uno de cada 3 500 nacimientos de varones. Los pasos para identificar el gen se resumen en la **FIGURA 4.9**.

Mediante el análisis del ligamiento se localizó el locus correspondiente en la banda cromosómica Xp21. Los afectados por la enfermedad con frecuencia presentan reestructuraciones cromosómicas en dicha banda. Al comparar la capacidad de las sondas de DNA ligadas al cromosoma X para hibridarse con el DNA de pacientes con DNA normal, se obtuvieron fragmentos clonados correspondientes a la región reestructurada o eliminada del DNA de los pacientes.

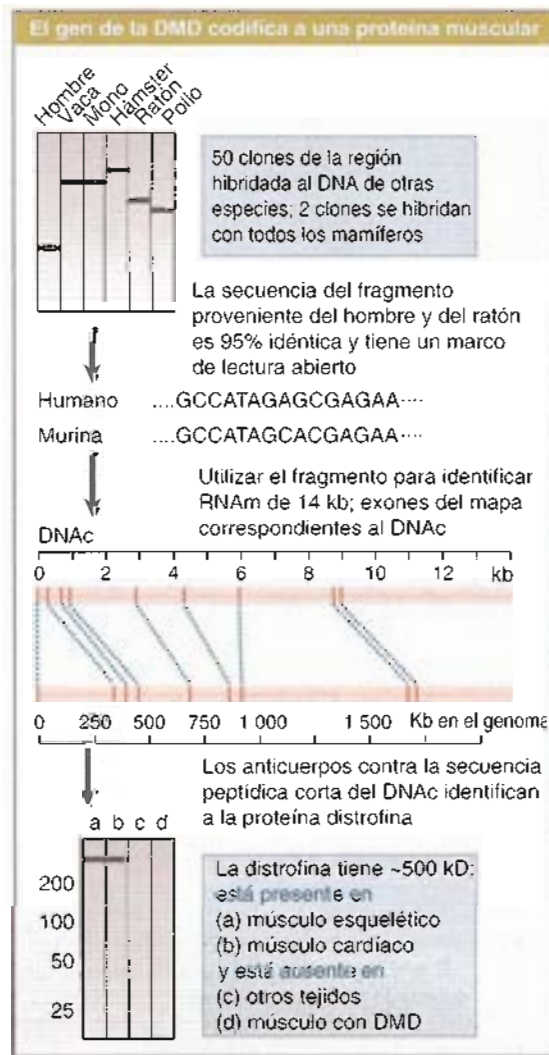


FIGURA 4.10 El gen de la distrofia muscular de Duchenne se caracterizó por medio de hibridación inespecífica, hibridación de DNAC, hibridación genómica e identificación de la proteína.

Una vez que se ha obtenido un poco del DNA localizado en la vecindad general del gen diana, es posible "caminar" a lo largo del cromosoma hasta llegar al gen. Mediante una caminata cromosómica se construyó un mapa de restricción de las regiones que flanquean a la sonda, el cual cubrió una región de >100 kb. Mediante el análisis del DNA de una serie de pacientes, en esta región se identificaron deleciones extensas en ambas direcciones; la más contundente está totalmente dentro de la región porque delinea un segmento importante para la función del gen e indica que éste, o cuando menos parte de él, yace en dicha región.

Una vez en la región del gen, se deben identificar los exones y los intrones. Mediante un análisis de hibridación inespecífica se identificaron fragmentos que presentan hibridación cruzada con el cromosoma X del ratón y con otras moléculas de DNA de

mamífero. Como se resume en la **FIGURA 4.10**, éstos se examinaron detalladamente para detectar marcos de lectura abiertos y las secuencias típicas de las uniones exón-intrón. Los fragmentos que cumplieran con estos criterios se utilizaron como sondas para identificar secuencias homólogas en una biblioteca de DNAC preparada a partir de RNAm muscular.

El DNAC correspondiente al gen identifica a un RNAm inusualmente largo, de aproximadamente 14 kb. La hibridación revertida hacia el genoma muestra que el RNAm está representado en >60 exones esparcidos en ~2 000 kb de DNA, lo cual hace del gen de la DMD el más largo identificado hasta ahora.

El gen codifica a una proteína de ~500 kD denominada distrofina, que es uno de los componentes del músculo, más bien escaso. Todos los afectados por la enfermedad presentan deleciones en este locus y carecen de distrofina (o es deficiente).

El músculo también se distingue por tener la proteína más grande conocida, la titina, formada por aproximadamente 27 000 aminoácidos. El gen correspondiente contiene el número mayor de exones (178) y el exón más largo del genoma humano (17 000 bp).

Otra técnica que permite examinar rápidamente los fragmentos genómicos para detectar los exones se denomina técnica de **captura de exones**. En la **FIGURA 4.11** se muestra que empieza con un vector que contiene un promotor fuerte y un solo intrón localizado entre los dos exones. Cuando este vector es introducido a las células por transfección, la transcripción del mismo genera grandes cantidades de RNA que contiene las secuencias de los dos exones. Dentro del intrón hay un sitio de restricción-clonación que se utiliza para insertar fragmentos genómicos de una región de interés. Si un fragmento no tiene exón, no hay cambios en el patrón de corte y empalme, y el RNA contendrá sólo las mismas secuencias que el vector progenitor. No obstante, si el fragmento genómico contiene un exón flanqueado por dos secuencias intrónicas parciales, se reconocen los sitios de corte y empalme a uno y otro lado de este exón y su secuencia es insertada en el RNA, entre los dos exones del vector. No es difícil detectar este proceso mediante transcripción inversa del RNA citoplásmico a DNAC utilizando la reacción en cadena de polimerasa (PCR, *polymerase chain reaction*) para amplificar las secuencias localizadas entre los dos exones del vector. Por ello, la aparición de secuencias del fragmento genómico en la población amplificada indica que un exón ha sido capturado. Como en las células animales los intrones suelen ser grandes y los exones pequeños, hay muchas probabilidades de que un fragmento

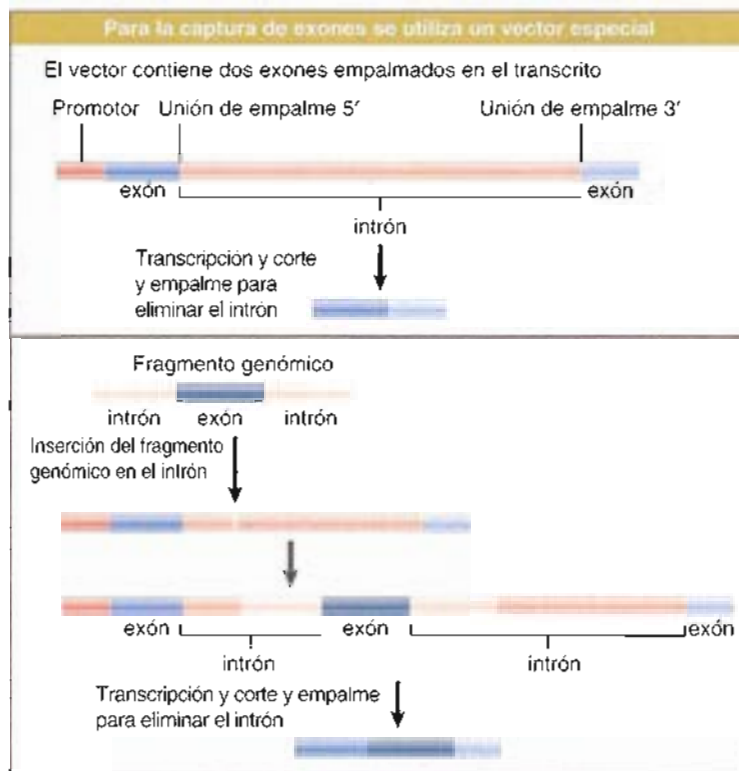


FIGURA 4.11 Para la captura de exones se utiliza un vector especial de corte y empalme. Si un exón está presente en el fragmento genómico, su secuencia será recuperada en el RNA citoplásmico, pero si el fragmento genómico está formado únicamente por secuencias de dentro de un intrón, no ocurre el corte y empalme y el RNA no es exportado al citoplasma.

aleatorio de DNA genómico contenga la estructura requerida de un exón rodeado de intrones parciales. De hecho, la captura de exones puede parecerse a los eventos ocurridos de forma natural durante la evolución de los genes (véase la sección 3.8, ¿Cómo evolucionaron los genes interrumpidos?)

4.8 La conservación de la organización del genoma ayuda a identificar genes

Conceptos principales

- Los algoritmos para identificar genes no son perfectos, de modo que son necesarias numerosas correcciones al conjunto inicial de datos.
- Los pseudogenes deben distinguirse de los genes activos.
- Hay amplias relaciones sinténicas entre los genomas del ratón y los del humano, y la mayoría de los genes activos se encuentran en una región sinténica.

Una vez ensamblada la secuencia de un genoma, todavía se tienen que identificar los genes que con-

tiene. Las secuencias codificadoras representan una fracción muy pequeña. Los exones pueden ser identificados como marcos de lectura abiertos e ininterrumpidos, flanqueados por secuencias apropiadas. ¿Cuáles son los criterios para identificar un gen activo en una serie de exones?

En la **FIGURA 4.12** se muestra que un gen activo debe estar formado por una serie de exones en la cual el primero de ellos sigue inmediatamente a un promotor; los exones internos son flanqueados por uniones apropiadas de corte y empalme, y el último va seguido de señales de procesamiento 3'; uniendo a los exones puede deducirse un solo marco de lectura abierto que empieza con un codón de iniciación y termina con uno de terminación. Los exones internos pueden identificarse como marcos de lectura abiertos flanqueados por uniones de corte y empalme. En los casos más sencillos, el primero y el último exón inician y terminan la región codificadora, respectivamente (así como las regiones no traducidas 5' y 3'). En casos más complejos, dichos exones pueden tener sólo regiones no traducidas y, por lo tanto, ser más difíciles de identificar.

Los algoritmos utilizados para conectar los exones no son realmente efectivos cuando el genoma es

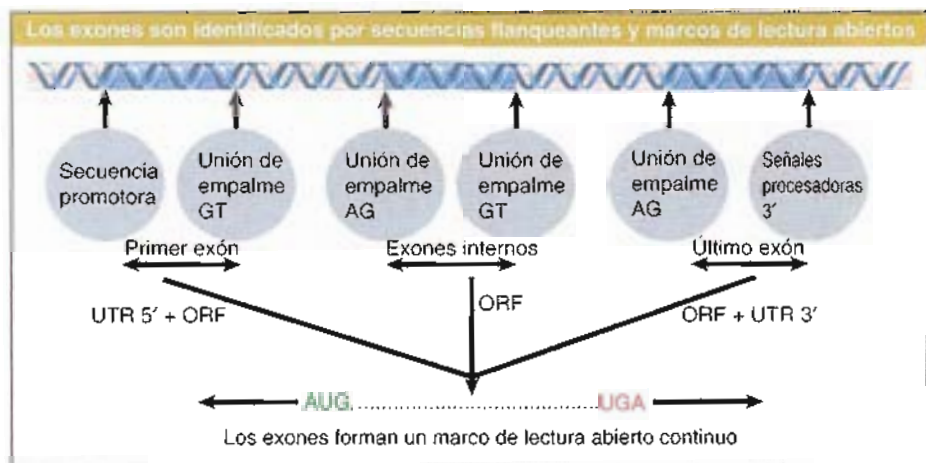


FIGURA 4.12 Los exones de los genes codificadores de proteínas son identificados como secuencias codificadoras flanqueadas por señales apropiadas (con regiones no traducidas en ambos extremos). La serie de exones debe generar un marco de lectura abierto con codones adecuados de inicio y final. ORF, marco de lectura abierto; UTR, regiones no traducidas; GT y AG se refieren a las bases nitrogenadas de las secuencias.

muy grande y los exones están muy separados. Por ejemplo, en el análisis inicial del genoma humano se mapearon 170 000 exones en 32 000 genes. No es difícil que estas cifras sean incorrectas porque resultan en un promedio de 5.3 exones por gen, mientras que la media de los genes caracterizados completamente es de 10.2, de modo que, o se han omitido muchos exones o deben conectarse de forma diferente en un número más pequeño de genes en toda la secuencia genómica.

Aun cuando la organización de un gen sea identificada correctamente, surge el problema de distinguir entre genes activos y pseudogenes. Muchos de estos últimos se reconocen por defectos obvios que dan lugar a mutaciones múltiples que resultan en una secuencia codificadora inactiva. Los pseudogenes que han surgido más recientemente no acumulan tantas mutaciones, de modo que pueden ser más difíciles de reconocer. En un ejemplo extremo, el ratón tiene sólo un gen *Gapdh* activo (que codifica a la deshidrogenasa de gliceraldehído fosfato), pero ~400 pseudogenes. De éstos, en un principio pareció que aproximadamente 100 estaban activos en la secuencia del genoma del ratón y fue necesario examinarlos uno por uno para excluirllos de la lista de genes activos.

La certeza de que un gen está activo se puede incrementar comparando regiones de los genomas de especies diferentes. Ha habido una reorganización general muy amplia de las secuencias entre los genomas de ratón y de humano, como se observa en el simple hecho de que el genoma haploide del humano tiene 23 cromosomas y el del ratón, 20. No obstante, localmente, el orden de los genes suele ser

el mismo: cuando se comparan pares de homólogos de humano y de ratón, los genes localizados a ambos lados también tienden a ser homólogos. A esta relación se le denomina **sintenia**.

En la **FIGURA 4.13** se muestra la relación entre el cromosoma 1 del ratón y el conjunto cromosómico humano; se observa que 21 segmentos de este cromosoma de ratón tienen contrapartes sinténicas en los cromosomas humanos. La extensión de la reorganización ocurrida entre los genomas se demuestra porque los segmentos están esparcidos en seis cromosomas humanos diferentes. El mismo tipo de relación se ha detectado en todos los cromosomas del ratón, excepto en el X, sinténico sólo con el cromosoma humano X, lo cual se explica por el hecho de que el cromosoma X constituye un caso especial, sujeto a compensación de dosis para ajustar las diferencias entre machos (una copia) y hembras (dos copias) (véase la sección 31.5, Los cromosomas X experimentan cambios globales). Esto puede ejercer presión selectiva contra la translocación de genes desde y hacia el cromosoma X.

La comparación de las secuencias de los genomas del ratón y del humano revela que >90% de cada genoma se encuentra en bloques sinténicos cuyo tamaño varía (de 300 kb a 65 Mb). Hay un total de 342 segmentos sinténicos, con una longitud promedio de 7 Mb (0.3% del genoma). El 99% por ciento de los genes del ratón tiene un homólogo en el genoma humano; para el 96% de dichos genes, ese homólogo se encuentra en una región sinténica.

La comparación de los genomas proporciona información interesante con respecto de la evolución de las especies. El número de familias de



FIGURA 4.13 El cromosoma 1 del ratón tiene 21 segmentos de 1 a 25 Mb que son sinténicos con regiones que corresponden a partes de seis cromosomas humanos.

genes de los genomas del ratón y del humano es el mismo, y una diferencia importante entre las especies es la expansión diferencial de familias específicas en uno de los genomas, lo cual es particularmente notable en los genes que afectan a características fenotípicas únicas de las especies. En el ratón, de las 25 familias en las cuales se ha observado aumento de tamaño, 14 contienen genes específicamente implicados en la reproducción de los roedores y cinco, genes específicos del sistema inmunológico.

Una validación de la importancia de los bloques sinténicos proviene de comparaciones de pares de los genes contenidos en ellos. Al buscar pseudogenes similares sobre la base de comparaciones de secuencias, para un gen que no está en una localización sinténica (esto es, su contexto es diferente en las dos especies) hay el doble de probabilidades de que sea un pseudogén. En otras palabras, la translocación lejos del locus original tiende a asociarse con la creación de pseudogenes, por lo tanto, la falta de un gen relacionado en una posición sinténica despierta sospechas de que un supuesto gen puede realmente ser un pseudogén. En total, >10% de los genes identificados inicialmente en el análisis del genoma probablemente resulten pseudogenes.

Como regla general, las comparaciones entre genomas incrementan significativamente la certeza del pronóstico génico. Cuando se conservan características de la secuencia que conducen a genes activos, por ejemplo, entre el hombre y el ratón, se incrementan las probabilidades de que identifiquen a homólogos activos.

La identificación de genes que codifican RNA es más complicada porque no es posible utilizar el criterio del marco de lectura abierto. También es verdad que con el análisis comparativo del genoma se incrementó el rigor del análisis. Por ejemplo, al analizar el genoma del humano o del ratón se identifican ~500 genes que codifican RNA, pero la comparación de características sugiere que <350 de estos genes están realmente activos en cada genoma.

4.9 Los organelos contienen DNA

Conceptos principales

- Las mitocondrias y los cloroplastos tienen genomas que muestran herencia no mendeliana. Típicamente son heredados de la madre.
- Los genomas de los organelos pueden ser sometidos a segregación somática en las plantas.
- Las comparaciones del DNA mitocondrial sugieren que los humanos descienden de una sola hembra, la cual vivió hace 200 000 años en África.

La primera evidencia de genes fuera del núcleo provino de la herencia no mendeliana detectada en plantas (en los primeros años del siglo xx, después del redescubrimiento de la herencia mendeliana). En ocasiones, dicha herencia está relacionada con el fenómeno de la segregación somática, en ambos casos la causa es similar:

- La herencia no mendeliana se define como la incapacidad de la progenie de un apareamiento de mostrar la segregación mendeliana de los caracteres de los progenitores. Esta limitante refleja que falta la asociación entre el carácter segregante y el huso meiótico.
- La segregación somática describe un fenómeno en el cual los caracteres de los progenitores se segregan en las células somáticas, de modo que exhiben la heterogeneidad del organismo. Ésta es una característica notable del desarrollo de las plantas porque refleja la falta de asociación entre el carácter segregante y el huso mitótico.

Por lo tanto, se considera que la herencia no mendeliana y la segregación somática indican la presencia de genes que residen fuera del núcleo y que no recurren a la segregación en el huso mitótico y el meiótico para distribuir réplicas a los gametos o a las células hijas, respectivamente. En la **FIGURA 4.14** se muestra que esto sucede cuando las mitocondrias heredadas de los progenitores macho y hembra tienen alelos diferentes, y por casualidad, una célula hija recibe una distribución desequilibrada de mitocondrias que representa sólo a un progenitor (véase la sección 17.12. ¿Cómo se replican y segregan las mitocondrias?).

La forma extrema de la herencia no mendeliana es la herencia uniparental, cuando se hereda el genotipo de sólo uno de los progenitores y el del otro progenitor se pierde permanentemente. En ejemplos menos extremos, la progenie de un genotipo progenitor excede a la del otro genotipo; por lo general, es el genotipo de la madre el que se hereda preferentemente (o únicamente). Este efecto se define en ocasiones como **herencia materna**.

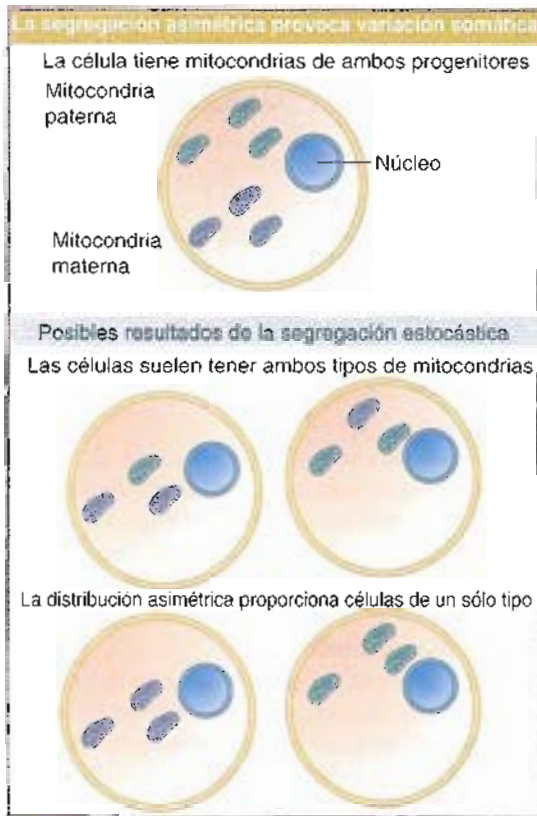


FIGURA 4.14 Cuando los alelos mitocondriales paternos y maternos difieren, una célula tiene dos conjuntos de DNA mitocondrial. La mitosis suele generar células hijas con ambos conjuntos. La variación somática resulta de una segregación desigual que genera células hijas con sólo un conjunto.

El punto importante es que predomina el genotipo proporcionado por el progenitor de determinado género, como se observa en los índices de segregación anormales cuando se realiza una cruce entre un mutante y un tipo silvestre, lo cual contrasta con el comportamiento de la genética mendeliana, que se presenta cuando cruces recíprocas muestran que las contribuciones de ambos padres son heredadas de forma equivalente.

El sesgo en los genomas progenitores se establece al formarse un cigoto, o poco después, por varias causas posibles. La aportación de información paterna o materna a los organelos del cigoto puede ser desigual; en el caso más extremo, sólo contribuye un progenitor. En otros casos, las aportaciones son equivalentes, pero la información provista por un progenitor no sobrevive. Es posible que se combinen ambos efectos, pero independientemente de la causa, la representación desigual de la información proveniente de los progenitores contrasta con la información genética nuclear, que se deriva de manera equivalente de uno y otro progenitor.

La herencia no mendeliana resulta de la presencia de genomas de DNA en mitocondrias y cloroplastos, heredados independientemente de los ge-

nes nucleares. En efecto, el genoma de los organelos es un fragmento de DNA físicamente secuestrado en una parte definida de la célula y sujeto a su propia forma de expresión y regulación. Un genoma de organelo puede codificar algunas o todas las moléculas de RNA, pero sólo algunas de las proteínas necesarias para perpetuar al organelo. Las otras proteínas son codificadas en el núcleo, expresadas a través del mecanismo citoplásmico de síntesis de proteínas e importadas al interior del organelo.

Los genes que no residen dentro del núcleo generalmente se describen como **genes extranucleares** que se transcriben y traducen en el mismo compartimiento del organelo (mitocondria o cloroplasto) en el cual residen. Por el contrario, los genes nucleares son expresados a través de la síntesis citoplásmica de proteínas. (El término **herencia citoplásmica** suele utilizarse para describir el comportamiento de los genes en los organelos, pero esta descripción no debe utilizarse porque es importante poder distinguir entre lo que sucede en el citosol general y lo que acontece en organelos específicos.)

Los animales superiores muestran herencia materna, que se explica si las mitocondrias provienen todas del óvulo y en absoluto del esperma. En la **FIGURA 4.15** se muestra que el esperma contribuye sólo con una copia de DNA nuclear, de modo que los genes mitocondriales son derivados exclusivamente de la madre, y en los machos, son desechados en cada generación.

Las condiciones del organelo son diferentes de las observadas en el núcleo, y por lo tanto, el DNA de los organelos evoluciona a su propia velocidad. Si la herencia es uniparental, puede no haber recombinación entre los genomas progenitores. De hecho, la recombinación suele no presentarse cuando los genomas de los organelos son heredados de ambos progenitores. El DNA de los organelos tiene un sistema de replicación diferente al del núcleo, por lo que la tasa de error durante la replicación puede ser diferente. El DNA mitocondrial acumula mutaciones más rápidamente que el DNA nuclear en los mamíferos, sin embargo, en las plantas, la acumulación en la mitocondria es más lenta que en el núcleo (en el cloroplasto la velocidad es intermedia).

Una consecuencia de la herencia materna es que la secuencia del DNA mitocondrial es más sensible a reducciones en el tamaño de la población de reproducción que la del DNA nuclear. La comparación de las secuencias de DNA mitocondrial en un rango de poblaciones humanas permite construir un árbol evolutivo. La divergencia entre las moléculas de DNA mitocondrial humano cubre el 0.57%. Es posible elaborar un árbol en el cual las variantes mitocondriales se derivaron de un ancestro (africano) común. La tasa a la cual el DNA mitocondrial

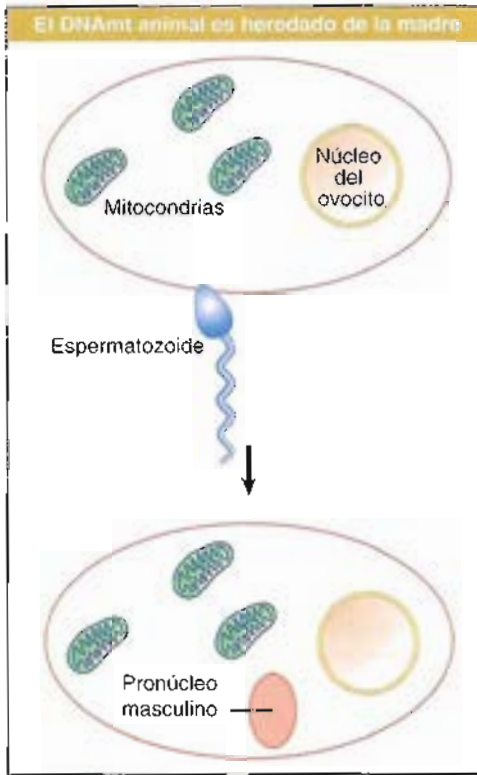


FIGURA 4.15 El DNA del esperma entra al ovocito para formar el pronúcleo masculino en el óvulo fertilizado, sin embargo, todas las mitocondrias son proporcionadas por el ovocito.

de los mamíferos acumula mutaciones es de 2 a 4% por cada millón de años, es decir, >10 veces más rápida que la tasa de la globina. Dicha tasa generará la divergencia observada en un periodo evolutivo de 140 000 a 280 000 años, lo cual implica que la raza humana desciende de una sola hembra que vivió en África hace ~200 000 años.

4.10 Los genomas de los organelos son moléculas circulares de DNA que codifican proteínas de los organelos

Conceptos principales

- Los genomas de los organelos suelen ser moléculas circulares de DNA (pero no siempre).
- Los genomas de los organelos codifican algunas de las proteínas que se encuentran en el organelo, pero no a todas.

La mayoría de los genomas de los organelos asumen la forma de una sola molécula circular de DNA de secuencia única (llamada **DNAm** en la mitocondria y **DNAct** en el cloroplasto). Hay pocas excep-

ciones en las que el DNA de la mitocondria es una molécula lineal; estas estructuras generalmente se presentan en las eucariotas inferiores.

En general, en un organelo hay numerosas copias del genoma, y en cada célula hay múltiples organelos, de modo que son numerosos los genomas de organelos por célula. Aunque el genoma de los organelos es único, constituye una secuencia repetitiva respecto de cualquier secuencia nuclear no repetitiva.

Los genomas de los cloroplastos son relativamente grandes, casi siempre ~140 kb en las plantas superiores y <200 kb en las eucariotas inferiores, dimensiones comparables con las de un bacteriófago grande, el T4, por ejemplo, mide ~165 kb. Son varias las copias del genoma por organelo, típicamente de 20 a 40 en una planta superior, y múltiples copias del organelo en cada célula, por lo general, de 20 a 40.

El tamaño total de los genomas mitocondriales varía en más de un orden de magnitud. Así, los genomas mitocondriales de las células animales son pequeños, aproximadamente de 16.5 kb en los mamíferos. Hay muchos cientos de mitocondrias por célula, y cada mitocondria tiene múltiples copias de DNA. La cantidad total de DNA mitocondrial respecto del DNA del núcleo es pequeña, se estima en <1%.

Por el contrario, en las levaduras, el genoma mitocondrial es mucho más grande, por ejemplo, en *Saccharomyces cerevisiae* el tamaño exacto varía de una cepa a otra, pero en promedio es de ~80 kb. Hay ~22 mitocondrias por célula, que corresponden a ~4 genomas por organelo. En células de cultivo, la proporción de DNA mitocondrial puede llegar al 18 por ciento.

En las plantas, el rango de dimensiones del DNA mitocondrial es extremadamente amplio, con un mínimo de ~100 kb. Las dimensiones del genoma hacen difícil aislarlo intacto, pero el mapeo de restricción de muchas plantas sugiere que el genoma mitocondrial suele ser una secuencia individual organizada en un círculo, dentro del cual hay secuencias homólogas cortas. La recombinación entre estos elementos genera moléculas circulares subgenómicas más pequeñas que coexisten con el genoma "maestro" completo, buen ejemplo de la complejidad aparente de las moléculas de DNA mitocondrial de las plantas.

Ahora que se cuenta con la secuencia de los genomas mitocondriales de numerosos organismos, es posible observar algunos patrones generales en la representación de las funciones del DNA mitocondrial. En la **FIGURA 4.16** se resume la distribución de los genes de ese tipo de genomas. El número total de genes codificadores de proteínas es bastante pequeño y no se correlaciona con el tamaño del geno-

Las mitocondrias codifican a moléculas de RNA y proteínas			
Especie	Tamaño (en kb)	Genes codificadores de proteínas	Genes codificadores de RNA
Hongos	19-100	8-14	10-28
Protistas	6-100	3-62	2-29
Plantas	186-366	27-34	21-30
Animales	16-17	13	4-24

FIGURA 4.16 Los genomas mitocondriales tienen genes (principalmente de los complejos I-IV) que codifican proteínas, moléculas de RNAr y moléculas de RNAt.

ma. Las mitocondrias de los mamíferos utilizan sus genomas de 16 kb para codificar a 13 proteínas, en tanto que las mitocondrias de las levaduras utilizan los de 60 a 80 kb para codificar a sólo ocho proteínas. Las plantas, cuyos genomas mitocondriales son mucho más grandes, codifican a más proteínas. En la mayoría de los genomas mitocondriales se encuentran intrones, aunque no en los genomas muy pequeños de mamífero.

Los dos RNAr principales siempre son codificados por el genoma mitocondrial. El número de moléculas de RNAt codificado por el genoma mitocondrial fluctúa entre ninguno y todo el complemento (25 a 26 en las mitocondrias), lo cual justifica la variación en la Figura 4.16.

La parte principal de la actividad codificadora de proteínas está dedicada a los componentes de los ensambles de múltiples subunidades de los complejos de respiración I-IV. Muchas proteínas ribosómicas son codificadas en los genomas protistas y mitocondriales de las plantas, sin embargo, muy pocas o ninguna se codifican en los genomas de hongos y animales. En muchos genomas mitocondriales protistas hay genes que codifican proteínas y que están implicados en la importación.

4.11 La organización del DNA mitocondrial es variable

Conceptos principales

- El DNA mitocondrial de las células animales es extremadamente compacto y suele codificar a 13 proteínas, dos moléculas de RNAr y 22 de RNAt.
- El DNA mitocondrial de las levaduras es cinco veces más largo que el DNAmt de las células animales por la presencia de intrones largos.

El DNA mitocondrial de los animales es extremadamente compacto. En diferentes filos animales se observan diferencias notables en la organización genética detallada, pero se conserva el principio general de un genoma pequeño que codifica a un número

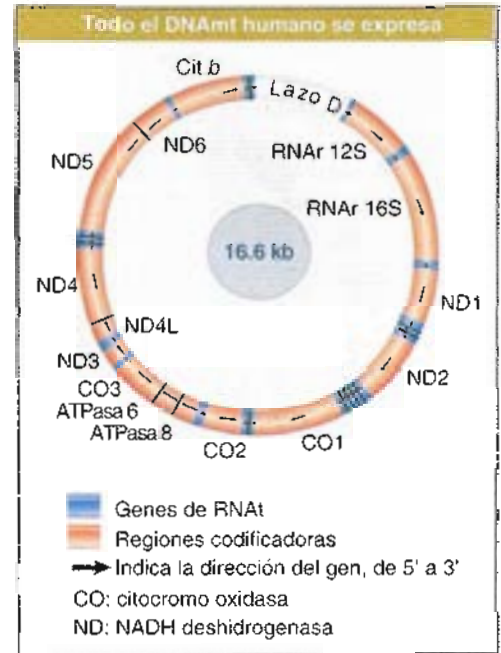


FIGURA 4.17 El DNA mitocondrial humano tiene 22 genes de RNAt y 13 regiones codificadoras de proteínas. Catorce de las 15 regiones codificadoras de proteínas o codificadoras de RNA se transcriben en la misma dirección. Catorce de los genes de RNAt se expresan en la dirección de las manecillas del reloj y ocho en dirección contraria.

restringido de funciones. En los genomas mitocondriales de los mamíferos, la organización es extremadamente compacta, no hay intrones, de hecho, algunos genes se superponen, y casi cada par de bases puede ser asignado a un gen. Con excepción del lazo D, región implicada en el inicio de la replicación del DNA, no puede considerarse que más de 87 de los 16 569 bp del genoma mitocondrial humano se encuentren en regiones intercistricas.

Las secuencias de nucleótidos completas de los genomas mitocondriales de las células animales muestran una homología extensa en cuanto a organización. El mapa del genoma mitocondrial humano se resume en la FIGURA 4.17. Hay 13 regiones codificadoras de proteínas; todas éstas forman parte del mecanismo implicado en la respiración, entre otras, el citocromo *b*, las tres subunidades de la citocromo oxidasa, una de las subunidades de la ATPasa y siete unidades (o proteínas asociadas) de la NADH deshidrogenasa.

La discrepancia en el tamaño de los genomas mitocondriales de *S. cerevisiae* (84 kb), cinco veces más grandes que los de los mamíferos (16 kb), alerta ante el hecho de que debe haber una gran diferencia en su organización genética, a pesar de su función común. El número de productos relacionados con las funciones enzimáticas mitocondriales y sintetizados de forma endógena parece ser similar. ¿El

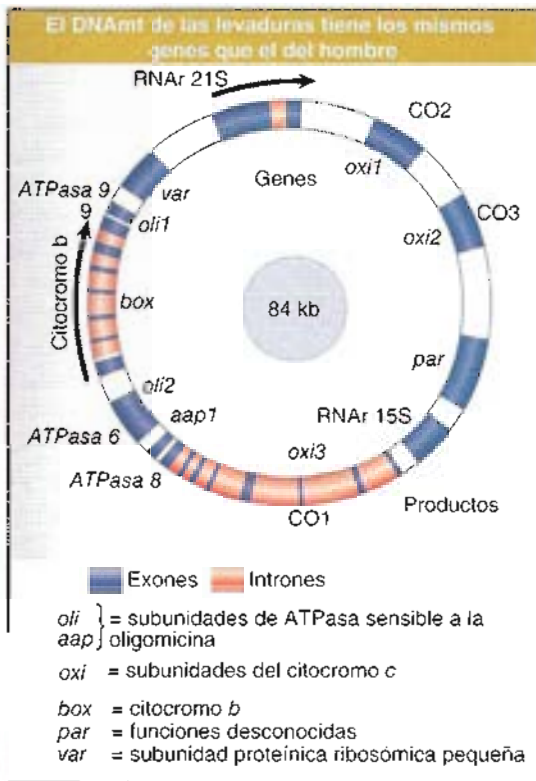


FIGURA 4.18 El genoma mitocondrial de *S. cerevisiae* contiene genes codificadores de proteínas, genes de RNAr y genes de RNAt interrumpidos y continuos (no se indican las posiciones). Las flechas indican la dirección de la transcripción.

material genético adicional de las levaduras mitocondriales representa a otras proteínas, quizá relacionadas con la regulación, o no se expresa?

El mapa de la **FIGURA 4.18** representa al RNA principal y los productos proteínicos de la mitocondria de la levadura, en el cual la característica más notable es la dispersión de los loci.

Los dos más prominentes son los genes interrumpidos *box* (que codifican al citocromo *b*) y *oxi3* (que codifica a la subunidad 1 de la citocromo oxidasa), que juntos son ¡casi tan largos como el genoma mitocondrial entero de los mamíferos! Muchos de los intrones largos de estos genes tienen marcos de lectura abiertos en registro con el exón precedente (véase la sección 27.5. Algunos intrones del grupo I codifican endonucleasas que promueven la movilidad), fenómeno que agrega numerosas proteínas, todas sintetizadas en cantidades bajas, al complemento de la mitocondria de la levadura.

Los genes restantes son ininterrumpidos y corresponden a otras dos subunidades de la citocromo oxidasa codificadas por la mitocondria, a la(s) subunidad(es) de la ATPasa y (en el caso del gen *var1*) a una proteína ribosómica mitocondrial. Es poco probable que el número total de genes mitocondriales de las levaduras exceda de ~25.

Los cloroplastos tienen >100 genes

Genes	Tipos
Codificadores de RNA	
RNAr 16S	1
RNAr 23S	1
RNAr 4,5S	1
RNAr 5S	1
RNAT	30-32
Expresión génica	
Proteínas r	20-21
RNA polimerasa	3
Otras	2
Funciones del cloroplasto	
Rubisco y tilacoides	31-32
NADH deshidrogenasa	11
Total	105-113

FIGURA 4.19 El genoma de los cloroplastos de las plantas terrestres codifica a cuatro moléculas de RNAr, 30 moléculas de RNAT y ~60 proteínas.

4.12 El genoma de los cloroplastos codifica a numerosas proteínas y moléculas de RNA

Concepto principal

- Los genomas de los cloroplastos varían en cuanto a tamaño, pero son lo suficientemente grandes como para codificar de 50 a 100 proteínas, así como a los RNAr y los RNAT.

¿Qué genes portan los cloroplastos? Las moléculas de DNA de los cloroplastos varían en longitud de 120 a 190 kb. Los genomas secuenciados de cloroplastos (>10 en total) tienen de 87 a 183 genes. En la **FIGURA 4.19** se resumen las funciones codificadas por el genoma de los cloroplastos de las plantas terrestres. Hay más variaciones en los genomas de los cloroplastos de las algas.

En general, la situación es similar a la de las mitocondrias, excepto que hay más genes implicados. El genoma del cloroplasto codifica a todas las especies de RNAr y de RNAT necesarias para la síntesis de proteínas. El ribosoma incluye dos moléculas de RNAr pequeñas, además de las especies principales. El conjunto de RNAT puede incluir todos los genes necesarios. El genoma de los cloroplastos codifica a ~50 proteínas, incluidas la RNA polimerasa y las proteínas ribosómicas. También en este caso, la regla es que los genes de los organelos son transcritos y traducidos por el mecanismo del organelo.

Cerca de la mitad de los genes de los cloroplastos codifican a proteínas implicadas en la síntesis de proteínas. El origen endosimbiótico de los cloroplastos resalta por las relaciones entre estos genes y sus

contrapartes en las bacterias. La organización de los genes de RNAr en particular está íntimamente relacionada con la de una cianobacteria, la cual define con mayor precisión al último ancestro común de cloroplastos y bacterias.

Intrones y cloroplastos caben en dos clases generales; los que se encuentran en los genes de RNAr suelen estar en el lazo anticodón (aunque no inevitablemente), como los intrones localizados en los genes de RNAr nucleares de las levaduras (véase la sección 26.14, El corte y empalme del RNAr de las levaduras implica procedimientos de corte y empalme). Los intrones localizados en los genes codificadores de proteínas se asemejan a los de los genes mitocondriales (véase Cap. 27, El RNA catalítico), por lo que el evento endosimbiótico se ubica en un periodo de la evolución anterior a la separación de las procariontas con genes ininterrumpidos.

La función de los cloroplastos es la fotosíntesis. Muchos de sus genes codifican a proteínas de los complejos localizados en las membranas de los tilacoides, cuya constitución muestra un balance diferente del de los complejos mitocondriales. Si bien algunos de estos complejos son similares a los mitocondriales debido a que algunas subunidades son codificadas por el genoma de los organelos y otras por el nuclear, otros complejos de cloroplastos son codificados completamente por un genoma.

4.13 Las mitocondrias evolucionaron por endosimbiosis

¿Cómo evolucionó una situación en la cual un organelo contiene información genética para algunas de sus funciones, mientras que otras son codificadas en el núcleo? En la FIGURA 4.20 se ilustra el modelo de la endosimbiosis de la evolución de las mitocondrias, durante la cual, células primitivas capturaron a bacterias que proporcionaron las funciones y las estructuras que evolucionaron hacia mitocondrias y cloroplastos. En ese momento, el proto-organelo debe haber contenido todos los genes necesarios para especificar sus funciones.

Las homologías secuenciales sugieren que mitocondrias y cloroplastos evolucionaron separadamente a partir de linajes comunes a las eubacterias, pero las mitocondrias comparten su origen con las bacterias púrpura α y los cloroplastos, con las cianobacterias. Entre las bacterias, el pariente de las mitocondrias más cercano que se conoce es *Rickettsia* (agente causal de tifus), parásito intracelular obligado que probablemente desciende de bacterias de vida libre, hipótesis que refuerza la idea de que las mitocondrias se originaron en un evento endosim-

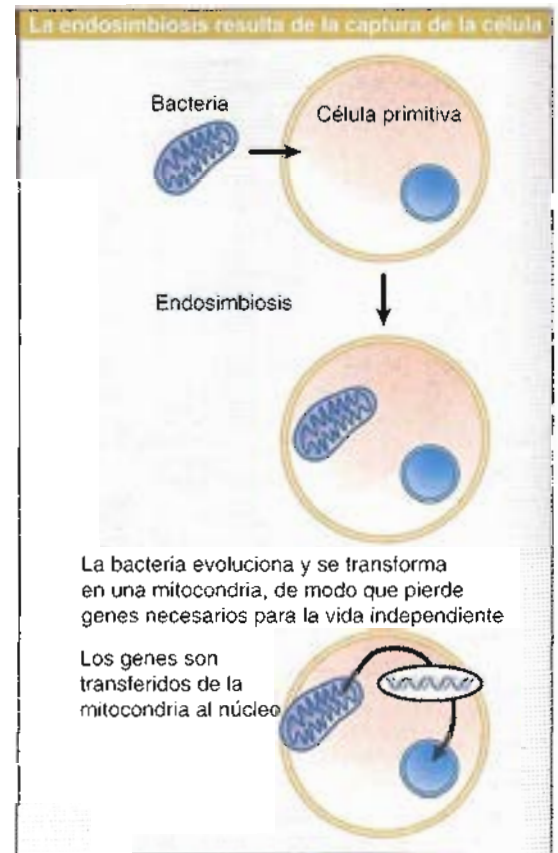


FIGURA 4.20 Las mitocondrias se originaron por un evento endosimbiótico en el cual una bacteria fue capturada por una célula eucariótica.

biótico en que también estuvo involucrado un ancestro común.

Deben haber ocurrido dos cambios, conforme las bacterias se integraron en la célula recipiente y evolucionaron hacia mitocondrias (o cloroplastos). Los organelos tienen mucho menos genes que una bacteria independiente y han perdido muchas de las funciones génicas necesarias para la vida independiente (como las rutas metabólicas). La mayoría de los genes que codifican funciones de los organelos se localizan de hecho en el núcleo, por lo tanto, estos genes deben haber sido transferidos ahí desde el organelo.

La transferencia de DNA entre organelo y núcleo ha ocurrido a lo largo de la evolución, y aún continúa. La velocidad de transferencia puede medirse directamente al introducir en un organelo un gen que sólo puede funcionar en el núcleo, porque, por ejemplo, contiene un intrón nuclear o la proteína debe funcionar en el citosol. Respecto de la provisión del material necesario para la evolución, la velocidad de transferencia del organelo al núcleo equivale aproximadamente a la de una mutación génica única. El DNA introducido en las mitocondrias

es transferido al núcleo a una tasa de 2×10^{-5} por generación. Los experimentos para medir la transferencia en dirección opuesta, del núcleo a la mitocondria, sugieren que la tasa es mucho más baja, $<10^{-10}$. Cuando se introduce un gen nuclear específico de resistencia a antibióticos en un cloroplasto, la transferencia al núcleo y el éxito de la expresión suelen vigilarse mediante el análisis de plántulas, de modo de evaluar la resistencia al antibiótico; este proceso muestra que la transferencia ocurre a una velocidad de 1 en 16 000 plántulas, o 6×10^{-5} .

Obviamente, la transferencia de un gen de un organelo al núcleo requiere de movimiento físico del DNA, pero la expresión exitosa también implica cambios en la secuencia codificadora. Las proteínas de organelos codificadas por genes nucleares tienen secuencias especiales que les permiten ser importadas al interior del organelo después de haber sido sintetizadas en el citoplasma (véase la sección 10.16, La inserción postransduccional de la membrana depende de secuencias líder). Estas secuencias no son necesarias para las proteínas sintetizadas dentro del organelo. Quizá el proceso de transferencia efectiva de genes tuvo lugar en un periodo en el cual los compartimientos se definían de forma menos rígida, de modo que era más fácil que se reubicara el DNA y que las proteínas se incorporaran en el organelo, independientemente del sitio en que se sintetizaran.

Los mapas filogenéticos muestran que las transferencias de genes han sido independientes en muchos linajes diferentes. Parece que las transferencias de genes mitocondriales al núcleo ocurrieron sólo al principio de la evolución de las células animales, pero es posible que el proceso todavía continúe en las células de las plantas. El número de transferencias puede ser alto; en el organismo *Arabidopsis* hay >800 genes nucleares cuyas secuencias están relacionadas con las de genes ubicados en los cloroplastos de otras plantas, los cuales podrían haber evolucionado a partir de genes originados en el cloroplasto.

4.14 Resumen

Las secuencias de DNA que componen a un genoma eucariótico pueden ser clasificadas en tres grupos:

- secuencias no repetitivas únicas;
- secuencias moderadamente repetitivas que se dispersan y repiten pocas veces respecto de copias no idénticas;
- secuencias muy repetitivas, que son cortas y suelen repetirse en forma de secuencias en tándem.

Las proporciones de los tipos de secuencias son características de cada genoma, aunque más exten-

dos tienden a tener una proporción menor de DNA no repetitivo. Aproximadamente el 50% del genoma humano consta de secuencias repetitivas, y la vasta mayoría corresponde a secuencias de transposones. La mayoría de los genes estructurales se localiza en el DNA no repetitivo. La cantidad de DNA no repetitivo refleja mucho mejor la complejidad del organismo que el tamaño total del genoma; la mayor cantidad de DNA no repetitivo en los genomas es $\sim 2 \times 10^9$ bp.

La herencia no mendeliana se explica por la presencia de DNA en organelos del citoplasma. Las mitocondrias y los cloroplastos son sistemas delimitados por membranas en los cuales algunas proteínas son sintetizadas en el organelo, mientras que otras son importadas. En general, el genoma de los organelos es una molécula circular de DNA que codifica a todas las moléculas de RNA y a algunas de las proteínas requeridas por el organelo.

Los genomas mitocondriales varían enormemente de tamaño, de 16 kb en el genoma minimalista de los mamíferos, a 570 kb en el de plantas superiores. Los genomas más grandes pueden codificar funciones adicionales. Los genomas de los cloroplastos fluctúan entre 120 y 200 kb; en los secuenciados, la organización y las funciones codificadoras son similares. En las mitocondrias y los cloroplastos, muchas de las proteínas principales contienen algunas subunidades sintetizadas en el organelo y otras importadas del citosol.

En las levaduras, las reestructuraciones son muy frecuentes en el DNA mitocondrial, y se ha encontrado recombinación entre los genomas mitocondriales o los genomas de los cloroplastos. Se han observado transferencias de DNA entre los genomas de los cloroplastos o de las mitocondrias y los genomas nucleares.

Referencias

4.3 Los genomas individuales muestran una variación extensa

Artículos de investigación

- Altshuler, D., Brooks, L. D., Chakravarti, A., Collins, F. S., Daly, M. J., and Donnelly, P. (2005). A haplotype map of the human genome. *Nature* 437, 1299–1320.
- Altshuler, D., Pollara, V. J., Cowles, C. R., Van Euen, W. J., Baldwin, J., Linton, L., and Lander, E. S. (2000). An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature* 407, 513–516.
- Mullikin, J. C., Hunt, S. E., Cole, C. G., Mortimore, B. J., Rice, C. M., Burton, J., Matthews, L. H., Pavitt, R., Plumb, R. W., Sims, S. K., Ainscough, R. M., Atwood,

J., Bailey, J. M., Barlow, K., Bruskiwich, R. M., Butcher, P. N., Carter, N. P., Chen, Y., and Clee, C. M. (2000). An SNP map of human chromosome 22. *Nature* 407, 516–520.

4.4 Los RFLP y los SNP pueden ser utilizados para el mapeo genético

Artículos de revisión

Gusella, J. F. (1986). DNA polymorphism and human disease. *Annu. Rev. Biochem.* 55, 831–854.

White, R., Leppert, M., Bishop, D. T., et al. (1985). Construction of linkage maps with DNA markers for human chromosomes. *Nature* 313, 101–105.

Artículos de investigación

Altshuler, D., Brooks, L. D., Chakravarti, A., Collins, F. S., Daly, M. J., and Donnelly, P. (2005). A haplotype map of the human genome. *Nature* 437, 1299–1320.

Dib, C., Faure, S., Fizames, C., et al. (1996). A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* 380, 152–154.

Dietrich, W. F., Miller, J., Steen, R., et al. (1996). A comprehensive genetic map of the mouse genome. *Nature* 380, 149–152.

Donis-Keller, J., Green, P., Helms, C., et al. (1987). A genetic linkage map of the human genome. *Cell* 51, 319–337.

Hinds, D. A., Stuve, L. L., Nilsen, G. B., Halperin, E., Eskin, E., Ballinger, D. G., Frazer, K. A., and Cox, D. R. (2005). Whole-genome patterns of common DNA variation in three human populations. *Science* 307, 1072–1079.

Sachidanandam, R., Weissman, D., Schmidt, S., et al. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. The International SNP Map Working Group. *Nature* 409, 928–933.

4.5 ¿Por qué los genomas son tan grandes?

Artículos de revisión

Gall, J. G. (1981). Chromosome structure and the C-value paradox. *J. Cell Biol.* 91, 3s–14s.

Gregory, T. R. (2001). Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. *Biol. Rev. Camb. Philos. Soc.* 76, 65–101.

4.6 Los genomas eucarióticos contienen secuencias de DNA repetitivas y no repetitivas

Artículos de revisión

Britten, R. J. and Davidson, E. H. (1971). Repetitive and nonrepetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Q. Rev. Biol.* 46, 111–133.

Davidson, E. H. and Britten, R. J. (1973). Organization, transcription, and regulation in the animal genome. *Q. Rev. Biol.* 48, 565–613.

4.7 Los genes pueden ser aislados por la conservación de los exones

Artículos de investigación

Buckler, A. J., Chang, D. D., Graw, S. L., Brook, J. D., Haber, D. A., Sharp, P. A., and Housman, D. E. (1991). Exon amplification: a strategy to isolate mammalian genes based on RNA splicing. *Proc. Natl. Acad. Sci. USA* 88, 4005–4009.

Kunkel, L. M., Monaco, A. P., Middlesworth, W., Ochs, H. D., and Latt, S. A. (1985). Specific cloning of DNA fragments absent from the DNA of a male patient with an X chromosome deletion. *Proc. Natl. Acad. Sci. USA* 82, 4778–4782.

Monaco, A. P., Bertelson, C. J., Middlesworth, W., Colletti, C. A., Aldridge, J., Fischbeck, K. H., Bartlett, R., Pericak-Vance, M. A., Roses, A. D., and Kunkel, L. M. (1985). Detection of deletions spanning the Duchenne muscular dystrophy locus using a tightly linked DNA segment. *Nature* 316, 842–845.

4.9 Los organelos contienen DNA

Artículos de investigación

Cann, R. L., Stoneking, M., and Wilson, A. C. (1987). Mitochondrial DNA and human evolution. *Nature* 325, 31–36.

4.10 Los genomas de los organelos son moléculas circulares de DNA que codifican proteínas de los organelos

Artículos de revisión

Lang, B. F., Gray, M. W., and Burger, G. (1999). Mitochondrial genome evolution and the origin of eukaryotes. *Annu. Rev. Genet.* 33, 351–397.

4.11 La organización del DNA mitocondrial es variable

Artículos de revisión

Attardi, G. (1985). Animal mitochondrial DNA: an extreme example of economy. *Int. Rev. Cytol.* 93, 93–146.

Boore, J. L. (1999). Animal mitochondrial genomes. *Nucleic Acids Res.* 27, 1767–1780.

Clayton, D. A. (1984). Transcription of the mammalian mitochondrial genome. *Annu. Rev. Biochem.* 53, 573–594.

Gray, M. W. (1989). Origin and evolution of mitochondrial DNA. *Annu. Rev. Cell Biol.* 5, 25–50.

Artículos de investigación

Anderson, S., Bankier, A. T., Barrell, B. G., et al. (1981). Sequence and organization of the human mitochondrial genome. *Nature* 290, 457–465.

4.12 El genoma de los cloroplastos codifica a numerosas proteínas y moléculas de RNA

Artículos de revisión

- Palmer, J. D. (1985). Comparative organization of chloroplast genomes. *Annu. Rev. Genet.* 19, 325–354.
- Shimada, H. and Sugiura, M. (1991). Fine structural features of the chloroplast genome: comparison of the sequenced chloroplast genomes. *Nucleic Acids Res.* 11, 983–995.
- Sugiura, M., Hirose, T., and Sugita, M. (1998). Evolution and mechanism of translation in chloroplasts. *Annu. Rev. Genet.* 32, 437–459.

4.13 Las mitocondrias evolucionaron por endosimbiosis

Artículos de revisión

- Lang, B. F., Gray, M. W., and Burger, G. (1999). Mitochondrial genome evolution and the origin of eukaryotes. *Annu. Rev. Genet.* 33, 351–397.

Artículos de investigación

- Adams, K. L., Daley, D. O., Qiu, Y. J., Whelan, J., and Palmer, J. D. (2000). Repeated, recent and diverse transfers of a mitochondrial gene to the nucleus in flowering plants. *Nature* 408, 354–357.
- Arabidopsis Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815.
- Huang, C. Y., Ayliffe, M. A., and Timmis, J. N. (2003). Direct measurement of the transfer rate of chloroplast DNA into the nucleus. *Nature* 422, 72–76.
- Thorsness, P. E. and Fox, T. D. (1990). Escape of DNA from mitochondria to the nucleus in *S. cerevisiae*. *Nature* 346, 376–379.

Secuencias genómicas y números de genes

ESQUEMA DEL CAPÍTULO

5.1 Introducción

5.2 El número de genes bacterianos abarca un rango de un orden de magnitud

5.3 Se conoce el número total de genes de numerosas eucariotas

- Hay 6 000 genes en una levadura; 18 500 en un gusano; 13 600 en una mosca; 25 000 en la pequeña planta *Arabidopsis* y, probablemente de 20 000 a 25 000 en el ratón y el hombre.

5.4 ¿Cuántos tipos diferentes de genes hay?

5.5 El genoma humano tiene menos genes de los esperados

- Sólo 1% del genoma humano está formado por regiones codificadoras.
- Los exones forman ~5% de cada gen, por lo tanto, los genes (exones más intrones) constituyen ~25% del genoma.
- El genoma humano tiene de 20 000 a 25 000 genes.
- ~60% de los genes humanos son sometidos a corte y empalme alternativo.
- Hasta el 80% de los cortes y empalmes alternativos cambian la secuencia proteínica, de manera que el proteoma tiene ~50 000 a 60 000 miembros.

5.6 ¿Cómo se distribuyen los genes y otras secuencias en el genoma?

- Las secuencias repetidas (presentes en más de una copia) representan >50% del genoma humano.
- La mayor parte de las secuencias repetidas está formada por copias de transposones no funcionales.
- Hay numerosas duplicaciones de extensas regiones cromosómicas.

5.7 El cromosoma Y tiene varios genes específicos de la masculinidad

- El cromosoma Y tiene ~60 genes que se expresan específicamente en los testículos.
- Los genes específicos de la masculinidad están en múltiples copias de segmentos cromosómicos repetidos.
- La conversión génica entre copias múltiples permite el mantenimiento de los genes activos durante la evolución.

5.8 Las especies más complejas evolucionan agregando nuevas funciones génicas

- La comparación de diferentes genomas muestra un incremento constante en el número de genes conforme

se añaden genes adicionales para formar eucariotas, organismos multicelulares, animales y vertebrados.

- La mayoría de los genes únicos de los vertebrados están relacionados con el sistema nervioso o el inmunológico.

5.9 ¿Cuántos genes son esenciales?

- No todos los genes son esenciales. En las levaduras y las moscas, las deleciones de <50% de los genes tienen efectos detectables.
- Cuando dos o más genes son redundantes, una mutación en cualquiera de ellos puede no tener efectos detectables.
- No se sabe porqué sobreviven los genes que aparentemente son prescindibles en el genoma.

5.10 Los genes se expresan en niveles muy diferentes

- En una célula dada, la mayoría de los genes se expresa en niveles bajos.
- Sólo un pequeño número de genes, cuyos productos son especializados para un tipo celular específico, se expresan ampliamente.

5.11 ¿Cuántos genes se expresan?

- Los RNAm expresados en niveles bajos se superponen ampliamente en comparaciones de tipos de células.
- Las moléculas de RNAm abundantemente expresadas suelen ser específicas del tipo de célula.
- ~10 000 genes expresados pueden ser comunes a la mayoría de los tipos celulares de una eucariota superior.

5.12 El número de genes expresados puede medirse en masa

- La tecnología del chip permite tomar una instantánea de la expresión del genoma completo de una unidad celular de levadura.
- ~75% (~4 500 genes) del genoma de las levaduras se expresa en condiciones de cultivo normales.
- La tecnología del chip permite comparar detalladamente células animales relacionadas para determinar (por ejemplo) las diferencias de expresión entre una célula normal y una cancerígena.

5.13 Resumen

5.1 Introducción

Desde que en 1995 se secuenciaron los primeros genomas, la velocidad y el rango de secuenciación han mejorado extraordinariamente. Los primeros genomas secuenciados fueron genomas bacterianos pequeños, de <2 Mb, hacia 2002, el genoma humano de 3 000 Mb había sido secuenciado. A la fecha se han secuenciado los genomas de una gran variedad de organismos, entre otros, las bacterias, las arqueobacterias, las levaduras, las eucariotas inferiores, las plantas y los animales, incluidos, en este último caso, gusanos, moscas, roedores y algunos mamíferos.

Quizá la información más importante derivada de la secuencia de un genoma sea el número de genes. *Mycoplasma genitalium*, bacteria intracelular obligada, tiene el genoma más pequeño conocido de cualquier organismo, de sólo ~470 genes; el genoma de las bacterias de vida libre oscila entre 1 700 y 7 500. Los genomas de las arqueobacterias abarcan un rango similar; los de las eucariotas unicelulares empiezan en aproximadamente 5 300 genes, en tanto que los de gusanos y moscas tienen unos 18 500 y 13 500 genes, respectivamente; sin embargo, ese número se incrementa sólo a ~25 000 en el ratón y el hombre.

En la FIGURA 5.1 se resume el número mínimo de genes encontrado en seis grupos de organismos. Para formar una célula se necesitan ~500 genes; ~1 500 para una célula de vida libre; >5 000 para una célula con núcleo; >10 000 para un organismo multicelular y >13 000 para un organismo con sistema nervioso, pero obviamente muchas especies pueden tener más del número mínimo necesario para su tipo, por lo que el número de genes varía enormemente, aun entre especies íntimamente relacionadas.

En las bacterias y las eucariotas inferiores, la mayoría de los genes son únicos, sin embargo, en los genomas de las eucariotas superiores, los genes pueden dividirse en familias de miembros relacionados. No hay duda de que algunos genes son únicos (formalmente la familia tiene un solo miembro), pero muchos pertenecen a familias de diez o más miembros. El número de familias puede estar mejor relacionado con la complejidad global del organismo que el número de genes.

Parte de la información más reveladora proviene de la comparación de secuencias de genomas. Con las secuencias del genoma humano y del genoma del chimpancé de que se dispone actualmente, es posible empezar a abordar algunos de los aspectos de la interrogante sobre qué hace únicos a los humanos.

El número mínimo de genes oscila entre 500 y 30 000

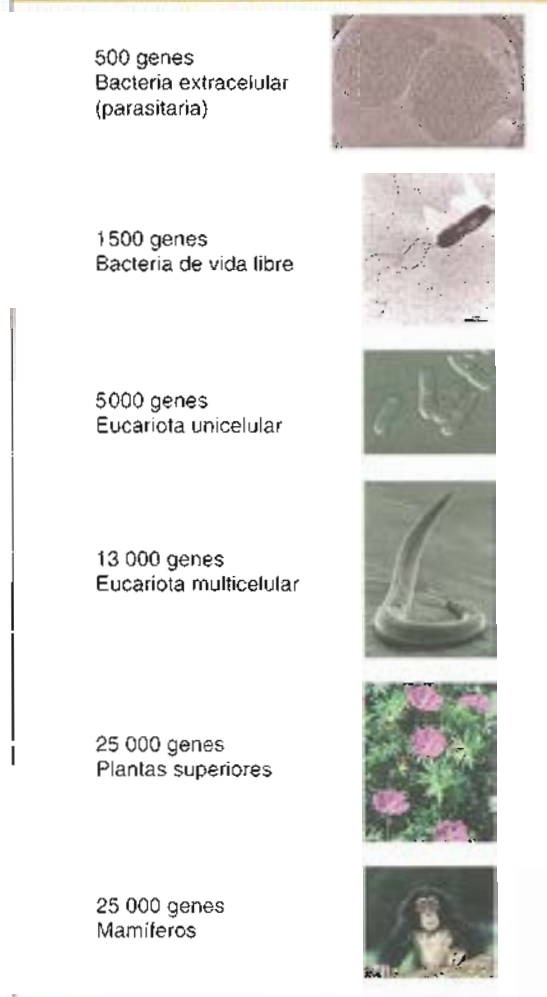


FIGURA 5.1 El número mínimo de genes requerido por cualquier tipo de organismo incrementa con su complejidad. Fotografía de bacteria extracelular cortesía de Gregory P. Henderson and Grant J. Jensen, California Institute of Technology. Fotografía de célula de vida libre cortesía de Karl O. Stetter, Universität Regensburg. Fotografía de eucariota unicelular cortesía de Eishi Noguchi, Drexel University College of Medicine. Fotografía de eucariota multicelular cortesía de Carolyn B. Marks and David H. Hall, Albert Einstein College of Medicine, Bronx, NY. Fotografía de planta superior cortesía de Keith Weller/USDA. Fotografía de mamífero © Photodisc.

5.2 El número de genes bacterianos abarca un rango de un orden de magnitud

Los intensos estudios actuales han permitido la secuenciación de numerosos genomas, entre otros, los incluidos en la FIGURA 5.2, en la cual se describe el tamaño de algunos de ellos, que va de 0.6×10^6 bp del micoplasma a 3.3×10^9 bp del genoma humano; se incluyen numerosos animales experimentales importantes, como las levaduras, la mosca de la fruta y un gusano nematodo.

El contenido de los genomas secuenciados varía de 470 a 30 000 genes			
Especies	Genomas (Mb)	Genes	Loci letales
<i>Mycoplasma genitalium</i>	0.58	470	~300
<i>Rickettsia prowazekii</i>	1.11	834	
<i>Haemophilus influenzae</i>	1.83	1 743	
<i>Methanococcus jannaschii</i>	1.66	1 738	
<i>B. subtilis</i>	4.2	4 100	
<i>E. coli</i>	4.6	4 288	1 800
<i>S. cerevisiae</i>	13.5	6 034	1 090
<i>S. pombe</i>	12.5	4 929	
<i>A. thaliana</i>	119	25 498	
<i>O. sativa</i> (arroz)	466	~30 000	
<i>D. melanogaster</i>	165	13 601	3 100
<i>C. elegans</i>	97	18 424	
<i>H. sapiens</i>	3 300	~25 000	

FIGURA 5.2 En numerosos organismos se conocen los tamaños de los genomas y los números de genes a partir de secuencias completas. Los loci letales se estiman a partir de bases de datos genéticos.

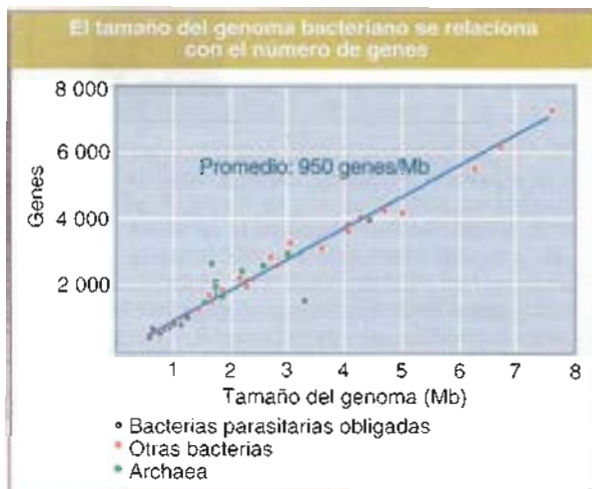


FIGURA 5.3 El número de genes de los genomas de las bacterias y de las archaea es proporcional al tamaño del genoma.

Las secuencias de los genomas de las bacterias y de las arqueobacterias demuestran que virtualmente todo el DNA (típicamente del 85 al 90%) codifica a moléculas de RNA o a proteínas. En la **FIGURA 5.3** se observa que el rango de dimensiones de los genomas es aproximadamente de un orden de magnitud y que las del genoma son proporcionales al número de genes, que suele tener una longitud aproximada de 1 000 bp.

Todas las bacterias con genomas de menos de 1.5 Mb son parásitos intracelulares obligados, es decir, viven en un hospedador eucariótico que les proporciona moléculas pequeñas. Sus genomas corresponden al número mínimo de funciones necesarias para construir una célula. El número de todas las clases de genes es más reducido respecto de las bacterias con genomas más grandes, pero la reducción más significativa se observa en los loci que codifican a enzimas relacionadas con funciones metabólicas (que dependen principalmente de la célula hospedadora) y con la regulación de la expresión génica. *Mycoplasma genitalium* tiene el genoma más pequeño, formado por ~470 genes.

Las propiedades biológicas de las archaea están entre las de las procariontes y las de las eucariotas, no obstante, el tamaño de su genoma y el número de genes están en el mismo rango que el de las bacterias. Dichas dimensiones fluctúan entre 1.5 y 3 Mb, que corresponde a 1 500 a 2 700 genes. *Methanococcus jannaschii* es una especie productora de metano que vive en condiciones de alta presión y alta temperatura, y el número total de genes que contiene es similar al de *Haemophilus influenzae*, aunque apenas unos cuantos pueden ser identificados comparándolos con genes conocidos de otros organismos. Sus mecanismos de expresión génica son más parecidos a los de las eucariotas que a los de las procariontes, pero el mecanismo de división celular se asemeja más al de estas últimas.

En las archaea y las bacterias de vida libre más pequeñas se identifica el número mínimo de genes necesarios para formar una célula susceptible de funcionar de manera independiente en el entorno; su genoma más pequeño tiene ~1 500 genes. La bacteria de vida libre con el genoma más pequeño conocido es el termófilo *Aquifex aeolicus*, que mide 1.5 Mb y tiene 1 512 genes. Una bacteria gramnegativa "típica", *H. influenzae*, tiene 1 743 genes, cada uno de ~900 bp. Por ello, se puede concluir que el número de genes necesarios para formar un organismo de vida libre es de ~1 500.

Las dimensiones de los genomas bacterianos cubren más o menos un orden de magnitud, de 0.6 Mb a <8 Mb, y los más grandes contienen más genes. Las bacterias cuyos genomas son los más grandes, *Sinorhizobium meliloti* y *Mesorhizobium loti*, son bacterias fijadoras de nitrógeno que viven en las raíces de las plantas. El tamaño de sus genomas (~7 Mb) y el número total de genes (>7 500) son similares a los de las levaduras.

El tamaño del genoma de *E. coli* es intermedio; la cepa común de laboratorio tiene 4 288 genes y una longitud promedio de ~950 bp; la separación promedio entre genes es de 118 bp. No obstante, puede haber diferencias muy significativas entre cepas: los

extremos conocidos van de 4.6 Mb y 4 249 genes de la cepa más pequeña, a 5.5 Mb y 5 361 genes de la más grande.

Todavía se desconocen las funciones de todos los genes. En la mayoría de estos genomas, ~60% de los genes pueden ser identificados con base en la homologación con genes conocidos de otras especies, los cuales se distribuyen más o menos de manera equivalente en clases cuyos productos se relacionan con el metabolismo, la estructura celular y el transporte de componentes, así como con la expresión génica y su regulación, si bien virtualmente a >25% de los genes es imposible atribuirle una función; muchos se encuentran en organismos relacionados, lo cual implica que se ha conservado la función.

Por su importancia médica, se ha hecho énfasis en la secuenciación de los genomas de las bacterias patógenas, y una perspectiva importante de la naturaleza de la patogenia proviene de la demostración de que las "islas de patogenicidad" son un rasgo característico de sus genomas. Dichas "islas" son regiones extensas, de ~10 a 200 kb, presentes en los genomas de especies patógenas, pero no en los genomas de las variantes no patógenas de la misma especie o de una especie relacionada. Su contenido de G-C suele diferir del contenido del resto del genoma, y es probable que estas bases migren entre bacterias merced a un proceso de transferencia horizontal. Por ejemplo, la bacteria que provoca el ántrax (*Bacillus anthracis*) tiene dos plásmidos (DNA extracromosómico) grandes, uno de los cuales tiene una isla de patogenicidad que incluye el gen codificador de la toxina del ántrax.

5.3 En numerosas eucariotas se conoce el número total de genes

Concepto principal

- Hay 6 000 genes en una levadura; 18 500 en un gusano; 13 600 en una mosca; 25 000 en la pequeña planta *Arabidopsis* y, probablemente ~25 000 en el ratón y el hombre.

Tan pronto como se observan los genomas eucarióticos, se pierde la relación entre su tamaño y el número de genes. Los genomas de las eucariotas unicelulares están en el mismo rango de tamaño que los genomas bacterianos más grandes. Las eucariotas superiores tienen más genes, pero el número no se correlaciona con el tamaño del genoma, como puede observarse en la FIGURA 5.4.

La información más amplia sobre las eucariotas inferiores se relaciona con las secuencias de los genomas de las levaduras *Saccharomyces cerevisiae* y

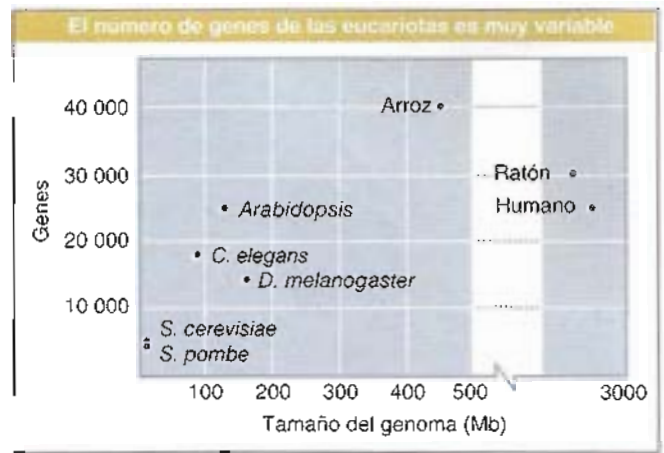


FIGURA 5.4 El número de genes de una eucariota varía de 6 000 a 40 000, pero no se correlaciona con el tamaño del genoma ni con la complejidad del organismo.

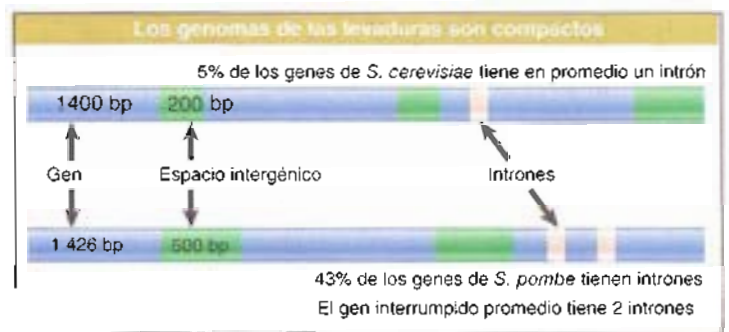


FIGURA 5.5 El genoma de *S. cerevisiae*, de 13.5 Mb, tiene 6 000 genes, casi todos interrumpidos. El genoma de *S. pombe*, de 12.5 Mb, tiene 5 000, de los cuales, casi la mitad tiene intrones. Las dimensiones de los genes y el espaciamiento son bastante similares.

Schizosaccharomyces pombe, cuyas características más importantes se resumen en la FIGURA 5.5. Los genomas de las levaduras de 12.5 Mb y 13.5 Mb tienen ~5 000 y ~6 000 genes, respectivamente. El marco de lectura abierto (ORF, *open reading frame*) promedio mide ~1.4 kb, de manera que ~70% del genoma lo ocupan regiones codificadoras. La diferencia principal entre estas cepas es que sólo el 5% de los genes de *S. cerevisiae* tiene intrones, respecto del 43% de los genes de *S. pombe*. La densidad de genes es alta, y la organización suele ser similar, aunque los espacios entre los genes son un poco más cortos en *S. cerevisiae*. Cerca de la mitad de los genes identificados por secuenciación ya se conocían o estaban relacionados con genes conocidos, los demás son nuevos, indicio del número de tipos nuevos que pueden ser descubiertos.

La identificación de marcos de lectura extensos con base en la secuencia es muy precisa, pero los ORF que codifican a <100 aminoácidos no pueden ser identificados únicamente mediante secuenciación por la elevada incidencia de falsos positivos. El



FIGURA 5.6 ~20% de los genes de *Drosophila* codifican a proteínas implicadas en el mantenimiento o la expresión de los genes; ~20% codifican a enzimas y <10% a proteínas relacionadas con el ciclo celular o con la transducción de señal. La mitad de los genes de *Drosophila* codifican a productos cuya función se desconoce.

análisis de la expresión génica sugiere que de ~300 a 600 de dichos ORF en *S. cerevisiae* tienden a ser genes genuinos.

Una forma eficaz de validar la estructura génica consiste en comparar secuencias de especies íntimamente relacionadas; si un gen está activo, es probable que se conserve. La comparación de las secuencias de cuatro especies de levaduras íntimamente relacionadas sugiere que 503 de los genes originalmente identificados en *S. cerevisiae* no tienen contraparte en las otras especies y, por lo tanto, deberían ser eliminadas del catálogo, con lo cual, el número total de genes para *S. cerevisiae* se reduce a 5 726.

El genoma del DNA de *Caenorhabditis elegans* varía entre regiones ricas en genes y otras en las cuales escasean. La secuencia total contiene ~18 500, pero sólo ~42% tiene contrapartes putativas fuera de los nematodos.

Aunque el genoma de la mosca es más grande que el del gusano, *D. melanogaster* tiene menos genes (13 600), y como resultado del corte y empalme alternativo, el número de transcritos diferentes es ligeramente mayor (14 100). No se sabe por qué la mosca, que es un organismo mucho más complejo, tiene sólo el 70% de genes del gusano, lo cual subraya enfáticamente la falta de una relación exacta entre el número de genes y la complejidad del organismo.

El tamaño del genoma de la planta *Arabidopsis thaliana* está entre el del gusano y el de la mosca, pero su número de genes es mayor (25 000), fenómeno que demuestra nuevamente que no hay una relación clara y subraya también la calidad especial de las plantas, que pueden tener más genes (debido a duplicaciones ancestrales) que las células animales. Gran parte del genoma de *Arabidopsis* se encuentra en segmentos duplicados, lo cual sugie-

re que hubo una duplicación ancestral del genoma (para dar lugar a un tetraploide); sólo el 35% de los genes de *Arabidopsis* son copias individuales.

El genoma del arroz (*Oryza sativa*) es ~4 veces más grande que el de *Arabidopsis*, pero el número de sus genes es sólo ~50% mayor, probablemente ~40 000. El DNA repetitivo ocupa del 42 al 45% del genoma. Más del 80% de los genes encontrados en *Arabidopsis* están representados en el arroz, y de estos genes comunes, ~8 000 se encuentran en *Arabidopsis* y en el arroz, pero no en los genomas bacterianos o animales que han sido secuenciados. Éste es probablemente el conjunto de genes que codifica funciones específicas de las plantas, como la fotosíntesis.

A partir del genoma de la mosca es posible darse una idea de cuantos genes están dedicados a cada tipo de función. En la **FIGURA 5.6** se dividen las funciones en categorías diferentes. Entre los genes identificados, se encuentran 2 500 enzimas, ~750 factores de transcripción, ~700 transportadores y canales de iones y ~700 proteínas implicadas en la transducción de señales. Poco más de la mitad de los genes codifica a productos de función desconocida. Aproximadamente el 20% de las proteínas reside en las membranas.

El tamaño de las proteínas se incrementa de las procariotas y las archaea, a las eucariotas. El tamaño promedio de las proteínas de la archaea *M. jannaschii* y la bacteria *E. coli* fluctúa entre 287 y 317 aminoácidos, respectivamente, mientras que *S. cerevisiae* y *C. elegans* tienen longitudes promedio de 484 y 442 aminoácidos, respectivamente. Las proteínas grandes (de 500 aminoácidos) son raras en las bacterias, pero constituyen un elemento importante (~1/3) de las eucariotas. El incremento de longitud se debe a la agregación de dominios adicionales, cada uno de los cuales constituye de 100 a 300 aminoácidos; no obstante, el mayor tamaño de las proteínas causa sólo una parte muy pequeña del incremento del tamaño del genoma.

Otra perspectiva del número de genes se obtiene al contar el número de genes expresados. Basándose en las estimaciones del número de especies de RNA_m que pueden contarse en una célula, se puede concluir que una célula promedio de un organismo vertebrado expresa de ~10 000 a 20 000 genes. Las significativas superposiciones entre poblaciones de mensajeros en diferentes tipos de células sugerirían que el número total de genes expresados en el organismo debería rebasar por poco esa cantidad. La estimación de 20 a 25 000 para el total del genoma humano (véase la sección 5.5, El genoma humano tiene menos genes de los esperados) implicaría que una proporción significativa del número total de genes de hecho se expresa en cualquier célula dada.

Los genes eucarióticos se transcriben de manera individual, y cada uno produce un mensajero monocistónico, y sólo hay una excepción general a esta regla: en el genoma de *C. elegans*, ~15% de los genes están organizados en unidades policistónicas (asociadas con el uso del empalme en *trans* para permitir la expresión de los genes en cascada en estas unidades; véase la sección 26.13. Las reacciones de empalme en *trans* utilizan moléculas pequeñas de RNA).

5.4 ¿Cuántos tipos diferentes de genes hay?

Algunos genes son únicos, en tanto que otros pertenecen a familias cuyos miembros están relacionados (pero usualmente no son idénticos). La proporción de genes únicos se reduce con el tamaño del genoma y la proporción de genes en familia se incrementa. El número mínimo de familias de genes necesario para codificar a una bacteria es >1 000, una levadura, >4 000 y una eucariota superior, de 11 000 a 14 000.

Algunos genes están presentes más de una vez o se relacionan entre sí, de modo que el número de tipos de genes es menor que el número total de genes. El número total de genes se puede dividir en conjuntos cuyos miembros están relacionados, según se define al comparar sus exones. (Una familia de genes surge por la duplicación de un gen ancestral seguida de la acumulación de cambios de secuencia entre copias. Es muy frecuente que los miembros de una familia estén relacionados, pero no son idénticos.) El número de tipos de genes se calcula sumando el número de genes únicos (para los cuales no hay otro gen relacionado) al número de familias que tienen dos o más miembros.

En la FIGURA 5.7 se compara el número total de genes con el número de familias distintas de cada uno de seis genomas. En las bacterias, la mayoría de los genes son únicos, así que el número de familias distintas se acerca al número total de genes. La situación es diferente incluso en la eucariota inferior *S. cerevisiae*, en la cual hay una proporción significativa de genes repetidos. El efecto más sobresaliente es que el número de genes se incrementa abruptamente en las eucariotas superiores, no así el de familias, que no cambia considerablemente.

En la FIGURA 5.8 se demuestra que la proporción de genes únicos disminuye bruscamente con el tamaño del genoma. Cuando los genes se presentan en familias, el número de miembros de una familia es reducido en las bacterias y las eucariotas inferiores, pero aumenta en las eucariotas superiores. Gran parte del tamaño adicional del genoma de *Arabidopsis* se debe a familias de >4 miembros.

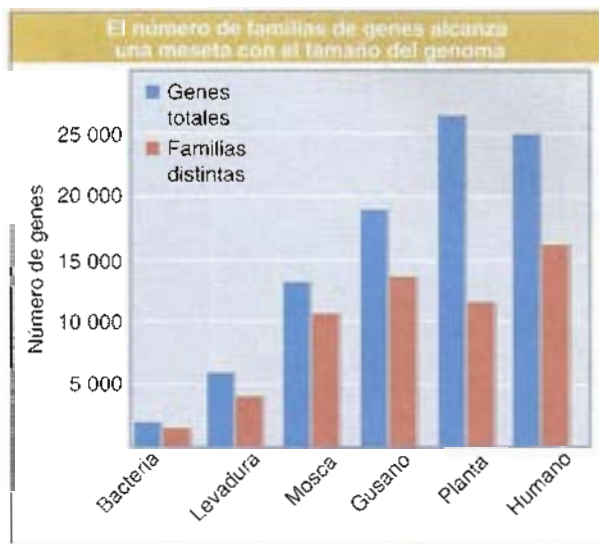


FIGURA 5.7 Numerosos genes están duplicados, de modo que el número de familias es mucho menor que el número total de genes. En el histograma se compara el número total de genes con el número de familias de genes.

	Genes únicos	Familias de 2-4 miembros	Familias de >4 miembros
<i>H. influenzae</i>	89%	10%	1%
<i>S. cerevisiae</i>	72%	19%	9%
<i>D. melanogaster</i>	72%	14%	14%
<i>C. elegans</i>	55%	20%	26%
<i>A. thaliana</i>	35%	24%	41%

FIGURA 5.8 La proporción de los genes presentes en múltiples copias se incrementa con el tamaño del genoma en las eucariotas superiores.

Si todos los genes se expresan, el número total representaría el número total de proteínas necesarias para constituir el organismo (proteoma), sin embargo, dos efectos significan que el proteoma es diferente del número total de genes. Por una parte, los genes están duplicados, y como resultado, algunos codifican a la misma proteína (aunque puede expresarse en un tiempo o lugar diferente) y otros pueden codificar a proteínas relacionadas que, nuevamente, tienen la misma función en diferentes tiempos y lugares. El proteoma puede ser más grande que el número de genes porque algunos pueden producir más de una proteína mediante corte y empalme alternativo.

¿Qué es el proteoma fundamental, el número básico de tipos diferentes de proteínas del organismo? El número de familias de genes proporciona



FIGURA 5.9 El genoma de la mosca puede dividirse en genes que podrían estar presentes en todas las eucariotas, en genes adicionales que probablemente están en todas las eucariotas multicelulares y en genes más específicos de subgrupos de especies incluídas las moscas.

un mínimo estimado, 1 400 en las bacterias, >4 000 en las levaduras y 11 000 a 14 000 en la mosca y el gusano.

¿Cuál es la distribución del proteoma entre tipos de proteínas? Las 6 000 proteínas del proteoma de las levaduras incluyen 5 000 proteínas solubles y 1 000 transmembrana; aproximadamente la mitad son citoplásmicas, un cuarto se encuentran en el nucleolo y el resto se dividen entre las mitocondrias y el retículo endoplásmico (ER, *endoplasmic reticulum*)/aparato de Golgi.

¿Cuántos genes son comunes a todos los organismos (o a grupos como las bacterias o las eucariotas superiores) y cuántos son específicos de cada tipo de organismo? En la **FIGURA 5.9** se resume la comparación entre levaduras, gusanos y moscas. Los genes que codifican a proteínas correspondientes en diferentes organismos se denominan **ortólogos**. Funcionalmente, suele considerarse que dos genes que se encuentran en organismos diferentes pueden realizar funciones correspondientes si sus secuencias son similares en >80% de la longitud. Con base en este criterio, ~20% de los genes de la mosca tienen ortólogos en las levaduras y los gusanos, genes que probablemente sean necesarios para todas las eucariotas. La proporción se incrementa a 30% cuando se compara a moscas y gusanos, lo cual probablemente representa la suma de las funciones génicas comunes a las eucariotas multicelulares. Esto aún deja una gran proporción de genes como codificadores de proteínas necesarias específicamente para las moscas o los gusanos, respectivamente.

El proteoma puede deducirse a partir del número y las estructuras de los genes, y también medir-

se directamente al analizar el contenido proteínico total de una célula o un organismo. Mediante dichas metodologías se ha identificado a algunas proteínas no sospechadas con base en el análisis del genoma, lo cual ha conducido a la identificación de genes nuevos. Para el análisis a gran escala de las proteínas se utilizan numerosos métodos. La espectrometría de masas permite separar e identificar proteínas en una mezcla obtenida directamente de células o tejidos. Las proteínas híbridas etiquetadas pueden obtenerse por expresión de moléculas de DNAc creadas al ligar las secuencias de los ORF con vectores de expresión apropiados que incorporan a las secuencias por etiquetas de afinidad, lo cual permite utilizar el análisis de matriz para evaluar los productos. Estos métodos también pueden ser efectivos para comparar proteínas provenientes de dos tejidos, por ejemplo, un tejido de un individuo sano y el de un enfermo, para determinar con precisión las diferencias.

Una vez que se conoce el número total de proteínas, entonces se puede preguntar cómo interactúan. Por definición, las proteínas que se encuentran en ensamblajes estructurales de proteínas múltiples deben formar interacciones estables entre ellas, pero las proteínas de las rutas de señalización interactúan de manera transitoria. En ambos casos, dichas interacciones pueden detectarse en sistemas experimentales en que un sistema de lectura amplía esencialmente el efecto de la interacción. Un sistema popular es el de dos híbridos, analizado en la sección 25.3. Los dominios independientes se unen al DNA y activan la transcripción. Dichos sistemas no detectan a todas las interacciones: por ejemplo, si en una ruta metabólica una enzima libera a un metabolito soluble que interactúa posteriormente con la siguiente enzima, las proteínas pueden no interactuar de forma directa.

En términos prácticos, los análisis de interacciones en pares pueden indicar el número mínimo de estructuras independientes o rutas. Un análisis de la capacidad de las 6 000 proteínas (pronosticadas) de las levaduras para interactuar en pares muestra que ~1 000 pueden unirse cuando menos a otra proteína. Mediante análisis directos de formación de complejos se han identificado 1 440 proteínas diferentes en 232 complejos multiproteínicos. Éste es el principio de un análisis que conducirá a la definición del número de ensamblajes funcionales o rutas. En un análisis comparable de 8 100 proteínas humanas se identificaron 2 800 interacciones, pero es más difícil de interpretar debido a las mayores dimensiones del proteoma.

Además de los genes funcionales, hay copias de genes que se han tornado no funcionales (identificados como tales por interrupciones de sus secuencias

codificadoras de proteínas) y que se conocen como pseudogenes (véase la sección 6.6, Los pseudogenes son callejones sin salida de la evolución), cuyo número puede ser alto. En el genoma del ratón y el humano, el número de pseudogenes es ~10% del número de genes (potencialmente) activos (véase la sección 4.8, La conservación de la organización del genoma ayuda a identificar genes).

Además de la necesidad de conocer la densidad de los genes para estimar el número total de éstos, se impone preguntar ¿cuál es su importancia?; ¿qué restricciones estructurales hacen necesario que estén espaciados, y si esto contribuye al gran tamaño de los genomas eucarióticos?

5.5 El genoma humano tiene menos genes de los esperados

Conceptos principales

- Sólo 1% del genoma humano está formado por regiones codificadoras.
- Los exones forman ~5% de cada gen, por lo tanto, los genes (exones más intrones) constituyen ~25% del genoma.
- El genoma humano tiene de 20 000 a 25 000 genes.
- ~60% de los genes humanos son sometidos a corte y empalme alternativo.
- Hasta el 80% de los cortes y empalmes alternativos cambian la secuencia proteínica, de manera que el proteoma tiene ~50 000 a 60 000 miembros.

El genoma humano fue el primer genoma de los organismos vertebrados en ser secuenciado. Esta enorme tarea ha revelado información valiosa respecto de la composición genética de nuestra especie y de la evolución del genoma en general, además de que los conocimientos se han profundizado por la capacidad de comparar la secuencia del genoma humano con el genoma del ratón, secuenciado más recientemente.

El genoma de los mamíferos y el de los roedores generalmente se encuentran en un rango de dimensiones estrecho, $\sim 3 \times 10^9$ bp (véase la sección 4.5, ¿Por qué los genomas son tan grandes?). El genoma del ratón es ~14% más pequeño que el humano, probablemente porque ha tenido una tasa de delección más alta. Los genomas contienen familias de genes y genes similares, y de éstos, la mayoría tiene un ortólogo en el otro genoma, pero con diferencias en el número de miembros de una familia, especialmente en los casos en que las funciones son específicas de cada especie (véase la sección 4.8, La conservación de la organización del genoma ayuda a identificar genes). Ahora se cree que el genoma

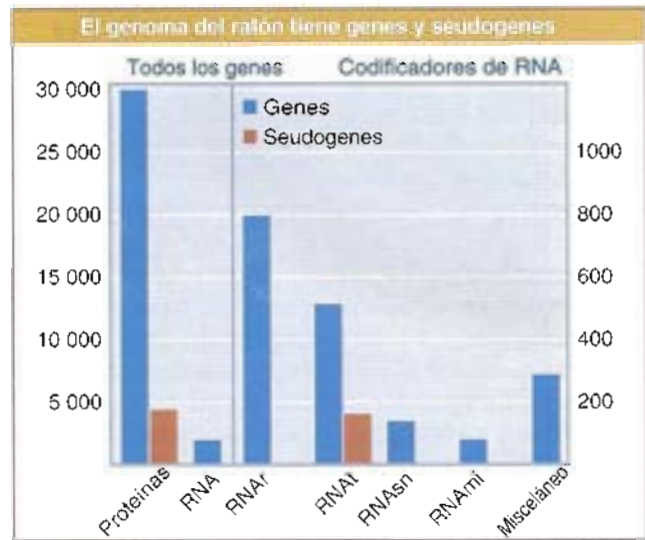


FIGURA 5.10 El genoma del ratón tiene ~30 000 genes codificadores de proteínas, los cuales tienen ~4 000 pseudogenes. Hay ~1 600 genes codificadores de RNA. Los datos de los genes codificadores de RNA se esquematizan a la derecha, en una escala ampliada, para demostrar que hay ~800 genes de RNAr, ~450 genes de RNAt, ~150 pseudogenes y ~350 genes misceláneos no codificadores de RNA, incluidos RNAsn y RNAmi.

del ratón, del cual originalmente se pensó que tenía ~30 000 genes, tiene casi el mismo número de genes que el genoma humano, de 20 000 a 25 000. En la **FIGURA 5.10** se representa la distribución de los genes del ratón. Los 30 000 genes codificadores de proteínas están acompañados por ~4 000 pseudogenes. Hay ~800 genes que representan a moléculas de RNA que no codifican proteínas, los cuales generalmente son pequeños (excepto las moléculas de RNA ribosómico). Aproximadamente la mitad de estos genes codifican a RNA de transferencia, para los cuales también ha sido identificado un número importante de pseudogenes.

El genoma humano (haploide) contiene 22 autosomas más el cromosoma X y el Y. El tamaño de los cromosomas varía de 45 a 279 Mb de DNA, que resulta en un contenido total del genoma de 3 286 Mb ($\sim 3.3 \times 10^9$ bp). Con base en la estructura cromosómica, el genoma total puede dividirse en regiones de eucromatina (que contienen potencialmente genes activos) y de heterocromatina (véase la sección 28.7, La cromatina se divide en eucromatina y en heterocromatina). La eucromatina comprende la mayor parte del genoma, $\sim 2.9 \times 10^9$ bp. La secuencia identificada del genoma representa ~90% de la eucromatina. Además de proporcionar información sobre el contenido genético del genoma, la secuencia también identifica características que pueden ser de importancia estructural (véase la sección 28.8, Los cromosomas tienen patrones de distribución en bandas).

En la FIGURA 5.11 se muestra que una pequeña proporción (~1%) del genoma humano es representada por exones que, de hecho, codifican a proteínas. Los intrones que constituyen las secuencias restantes en los genes llevan el total del DNA implicado en la producción de proteínas a ~25 por ciento. Como se observa en la FIGURA 5.12, el gen humano promedio tiene una longitud de 27 kb, con nueve exones que incluyen una secuencia codificadora total de 1 340 bp. La secuencia codificadora promedio representa, por lo tanto, sólo el 5% de la longitud del gen.

Dos esfuerzos independientes de secuenciar el genoma humano resultaron en estimados de ~30 000 y 40 000 genes, respectivamente. Una forma de determinar la precisión de los análisis consiste en averiguar si identifican a los mismos genes, y la sorprendente respuesta es que la superposición entre uno y otro conjunto de genes es de sólo ~50%, como se resume en la FIGURA 5.13. En un análisis anterior del conjunto de genes humanos, basado en transcritos de RNA, se identificaron ~11 000 genes,

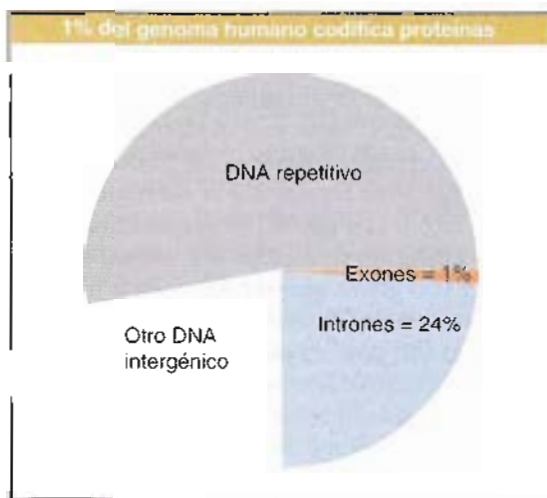


FIGURA 5.11 Los genes ocupan el 25% del genoma humano, si bien las secuencias codificadoras de proteínas son sólo una pequeña parte de esta fracción.

de los cuales casi todos están en los dos grandes conjuntos de genes humanos y representan la mayor parte de la superposición entre ellos. Así pues, no hay duda de la autenticidad de la mitad de cada conjunto de genes humanos, pero todavía se tiene que establecer la relación entre la mitad restante de cada uno de los conjuntos. ¡Las discrepancias ilustran los riesgos del análisis de secuencias a gran escala! Conforme se analiza la secuencia (y conforme son secuenciados otros genomas con los que puede ser comparada), el número de genes válidos parece declinar, de modo que ahora suele pensarse que son ~20 000 a 25 000.

Sea como sea, el número total de genes humanos es mucho menor de lo esperado; la mayor parte de las estimaciones anteriores a la secuenciación del genoma fueron de ~100 000. El número total de genes muestra un incremento relativamente pequeño respecto de las moscas y los gusanos (13 600 y 18 500, respectivamente), sin mencionar a la planta *Arabidopsis* (25 000) (véase la Fig. 5.2). Sin embargo, no debe sorprender la noción de que no se necesita un gran número de genes adicionales para formar un organismo más complejo. La diferencia de las secuencias de DNA del hombre y el chimpancé es extremadamente pequeña (la similitud es >99%), de modo que es obvio que las funciones y las interacciones entre un conjunto similar de genes pueden producir resultados muy diferentes. Las funciones de grupos específicos de genes pueden ser especialmente importantes, debido a que comparaciones detalladas de genes ortólogos en el hombre y el chimpancé sugieren una evolución acelerada de ciertas clases de genes, incluidas algunas implicadas en el desarrollo temprano, el olfato, la audición, todas funciones relativamente específicas de las especies.

El número de genes es menor que el número de proteínas potenciales debido al corte y empalme alternativo, cuya extensión es mayor en el hombre que en las moscas o los gusanos y puede afectar hasta el 60% de los genes, de manera que el incremen-

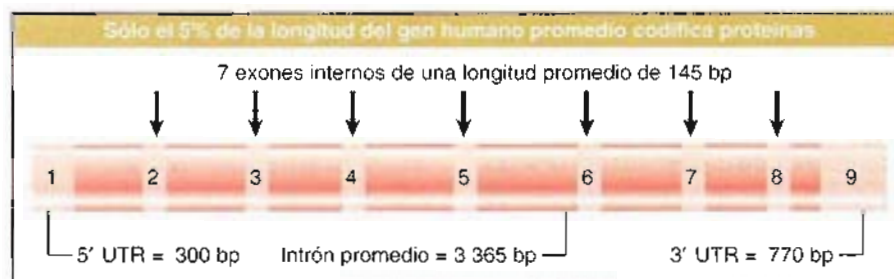


FIGURA 5.12 El gen humano promedio tiene una longitud de 27 kb y nueve exones, de los cuales, dos suelen ser más largos y estar en cada extremo, y siete exones internos. Las UTR de los exones terminales son regiones no traducidas (no codificadoras) localizadas en cada extremo del gen. (Estos datos se basan en el promedio. Algunos genes son extremadamente largos, de modo que la mediana de la longitud es de 14 kb, con siete exones.)

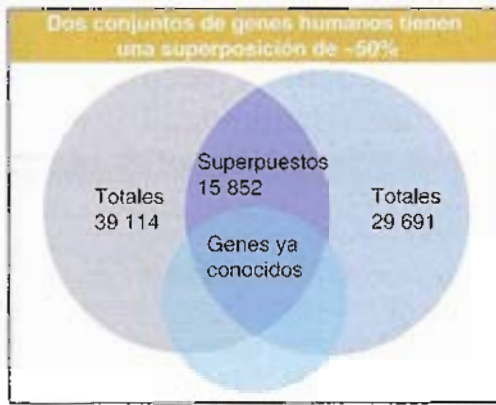


FIGURA 5.13 Los dos conjuntos de genes identificados en el genoma humano se superponen sólo parcialmente, como se muestra en los dos círculos grandes de la parte superior, no obstante, éstos incluyen casi todos los genes previamente conocidos, como lo demuestra la superposición con el círculo inferior más pequeño.

to de tamaño del proteoma humano en relación con el de otras eucariotas puede ser mayor que el incremento en el número de genes. Una muestra de genes provenientes de dos cromosomas sugiere que la proporción de cortes y empalmes alternativos, que de hecho resultan en cambios en las secuencias de las proteínas, puede ser hasta del 80 por ciento, lo cual puede incrementar el tamaño del proteoma a 50 000 a 60 000 miembros.

Sin embargo, en cuanto a la diversidad del número de familias de genes, la discrepancia entre el hombre y otras eucariotas puede no ser tan grande. Muchos de los genes humanos pertenecen a familias: en un análisis de ~25 000 genes se identificaron 3 500 únicos y 10 300 pares. Como puede observarse en la Figura 5.7, esto se extrapola a un número de familias de genes sólo un poco mayor que en los gusanos o las moscas.

5.6 ¿Cómo se distribuyen los genes y otras secuencias en el genoma?

Conceptos principales

- Las secuencias repetidas (presentes en más de una copia) representan >50% del genoma humano.
- La mayor parte de las secuencias repetidas está formada por copias de transposones no funcionales.
- Hay numerosas duplicaciones de extensas regiones cromosómicas.

¿Los genes se distribuyen uniformemente en el genoma? Algunos cromosomas cuentan con pocos

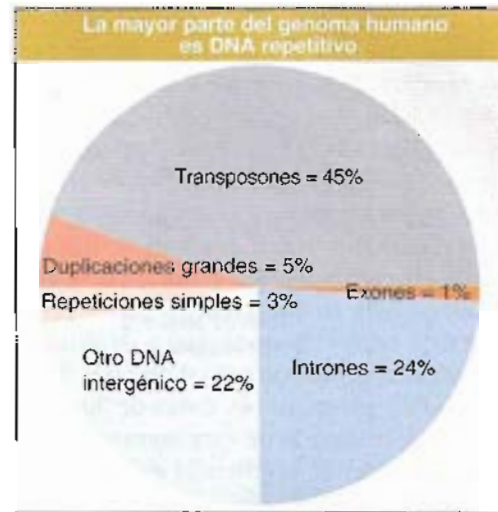


FIGURA 5.14 El componente más grande del genoma está formado por transposones. Otras secuencias repetitivas incluyen duplicaciones extensas y repeticiones sencillas.

genes y >25% de sus secuencias son "desiertos", o regiones de más de 500 kb en las que no hay genes, hasta >10% de las secuencias de los cromosomas con más genes lo son. Así, en total, ~20% del genoma humano consiste en desiertos carentes de genes.

Las secuencias repetitivas representan a >50% del genoma humano, como se observa en la **FIGURA 5.14**. Las secuencias repetitivas se agrupan en cinco clases:

- Los transposones (activos o inactivos) constituyen la gran mayoría (45% del genoma); todos se encuentran en múltiples copias.
- Los seudogenes procesados (~3 000, constituyen ~0.1% del DNA total). (Son secuencias que surgen por inserción de una copia de una secuencia de RNA_m en el genoma; véase la sección 6.6. Los seudogenes son los callejones sin salida de la evolución.)
- Repeticiones de secuencias simples (DNA muy repetitivo, como el (CA)_n que representa ~3%).
- Las duplicaciones segmentarias (bloques de 10 a 300 kb duplicados en una nueva región) representan a ~5%. Sólo una minoría de estas duplicaciones se encuentra en el mismo cromosoma; en los otros casos, las duplicaciones están en cromosomas diferentes.
- Las repeticiones en tándem forman bloques de un tipo de secuencia (especialmente en los centrómeros y los telómeros).

La secuencia del genoma humano subraya la importancia de los transposones (que tienen la capacidad de replicarse a sí mismos e insertarse en una ubicación diferente, además de que pueden funcionar exclusivamente como elementos del DNA [véase

Cap. 21. Los transposones] o tener una forma activa RNA [véase Cap. 22, Retrovirus y retroposones]. Su distribución en el genoma humano se resume en la Fig. 22.18). La mayoría de los transposones que se encuentran en el genoma humano no son funcionales; muy pocos están realmente activos, sin embargo, la gran proporción del genoma ocupada por estos elementos indica que han participado activamente en el molde del genoma. Una característica importante es que algunos genes presentes se originaron como transposones y evolucionaron hasta llegar a su condición actual después de perder la capacidad de transponerse. Cerca de 50 genes parecen haberse originado de esta manera.

La duplicación de segmentos en su forma más sencilla implica la duplicación en tándem de alguna región de un cromosoma (habitualmente debido a un evento de recombinación aberrante durante la meiosis; véase la sección 6.7, El entrecruzamiento desigual reorganiza a los agrupamientos de genes). Sin embargo, en muchos casos, las regiones duplicadas se encuentran en cromosomas diferentes, lo cual implica que originalmente hubo una duplicación en tándem seguida de la translocación de una copia a un sitio nuevo, o que la duplicación se debió a mecanismos totalmente diferentes. El caso extremo de la duplicación de segmentos se observa cuando se duplica todo el genoma, en cuyo caso, el genoma diploide inicialmente se convierte en tetraploide. Conforme las copias duplicadas desarrollan diferencias, el genoma puede convertirse en diploide nuevamente de forma gradual y efectiva, no obstante que las homologías entre las copias divergentes dejan evidencias del evento, fenómeno particularmente común en los genomas de las plantas. El estado actual del análisis del genoma humano identifica numerosas regiones duplicadas, pero no indica si hubo una duplicación completa del genoma en el linaje de los vertebrados.

Una característica intrigante del genoma humano es la presencia de secuencias que parecen no tener funciones codificadoras, pero que muestran una conservación evolutiva superior a la del nivel de fondo. Como se detectó mediante la comparación con otros genomas (inicialmente con el del ratón), dichas secuencias representan aproximadamente el 5% del genoma total. ¿Están conectadas funcionalmente con las secuencias codificadoras de proteínas? Su densidad en el cromosoma 18 es la misma que en cualquier otra región del genoma, aunque el cromosoma 18 tiene una concentración de genes codificadores de proteínas significativamente menor, lo cual sugiere indirectamente que su función no está relacionada con la estructura ni con la expresión de los genes codificadores de proteínas.

5.7

El cromosoma Y tiene varios genes específicos de la masculinidad

Conceptos principales

- El cromosoma Y tiene ~60 genes que se expresan específicamente en los testículos.
- Los genes específicos de la masculinidad están en múltiples copias de segmentos cromosómicos repetidos.
- La conversión génica entre copias múltiples permite el mantenimiento de los genes activos durante la evolución.

La secuenciación del genoma humano ha ampliado significativamente los conocimientos sobre la función de los cromosomas sexuales. En general se piensa que los cromosomas X y Y descienden de un autossoma común (muy antiguo). Su desarrollo ha implicado un proceso en el cual el cromosoma X ha retenido la mayor parte de los genes originales, mientras que el Y, ha perdido casi todos.

El cromosoma X se comporta como los autosomas en la medida en que las hembras tienen dos copias, y entre ellas puede tener lugar la recombinación. La densidad de genes localizados en el cromosoma X es comparable a la de otros cromosomas.

El cromosoma Y es mucho más pequeño que el X y tiene muchos menos genes. Su única función resulta de que sólo los machos portan el cromosoma Y, del cual sólo hay una copia, por lo tanto, los loci ligados a este último son efectivamente haploides y no diploides, como el resto de los genes humanos.

Durante muchos años se pensó que el cromosoma Y casi no portaba genes, excepto uno (o más) de los que determinan el sexo y definen la masculinidad. Gran parte del cromosoma Y (>95% de su secuencia) no experimenta entrecruzamiento con el cromosoma X, lo cual condujo a la perspectiva de que podía no contener genes activos porque no habría manera de evitar la acumulación de mutaciones deletéreas. Esta región está flanqueada por regiones pseudoautosómicas cortas que se intercambian frecuentemente con el cromosoma X durante la meiosis masculina. Originalmente, a esta región se le llamó no recombinante, pero ahora se le conoce como región específica determinante de la masculinidad.

La secuenciación detallada del cromosoma Y muestra que dicha región contiene tres tipos de regiones, como se ilustra en la **FIGURA 5.15**:

- Las *secuencias X transpuestas* están formadas por un total de 3.4 Mb que forman bloques grandes resultado de una transposición de la banda q21 en el cromosoma

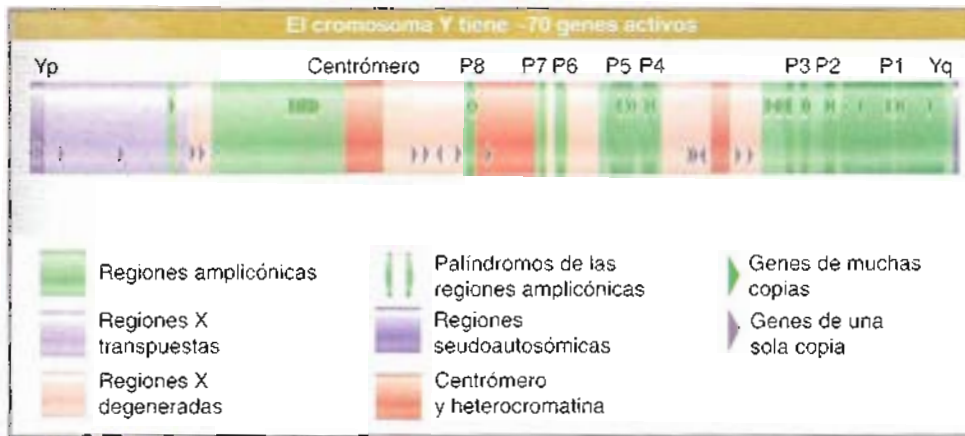


FIGURA 5.15 El cromosoma Y está formado por regiones X transpuestas, regiones X degeneradas y amplicones. Las regiones X transpuestas y degeneradas tienen dos y 14 genes únicos, respectivamente. Los amplicones tienen ocho palíndromos extensos (P1-P8), los cuales contienen a nueve familias de genes. Cada familia contiene por lo menos dos copias.

X hace aproximadamente tres o cuatro millones de años, lo cual es específico del linaje humano. Estas secuencias no se recombinan con el cromosoma X y prácticamente se han desactivado, ahora contienen sólo dos genes activos.

- Los *segmentos X degenerados* del cromosoma Y son secuencias que comparten un origen común con el cromosoma X (retorno al autósoma común del cual descienden los cromosomas X y Y) y contienen genes o pseudogenes relacionados con el cromosoma X. Hay 14 genes activos y 13 pseudogenes. En cierta forma, los activos han desafiado la tendencia a ser eliminados de las regiones cromosómicas que no pueden recombinarse durante la meiosis.
- Los *segmentos amplicónicos* tienen una longitud total de 10.2 Mb y se repiten internamente en el cromosoma Y. Hay ocho bloques palíndromos grandes que incluyen a nueve familias de genes codificadores de proteínas; el número de copias por familia fluctúa entre 2 y 35. El nombre "amplicón" refleja el hecho de que las secuencias han sido amplificadas internamente en el cromosoma Y.

Sumando los genes que se encuentran en estas tres regiones, el cromosoma Y contiene muchos más de lo que se hubiera esperado, hay 156 unidades de transcripción, de las cuales la mitad representa a genes codificadores de proteínas y la mitad representa a pseudogenes.

La presencia de los genes activos se explica por el hecho de que la existencia de copias de genes íntimamente relacionados en los segmentos amplicónicos permite que la conversión génica entre las

copias múltiples de un gen sea usada para regenerar copias activas. Las necesidades más comunes de las copias múltiples de un gen son cuantitativas (proporcionar más producto proteínico) o cualitativas (codificar proteínas con propiedades ligeramente distintas o expresadas en tiempos o lugares diferentes), aunque en este caso, la función esencial es evolutiva. De hecho, la existencia de copias múltiples permite la recombinación en el mismo cromosoma Y para sustituir a la diversidad evolutiva que suele ser proporcionada por recombinación entre los cromosomas alélicos.

La mayoría de los genes codificadores de proteínas en los segmentos amplicónicos se expresa específicamente en los testículos, y probablemente estén implicados en el desarrollo de la masculinidad. Si hay ~60 de dichos genes en un conjunto total de ~25 000 genes humanos, la diferencia genética entre varones y mujeres es de ~0.2%.

5.8 Las especies más complejas evolucionan al agregar nuevas funciones génicas

Conceptos principales

- La comparación de diferentes genomas muestra un incremento constante en el número de genes conforme se añaden genes adicionales para formar eucariotas, organismos multicelulares, animales y vertebrados.
- La mayoría de los genes únicos de los vertebrados están relacionados con el sistema nervioso o el inmunológico.

La comparación de la secuencia del genoma humano con secuencias encontradas en otras especies es reveladora respecto del proceso de evolución. En

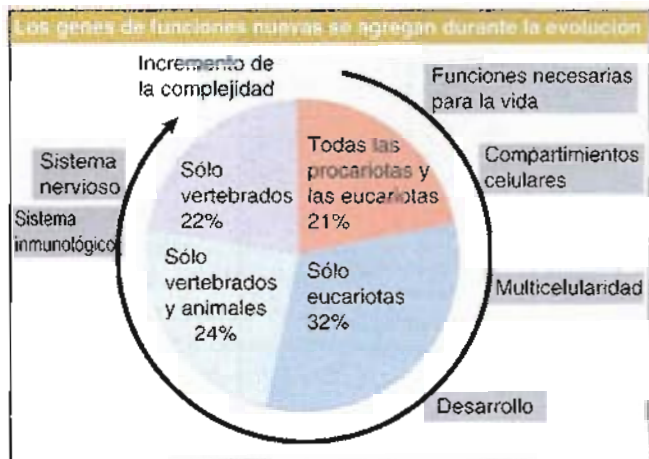


FIGURA 5.16 Los genes humanos pueden clasificarse según la amplitud de la distribución de sus homólogos en otras especies.

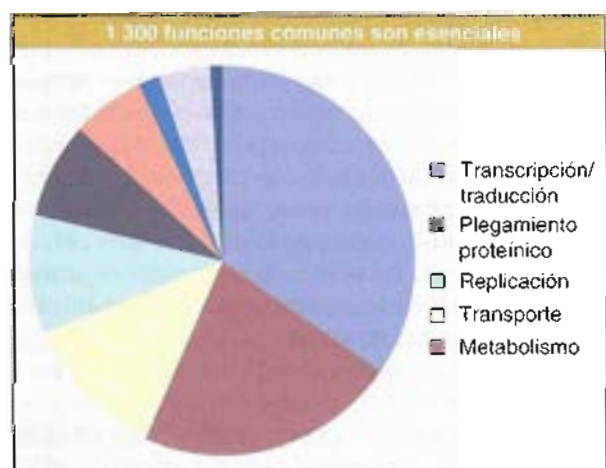


FIGURA 5.17 Las proteínas eucarióticas comunes están relacionadas con funciones celulares esenciales.

la FIGURA 5.16 se analizan los genomas humanos de acuerdo con la amplitud de su distribución en la naturaleza. Partiendo del más común (esquina superior derecha de la figura), el 21% de los genes es común a eucariotas y procariontas, y tienden a codificar proteínas esenciales para todas las formas vivas, típicamente para el metabolismo básico, la replicación, la transcripción y la traducción. Avanzando en el sentido de las manecillas del reloj, se agrega otro 32% de los genes a las eucariotas en general, por ejemplo, pueden encontrarse en las levaduras. Estos genes tienden a codificar proteínas implicadas en funciones generales de las células eucarióticas, pero no de las bacterias; podrían relacionarse, por ejemplo, con la especificación de organelos o con los componentes del citoesqueleto. Otro 24% de los genes son necesarios para la especificación de los animales, entre otros, los genes para la multicelularidad y el desarrollo de diferentes tipos de tejidos. Veintidós por ciento de los genes son únicos de vertebrados, y codifican principalmente a proteínas del

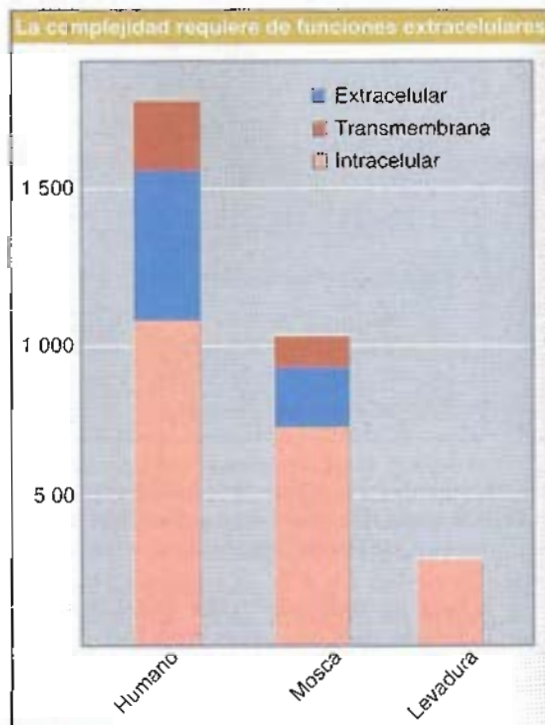


FIGURA 5.18 El incremento de la complejidad de las eucariotas se acompaña de la acumulación de proteínas nuevas para desempeñar funciones transmembrana e intracelulares.

sistema inmunológico y del nervioso, pero a muy pocas enzimas, lo cual coincide con la idea de que el origen de las enzimas es muy antiguo y que las rutas metabólicas se originaron al principio de la evolución. Por lo tanto, se observa que el avance de las bacterias a los vertebrados implica agregar grupos de genes que representan a las nuevas funciones necesarias en cada etapa.

Una manera de definir las proteínas comúnmente necesarias es identificar las que se encuentran en todos los proteomas. Comparando detalladamente el proteoma humano con los proteomas de otros organismos, 46% del de las levaduras, 43% del de los gusanos y 61% del de las moscas están representados en el humano; un grupo clave de ~1 300 proteínas está presente en los cuatro proteomas. Las proteínas comunes son proteínas domésticas básicas necesarias para funciones esenciales que corresponden a los tipos resumidos en la FIGURA 5.17. Las funciones principales están relacionadas con la transcripción y la traducción (35%), el metabolismo (22%), el transporte (12%), la replicación y modificación del DNA (10%), el plegamiento y la degradación de proteínas (8%) y los procesos celulares (6%).

Una de las sorprendentes características del proteoma humano es que tiene muchas proteínas nuevas respecto de otras eucariotas, pero relativamente pocos dominios proteínicos nuevos. La mayoría de los dominios proteínicos parecen ser comunes al rei-

no animal. Sin embargo, hay muchas arquitecturas proteínicas recientes definidas como combinaciones nuevas de dominios. En la FIGURA 5.18 se muestra que el mayor incremento tiene lugar en las proteínas transmembrana y las extracelulares. En las levaduras, la vasta mayoría de las arquitecturas tiene que ver con las proteínas intracelulares. En las moscas (o los gusanos) hay aproximadamente el doble de arquitecturas intracelulares, pero hay un incremento sorprendente de proteínas transmembrana y extracelulares, como es de esperar de la adición de funciones necesarias para las interacciones entre las células de un organismo multicelular. El incremento en las arquitecturas intracelulares necesario para formar un vertebrado (hombre) es relativamente pequeño, pero también en este caso se incrementan en gran medida la arquitectura transmembrana y la extracelular.

Desde hace mucho tiempo se sabe que la diferencia genética entre el hombre y el chimpancé (nuestro pariente más cercano) es muy pequeña, ~99% de los genomas son idénticos. Hoy día, la secuencia del genoma del chimpancé permite investigar con más detalle ese 1% de diferencias y detectar a las características responsables de "lo humano" pueden ser identificadas. La comparación muestra 35×10^6 sustituciones de nucleótidos (1.2% de diferencia global en la secuencia), 5×10^6 deleciones o inserciones (dejando en ~1.5% de la secuencia cromosómica, específica de cada especie) y muchas reestructuraciones cromosómicas. Las proteínas correspondientes suelen ser muy similares; 29% son idénticas y entre especies, en la mayoría de los casos hay sólo una o dos diferencias en los aminoácidos de la proteína. De hecho, las sustituciones de nucleótidos son menos frecuentes en los genes codificadores de proteínas que en los que tienden a estar involucrados en rasgos específicamente humanos, lo cual sugiere que la evolución de las proteínas no es un efecto importante en las diferencias entre el humano y el chimpancé, lo cual deja como posibilidades principales a los cambios de estructura génica en gran escala y los de la regulación. El 25% de las sustituciones de nucleótidos ocurre en los dinucleótidos CpG (entre los cuales hay muchos sitios reguladores potenciales).

- Cuando dos o más genes son redundantes, una mutación en cualquiera de ellos puede no tener efectos detectables.
- No se sabe por qué sobreviven los genes que aparentemente son prescindibles en el genoma.

La selección natural es la fuerza que garantiza que los genes útiles sean retenidos en el genoma; las mutaciones son aleatorias y su efecto más común en un ORF sería dañar al producto proteínico. Un organismo con mutaciones nocivas estaría en desventaja en la evolución, y en última instancia, la mutación sería eliminada por deficiencias de competitividad de los organismos que la portan. La frecuencia de un alelo deficiente en la población se equilibra entre la generación de mutaciones nuevas y la eliminación de mutaciones antiguas. Invirtiendo este argumento, siempre que se ve un ORF intacto en el genoma, se supone que su producto desempeña una función útil en el organismo. La selección natural debe haber impedido la acumulación de mutaciones en el gen. El destino final de un gen que deja de ser útil es acumular mutaciones hasta que deja de ser reconocible.

El mantenimiento de un gen implica que confiere una ventaja selectiva al organismo. Sin embargo, en el curso de la evolución, incluso una ventaja relativa pequeña puede ser objeto de la selección natural y un defecto fenotípico no ser necesariamente detectable de inmediato como resultado de una mutación. No obstante, sería importante saber cuántos genes son realmente *esenciales*, lo cual significa que su ausencia es letal para el organismo, lo cual, en el caso de los organismos diploides, significa por supuesto que la mutación nula homocigótica es letal.

Se puede suponer que la proporción de genes esenciales declinará con el incremento de tamaño del genoma porque los genomas más grandes pueden tener múltiples copias relacionadas con funciones génicas particulares, pero hasta ahora esta expectativa no ha sido confirmada por los datos (véase la Fig. 5.2).

Una forma de abordar el tema del número de genes es determinar mediante análisis mutacional el número de los que son esenciales. Si saturamos una región específica del cromosoma con mutaciones letales, las mutaciones deben mapearse en un número de grupos de complementación que corresponda al número de loci letales en dicha región. Al extrapolar este método al genoma como unidad global, se podrá calcular el número total de genes esenciales.

En el organismo que tiene el genoma más pequeño conocido (*M. genitalium*), las inserciones aleatorias tienen efectos detectables sólo en aproximadamente dos tercios de los genes; asimismo, menos

5.9 ¿Cuántos genes son esenciales?

Conceptos principales

- No todos los genes son esenciales. En las levaduras y las moscas, las deleciones de <50% de los genes tienen efectos detectables.

de la mitad de los genes de la bacteria *E. coli* parecen ser esenciales. La proporción es incluso más baja en la levadura *S. cerevisiae*. Cuando las inserciones se introdujeron aleatoriamente en el genoma en un

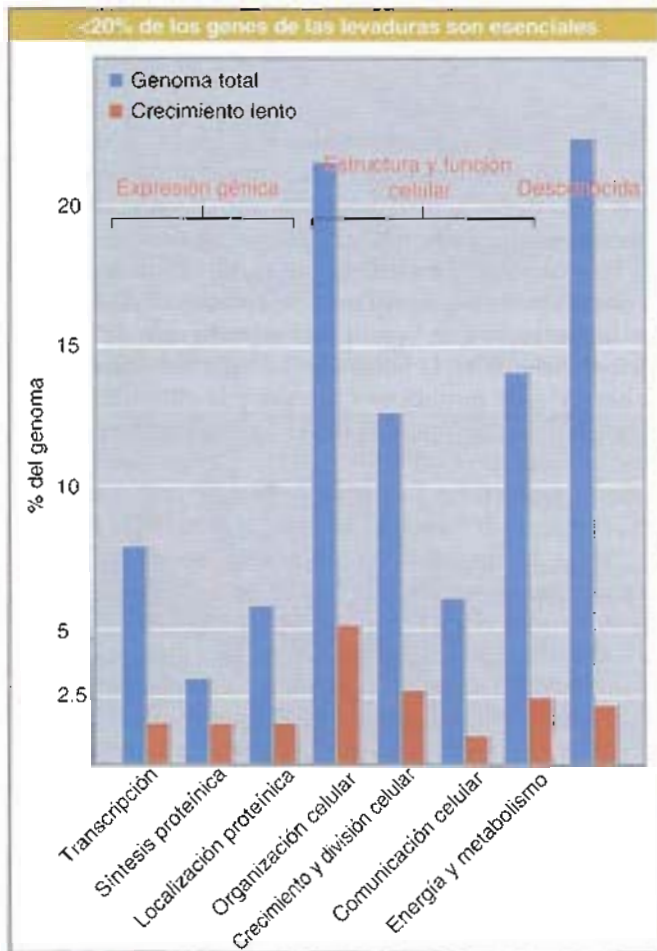


FIGURA 5.19 Los genes esenciales de las levaduras se encuentran en todas las clases. Las barras azules muestran la proporción total de cada clase de genes, en tanto que las rojas muestran a los esenciales.

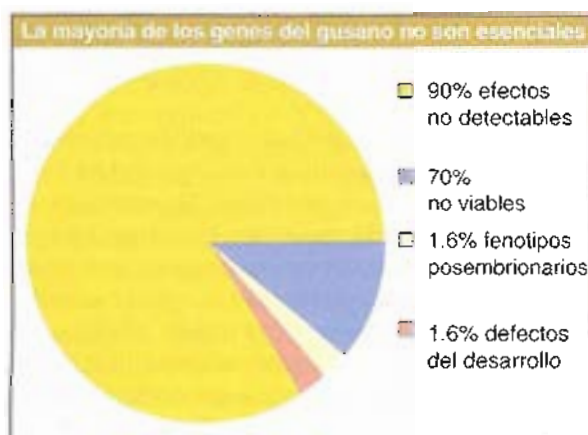


FIGURA 5.20 En un análisis sistemático de pérdida de función que se realizó en 86% de los genes del gusano se demostró que sólo el 10% tiene efectos detectables en el fenotipo.

análisis previo, sólo 12% fueron letales y otro 14% impedía el crecimiento. La mayor parte de las inserciones (70%) no tuvo ningún efecto. En un sondeo más sistemático, basado en la delección completa de cada uno de los 5 916 genes (>96% de los identificados), se muestra que sólo el 18.7% es esencial para el crecimiento en un medio de cultivo rico en nutrientes. En la FIGURA 5.19 se observa que las levaduras incluyen genes en todas las categorías. La única concentración notable de defectos está en los que codifican a productos implicados en la síntesis proteínica, en cuyo caso, ~50% son esenciales. Obviamente, en este método se subestima el número de genes esenciales para que la levadura viva en su hábitat natural, cuando no está tan bien provista de nutrientes.

En la FIGURA 5.20 se resumen los resultados de un análisis sistemático de los efectos de la pérdida de función génica en el gusano *C. elegans*. Las secuencias de genes individuales fueron pronosticadas a partir de la secuencia del genoma, y al dirigir un RNA inhibidor contra estas secuencias (véase la sección 13.10, La interferencia de RNA está relacionada con el silenciamiento de genes), se creó un conjunto grande de gusanos en el cual se impidió el funcionamiento de un gen (pronosticado) en cada gusano. Los efectos detectables en el fenotipo se observaron sólo en 10% de estos organismos con genes desactivados, lo cual sugiere que la mayoría de los genes no tiene funciones esenciales.

La proporción de genes esenciales es mayor (21%) entre los genes de gusanos que tienen contrapartes en otras eucariotas, lo cual sugiere que los genes ampliamente conservados tienden a desempeñar funciones más básicas. También se incrementa la proporción de genes esenciales entre los que se presentan individualmente en cada genoma haploide, respecto de los genomas que tienen muchas copias de genes relacionados o idénticos, lo cual sugiere que muchos de los genes múltiples pueden ser duplicaciones relativamente recientes que pueden sustituir mutuamente sus funciones.

En *Drosophila* se han realizado análisis extensos del número de genes esenciales en una eucariota superior mediante intentos de correlacionar los aspectos visibles de la estructura cromosómica con el número de unidades genéticas funcionales. La idea de que esto pueda ser posible provino originalmente de la presencia de bandas en los cromosomas politénicos de *D. melanogaster*. (Estos cromosomas se encuentran en determinadas etapas del desarrollo y representan una forma física inusualmente amplia, en la cual son evidentes una serie de bandas [denominadas más formalmente cromómeros]; véase la sección 28.10, Los cromosomas politénicos forman bandas.) A partir del concepto anterior de que

las bandas pueden representar un orden lineal de genes, se ha llegado al intento de correlacionar la organización de los genes con la organización de las bandas. En el conjunto haploide de *D. melanogaster* hay ~5 000 bandas cuyo tamaño varía en más de un orden de magnitud, pero en promedio hay ~20 kb de DNA por banda.

El método básico consiste en saturar una región cromosómica con mutaciones, las cuales suelen ser sencillamente reunidas como letales, sin analizar la causa de ello. Cualquier mutación letal se toma para identificar a un locus que es esencial para el organismo. En ocasiones, las mutaciones provocan efectos deletéreos visibles pero no letales, en cuyo caso, también se toman en consideración para identificar un locus esencial. Cuando las mutaciones se colocan en grupos de complementación, el número puede ser comparado con el número de bandas de la región, o incluso, cada grupo de complementación puede ser asignado a bandas individuales. El propósito de estos experimentos ha sido determinar si hay una relación coherente entre las bandas y los genes, por ejemplo, ¿cada banda tiene un solo gen?

Agrupando los análisis realizados en los últimos 30 años, el número de grupos de complementación letales es ~70% del número de bandas; aún se desconoce si esta relación tiene algún significado funcional. Independientemente de la causa, la equivalencia constituye un estimado razonable de ~3 600 genes letales; desde cualquier punto de vista, en *Drosophila*, el número de loci letales es significativamente menor que el número total de genes.

Si la proporción de genes esenciales humanos es similar a la observada en otras eucariotas, se pronosticaría un rango de 4000 a 8000 genes en los que las mutaciones serían letales o producirían efectos evidentemente dañinos. Actualmente se han identificado 1 300 genes en los cuales las mutaciones provocan defectos evidentes, proporción sustancial del total esperado, particularmente en vista de que muchos genes letales pueden actuar tan precozmente que nunca veríamos sus efectos. Este tipo de sesgos también suele explicar los resultados de la FIGURA 5.21, en la cual se observa que la mayoría de los defectos genéticos conocidos se debe a mutaciones puntuales (en donde es más probable que haya cuando menos alguna función residual del gen).

¿Cómo se explica la supervivencia de genes cuya delección parece no tener efecto alguno? La explicación más probable es que el organismo tiene formas alternas de cumplir con la misma función. La posibilidad más sencilla es la **redundancia**, y que hay muchas copias de algunos genes, lo cual es cierto en algunos casos, cuando muchos genes (relacionados) deben ser desactivados para producir un efecto. En un escenario un poco más complejo,

un organismo puede tener dos rutas independientes susceptibles de proporcionar cierta actividad. La desactivación de cualquiera de ellas no sería dañina, pero la ocurrencia simultánea de mutaciones en los genes de ambas sería deletérea.

Este tipo de situaciones puede evaluarse combinando mutaciones. Para empezar, en la misma cepa se introducen delecciones en dos genes, ninguna letal de por sí, pero si el mutante doble muere, la cepa se denomina **letal sintética**. Esta técnica ha resultado muy efectiva en las levaduras, en las cuales puede automatizarse el aislamiento de mutantes dobles, procedimiento denominado **análisis de matriz genética sintética (SGA, synthetic genetic array analysis)**. En la FIGURA 5.22 se resume el resultado de un análisis en el cual se realizó una SGA en cada una de las 132 delecciones viables para detectar si puede sobrevivir en combinación con cualquiera de las 4700 delecciones viables. Cada uno de los genes de prueba tuvo por lo menos un compañero con el cual la combinación fue letal, y en la mayor parte de los genes de prueba fue ese el caso; la mediana es de ~25 compañeros, y el número mayor se observó en un gen de prueba que tuvo 146 compañeros letales.

La mayoría de las mutaciones humanas que provocan defectos son pequeñas	
Sentido erróneo/sin sentido	58%
Corte y empalme	10%
Reguladoras	<1%
Delecciones pequeñas	16%
Inserciones pequeñas	6%
Delecciones grandes	5%
Reestructuraciones grandes	2%

FIGURA 5.21 La mayor parte de los defectos genéticos conocidos de los genes humanos se debe a mutaciones puntuales; casi todos afectan directamente a la secuencia proteínica. El resto se debe a inserciones, delecciones o reestructuraciones de diferentes tamaños.

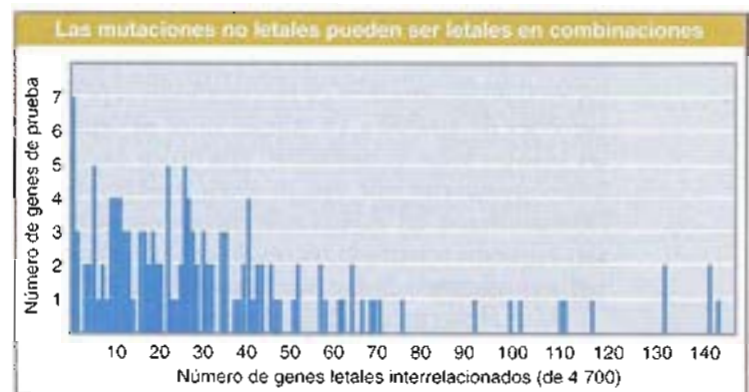


FIGURA 5.22 Los 132 genes mutantes de prueba presentan algunas combinaciones letales si se combinan con alguna de las 4700 mutaciones no letales. En la gráfica se ilustra cuántos genes letales interrelacionados hay por cada gen de prueba.

Una pequeña proporción (~10%) de los pares mutantes interrelacionados codificaron proteínas que interactúan físicamente.

Este resultado de alguna manera apunta hacia la explicación de la falta aparente de efecto de tantas deleciones. La selección natural actuará contra estas deleciones cuando se encuentren a sí mismas en combinaciones de pares letales. Hasta cierto punto, el organismo ha sido protegido contra los efectos dañinos de las mutaciones merced a la redundancia, pero paga el precio al acumular la "carga genética" de mutaciones que no son deletéreas por sí mismas, pero que pueden causar problemas serios cuando se combinen con otras mutaciones semejantes en generaciones futuras. La teoría de la selección natural sugeriría que la pérdida de genes individuales en dichas circunstancias produce una desventaja lo suficientemente grande como para mantener al gen activo durante la evolución.

5.10 Los genes se expresan en niveles muy diferentes

Conceptos principales

- En una célula dada, la mayoría de los genes se expresa en niveles bajos.
- Sólo un pequeño número de genes, cuyos productos son especializados para un tipo celular específico, se expresan ampliamente.

La proporción de DNA representada en una población de RNAm puede determinarse por la cantidad de DNA susceptible de hibridarse con el RNA. Dicho análisis de saturación habitualmente identifica a ~1% del DNA como proveedor de un molde para el RNAm. A partir de este descubrimiento es posible calcular el número de genes, siempre y cuando se conozca la longitud promedio de un RNAm. En una eucariota inferior, como una levadura, el número total de genes expresados es ~4 000, mientras que en los tejidos somáticos de las eucariotas superiores, suele ser de 10 000 a 15 000; el valor es similar en las plantas y los vertebrados. (La única excepción que coincide con este tipo de valor es el cerebro de los mamíferos, en el cual aparentemente se expresan números mucho mayores de genes, aunque no se ha confirmado la cantidad exacta.)

El análisis cinético de la reasociación de una población de RNA permite determinar la complejidad de su secuencia. Este tipo de análisis suele identificar a tres componentes de una célula eucariótica. Igual que en una curva de reasociación de DNA, un solo componente se hibrida en cerca de dos décadas de valores Rot (concentración de RNA × tiempo),

y una reacción que abarque un rango más amplio debe resolverse por ajuste computarizado de curvas en componentes individuales. También aquí, esto representa lo que realmente es un espectro continuo de secuencias.

Un ejemplo de una reacción RNAm excesivo × DNAc que genera tres componentes se muestra en la FIGURA 5.23:

- El primer componente tiene las mismas características que una reacción de control de RNAm de ovoalbúmina con su copia de DNA, lo cual sugiere que el primer componente es de hecho sólo RNAm de ovoalbúmina (que ciertamente ocupa aproximadamente la mitad de la masa de mensajero del tejido del oviducto).
- El siguiente componente proporciona 15% de la reacción, con una complejidad total de 15 kb; corresponde a 7 a 8 especies de RNAm con una longitud promedio de 2 000 bases.
- El último componente proporciona 35% de la reacción, que corresponde a una complejidad de 26 Mb, lo cual equivale a ~13 000 especies de RNAm con una longitud promedio de 2 000 bases.

A partir de este análisis se observa que aproximadamente la mitad de la masa del RNAm de la

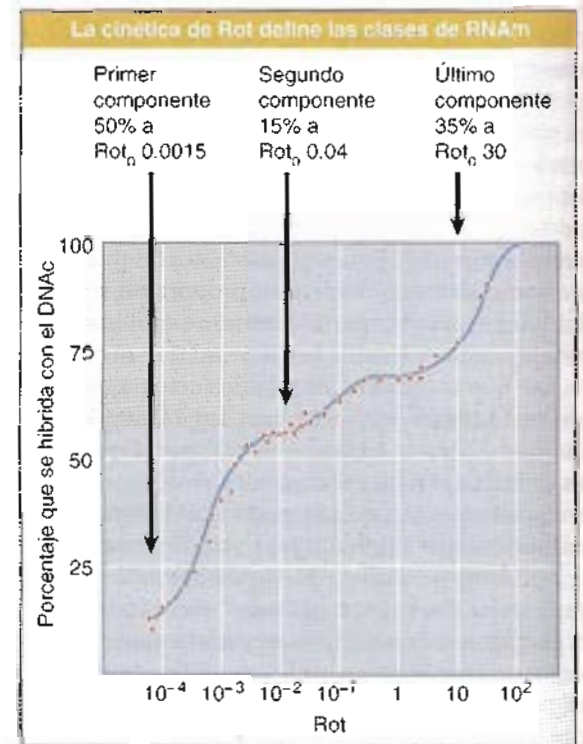


FIGURA 5.23 La hibridación entre RNAm excesivo y DNAc identifica a numerosos componentes de las células del oviducto de pollo caracterizadas por el $Rot_{1/2}$ de reacción.

célula representa un solo RNAm, ~15% de la masa es proporcionada por sólo 7 a 8 moléculas de RNAm y ~35% de la masa se divide en la enorme cifra de 13 000 especies de RNAm, por lo tanto, es obvio que las moléculas de RNAm que comprenden cada componente deben estar presentes en cantidades muy diferentes.

Al número promedio de moléculas de cada RNAm por célula se le denomina **abundancia**, que puede calcularse de forma muy sencilla si se conoce la masa total de RNA de la célula. En el ejemplo de la Figura 5.23, el RNAm total puede representarse como 100 000 copias del primer componente (RNAm de ovoalbúmina) y 4 000 de cada una de las 7 a 8 moléculas de RNAm del segundo componente, pero sólo ~5 copias de cada una de las 13 000 moléculas de RNAm que constituyen el último componente.

La población de RNAm se puede dividir en dos clases generales, de acuerdo a su abundancia:

- El oviducto es un caso extremo, con gran parte del RNAm representada en una sola especie, pero la mayor parte de las células contiene un número pequeño de moléculas de RNA en numerosas copias. Este **RNAm abundante** suele estar formado por <100 moléculas diferentes de RNAm en 1 000 a 10 000 copias por célula. Con frecuencia corresponde a gran parte de la masa, cerca de 50% del RNAm total.
- Aproximadamente la mitad de la masa de RNAm consiste en un número importante de secuencias, cerca de 10 000, cada una representada por sólo un pequeño número de copias en el RNAm, unas <10, que es la clase de **RNAm escaso** o **RNAm complejo**; es esta clase la que lleva a reacciones por saturación.

~13 000 a 15 000. ¿Cuántos de estos dos conjuntos de genes son idénticos? ¿Cuántos son específicos de cada tejido? Estas interrogantes suelen responderse al analizar el transcriptoma, o conjunto de secuencias representadas en el RNA.

De inmediato se observa que es probable que haya diferencias sustanciales entre los genes expresados en la clase abundante. La ovoalbúmina, por ejemplo, se sintetiza sólo en el oviducto y para nada en el hígado, lo cual significa que el 50% de la masa de RNAm del oviducto es específica de dicho tejido.

Sin embargo, las moléculas de RNAm abundante representan sólo una pequeña porción del número de genes expresados. Desde la perspectiva del número total de genes del organismo y del número de cambios que deben darse en la transcripción entre diferentes tipos de células, se necesita conocer la extensión de la superposición entre los genes representados en las clases de RNAm escaso de diferentes fenotipos celulares.

Las comparaciones entre tejidos diferentes muestran que, por ejemplo, ~75% de las secuencias expresadas en el hígado y en el oviducto son las mismas, en otras palabras, ~12 000 son expresados en el hígado y en el oviducto, ~5 000 genes adicionales sólo en el hígado y otros ~3 000 se expresan sólo en el oviducto.

Las moléculas de RNAm escaso se superponen ampliamente. Entre el riñón y el hígado del ratón, ~90% de las moléculas de RNAm escaso son idénticas, en función del número de genes expresados, una diferencia de sólo 1 000 a 2 000 entre tejido y tejido. El resultado general obtenido en varias comparaciones de este tipo revela que sólo ~10% de las secuencias de RNAm de una célula son únicas en ella. La mayoría de las secuencias son comunes en numerosos tipos celulares, quizá en todos.

Esto sugiere que el conjunto común de funciones de los genes expresados, tal vez ~10 000 en los mamíferos, comprende funciones necesarias en todos los tipos de células. En ocasiones, este tipo de genes se conoce como **genes de mantenimiento** o **genes constitutivos**; sus actividades contrastan con las representadas por funciones especializadas (como la ovoalbúmina o la globina), necesarias sólo para fenotipos celulares específicos. En ocasiones, a estos genes se les llama **genes especializados**.

5.11 ¿Cuántos genes se expresan?

Conceptos principales

- Los RNAm expresados en niveles bajos se superponen ampliamente en comparaciones de tipos de células.
- Las moléculas de RNAm abundantemente expresadas suelen ser específicas del tipo de célula.
- ~10 000 genes expresados pueden ser comunes a la mayoría de los tipos celulares de una eucariota superior.

El rango de genes expresados de muchos tejidos somáticos de eucariotas superiores es de 10 000 a 20 000. ¿Qué tanta es la superposición entre genes expresados en tejidos diferentes? Por ejemplo, el hígado de pollo se expresan de ~11 000 a 17 000 respecto del valor que se observa en el oviducto de

5.12 El número de genes expresados puede medirse *en masa*

Conceptos principales

- La tecnología del chip permite tomar una instantánea de la expresión del genoma completo de una unidad celular de levadura.

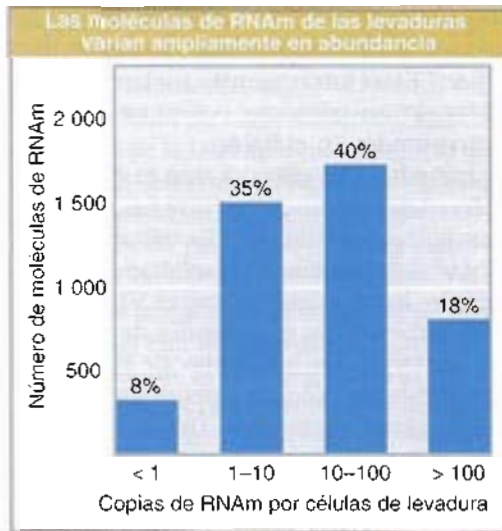


FIGURA 5.24 La abundancia de las moléculas de rRNA de las levaduras fluctúa entre <1 por célula (o sea que no todas las células tienen una copia de rRNA) y >100 por célula (que codifican a las proteínas más abundantes).

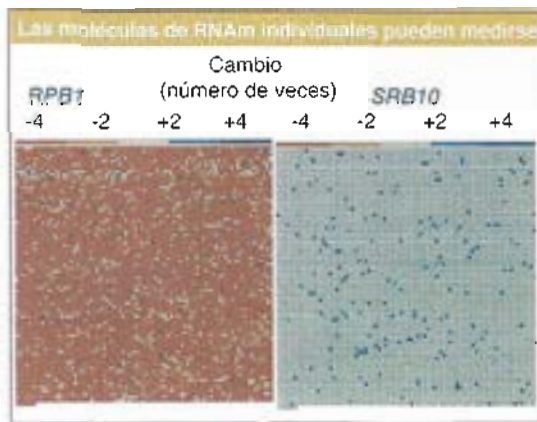


FIGURA 5.25 El análisis de HDA permite medir los cambios en la expresión de cada gen (la parte superior izquierda es el primer gen del cromosoma I y la inferior derecha, el último gen del cromosoma XVI). Los cambios en la expresión relativos a los tipos silvestres se indican en los colores rojo (reducción), blanco (sin cambios) o azul (incremento). Fotografías cortesía de Rick A. Young, Whitehead Institute, Massachusetts Institute of Technology.

- ~75% (~4 500 genes) del genoma de las levaduras se expresa en condiciones de cultivo normales.
- La tecnología del chip permite comparar detalladamente células animales relacionadas para determinar (por ejemplo) las diferencias de expresión entre una célula normal y una cancerígena.

La tecnología reciente permite estimar de manera más sistemática y precisa el número de genes expresados. Un método (análisis serial de la expresión génica o SAGE, *serial analysis of gene expression*) permite utilizar una etiqueta secuencial única para identificar a cada rRNA. La tecnología permite me-

dir posteriormente la abundancia de cada etiqueta. Mediante este método se identifican 4 665 genes expresados en *S. cerevisiae* cuando crece en condiciones normales, con abundancias que fluctúan entre 0.1 y >200 transcritos/célula, es decir, que ~75% del número total de genes (~6 000) se expresa en dichas condiciones. En la **FIGURA 5.24** se resume el número de moléculas diferentes de rRNA que se encuentran en cada nivel de abundancia.

La tecnología más poderosa utiliza chips que contienen matrices de oligonucleótidos de alta densidad (HDA, *high-density oligonucleotide arrays*) cuya construcción es posible porque se conoce la secuencia del genoma completo. En el caso de *S. cerevisiae*, cada uno de los 6 181 ORF están representados en la HDA por oligonucleótidos de 20-25-mer que corresponden perfectamente a la secuencia del mensaje y 20 oligonucleótidos no correspondientes que difieren en una posición de base. El nivel de expresión de cualquier gen se calcula sustrayendo la señal promedio de un apareamiento erróneo de su compañero perfecto. El genoma completo de las levaduras puede representarse en cuatro chips. Esta tecnología es lo suficientemente sensible como para detectar los transcritos de 5 460 genes (~90% del genoma) y mostrar que numerosos genes se expresan en niveles bajos, con abundancias de 0.1 a 0.2 transcritos/célula. Una abundancia de <1 transcrito/célula significa que no todas las células tienen una copia del transcrito en un momento dado.

La tecnología permite no sólo la medición de los niveles de expresión génica, sino también la detección de diferencias de expresión en células mutantes respecto de células de tipo silvestre que crecen en condiciones de cultivo diferentes, etcétera. Los resultados de la comparación de dos estados se expresan en una cuadrícula en la cual cada cuadro representa un gen específico y el cambio relativo en la expresión se indica mediante un color. La parte izquierda de la **FIGURA 5.25** muestra el efecto de una mutación en la RNA polimerasa II, la enzima que produce al rRNA, la cual, como puede esperarse, provoca que la expresión de la mayoría de los genes se reduzca mucho. La parte derecha de la figura, por el contrario, muestra que una mutación en un componente accesorio del mecanismo de transcripción (*SRB10*) tiene efectos mucho más restringidos y provoca incrementos en la expresión de algunos genes.

La extensión de esta tecnología a las células animales permitirá que las descripciones generales basadas en el análisis de hibridación del RNA sean remplazadas por descripciones exactas de los genes expresados y de las abundancias de sus productos, en cualquier tipo de célula dado. Un mapa de expresión génica de *D. melanogaster* detecta actividad de transcripción en alguna etapa del ciclo vital en casi

todos (93%) los genes pronosticados y muestra que el 40% tiene formas alternas de corte y empalme.

5.13 Resumen

Los genomas secuenciados incluyen numerosas bacterias y archaea, levaduras y un gusano, una mosca, un ratón y al hombre. El número mínimo de genes necesarios para formar una célula viviente (un parásito intracelular obligado) es ~470; para una célula de vida libre, ~1 700. Una bacteria gramnegativa típica tiene ~1 500 genes. El contenido génico de las cepas de *E. coli* fluctúa entre 4 300 y 5 400 genes. El gen bacteriano promedio tiene ~1 000 bp de longitud y se separa del siguiente por un espacio de ~100 bp. Las levaduras *S. pombe* y *S. cerevisiae* tienen 5 000 y 6 000 genes, respectivamente.

Aunque la mosca *D. melanogaster* es un organismo más complejo y tiene un genoma más grande que el del gusano *C. elegans*, la mosca tiene menos genes (13 600) que el gusano (18 500). La planta *Arabidopsis* tiene 25 000 genes y que no hay una relación clara entre el tamaño del genoma y el número de genes se demuestra porque el genoma del arroz es cuatro veces más grande, pero contiene sólo un incremento de 50% en el número de genes, a ~40 000. El ratón y el hombre tienen de 20 000 a 30 000 genes, cantidades mucho menores de lo esperado. La complejidad del desarrollo de un organismo puede depender de la naturaleza de las interacciones entre los genes, así como de su número total.

En las procariotas y en las eucariotas, aproximadamente 8 000 genes son comunes, y probablemente estén implicados en funciones básicas. En otros organismos multicelulares se pueden encontrar 12 000 genes. Para formar un animal, se agregan otros 8 000, y 8 000 más (profundamente relacionados con el sistema nervioso y el inmunológico) se encuentran en los vertebrados. En los organismos cuyo genoma ha sido secuenciado, sólo el 50% de los genes tiene funciones definidas. El análisis de genes letales sugiere que sólo una minoría de genes es esencial para cada organismo.

Las secuencias que constituyen un genoma eucariótico pueden clasificarse en tres grupos: las secuencias no repetitivas son únicas; las secuencias moderadamente repetitivas están dispersas y se repiten pocas veces como copias relacionadas, pero no idénticas, y las secuencias muy repetitivas son cortas, y se repiten como matrices en tándem. Las proporciones de los tipos de secuencias son características de cada genoma, aunque los genomas más grandes tienden a presentar una proporción menor de DNA no repetitivo. Aproximadamente el 50% del genoma humano consta de secuencias repetitivas, de las cuales, la gran mayoría corresponde

a secuencias de transposones. La mayor parte de los genes estructurales se localiza en el DNA no repetitivo, cuya complejidad es mejor indicador de la complejidad del organismo que la complejidad total del genoma: el DNA no repetitivo llega a tener una complejidad máxima de $\sim 2 \times 10^9$ bp.

Los genes se expresan en muy diversos niveles, de modo que puede haber 10^7 copias de rRNA de un gen abundante cuya proteína es el producto principal de la célula; 10^3 copias de cada rRNA de <10 mensajes moderadamente abundantes, y <10 copias de cada rRNA de >10 000 genes escasamente expresados. Las superposiciones entre poblaciones de rRNA de células con diferentes fenotipos son frecuentes; la gran mayoría de las moléculas de rRNA se encuentran en la mayoría de las células.

La herencia no mendeliana se explica por la presencia de DNA en organelos localizados en el citoplasma. Las mitocondrias y los cloroplastos representan a sistemas delimitados por membranas en los cuales algunas proteínas son sintetizadas en el organelo, mientras que otras son importadas. El genoma de los organelos suele ser un DNA circular que codifica a todos los rRNA y a algunas de las proteínas que necesita.

El tamaño de los genomas mitocondriales varía mucho, desde el genoma minimalista de los mamíferos, de 16 kb hasta el de 570 kb de las plantas superiores. Se supone que los genomas más grandes codifican funciones adicionales. Los genomas de los cloroplastos oscilan entre 120 y 200 kb; los que han sido secuenciados, tienen una organización y funciones codificadoras similares. Tanto en las mitocondrias como en los cloroplastos, muchas de las proteínas mayores contienen algunas subunidades sintetizadas en el organelo y otras importadas del citosol.

Las moléculas de DNA de los mamíferos se transcriben a un solo transcrito de la banda codificadora principal, en tanto que los productos individuales se generan por procesamiento de rRNA. En las levaduras, las reestructuraciones son bastante frecuentes en el DNA mitocondrial, además de que se han encontrado recombinaciones entre genomas mitocondriales o genomas de cloroplastos. Las transferencias de DNA han tenido lugar de los cloroplastos o las mitocondrias a los genomas nucleares.

Referencias

- 5.2 El número de genes bacterianos abarca un rango de un orden de magnitud

Artículos de revisión

- Bentley, S. D. and Parkhill, J. (2004). Comparative genomic structure of prokaryotes. *Annu. Rev. Genet.* 38, 771-792.

Hacker, J. and Kapur, J. B. (2000). Pathogenicity islands and the evolution of microbes. *Annu. Rev. Microbiol.* 54, 641–679.

Artículos de investigación

Blattner, F. R. et al. (1997). The complete genome sequence of *Escherichia coli* K-12. *Science* 277, 1453–1474.

Deckert, G. et al. (1998). The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature* 392, 353–358.

Galibert, F. et al. (2001). The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science* 293, 668–672.

5.3

Se conoce el número total de genes de numerosas eucariotas

Artículos de investigación

Adams, M. D. et al. (2000). The genome sequence of *D. melanogaster*. *Science* 287, 2185–2195.

Arabidopsis Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815.

C. elegans Sequencing Consortium (1998). Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 282, 2012–2022.

Duffy, A., and Grof, P. (2001). Psychiatric diagnoses in the context of genetic studies of bipolar disorder. *Bipolar Disord.* 3, 270–275.

Dujon, B. et al. (1994). Complete DNA sequence of yeast chromosome XI. *Nature* 369, 371–378.

Goff, S. A. et al. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* 296, 92–114.

Johnston, M. et al. (1994). Complete nucleotide sequence of *S. cerevisiae* chromosome VIII. *Science* 265, 2077–2082.

Kellis, M., Patterson, N., Endrizzi, M., Birren, B., and Lander, E. S. (2003). Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* 423, 241–254.

Oliver, S. G. et al. (1992). The complete DNA sequence of yeast chromosome III. *Nature* 357, 38–46.

Wilson, R. et al. (1994). 22 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*. *Nature* 368, 32–38.

Wood, V. et al. (2002). The genome sequence of *S. pombe*. *Nature* 415, 871–880.

5.4

¿Cuántos tipos diferentes de genes hay?

Referencias

Rual, J. F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Bertozzi, G. F., Gibbons, F. D., Dreze, M., Ayivi-Guedehoussou, N., et al. (2005). Towards a proteome-scale map of the human protein-protein interaction network. *Nature* 437, 1173–1178.

Artículos de revisión

Aebbersold, R. and Mann, M. (2003). Mass spectrometry-based proteomics. *Nature* 422, 198–207.

Hanash, S. (2003). Disease proteomics. *Nature* 422, 226–232.

Phizicky, E., Bastiaens, P. I., Zhu, H., Snyder, M., and Fields, S. (2003). Protein analysis on a proteomic scale. *Nature* 422, 208–215.

Sali, A., Glaeser, R., Earnest, T., and Baumcister, W. (2003). From words to literature in structural proteomics. *Nature* 422, 216–225.

Artículos de investigación

Agarwal, S., Heyman, J. A., Matson, S., Heidman, M., Piccirillo, S., Umansky, L., Drawid, A., Jansen, R., Liu, Y., Miller, P., Gerstein, M., Roeder, G. S., and Snyder, M. (2002). Subcellular localization of the yeast proteome. *Genes Dev.* 16, 707–719.

Arabidopsis Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815.

Gavin, A. C. et al. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415, 141–147.

Ho, Y. et al. (2002). Systematic identification of protein complexes in *S. cerevisiae* by mass spectrometry. *Nature* 415, 180–183.

Rubin, G. M. et al. (2000). Comparative genomics of the eukaryotes. *Science* 287, 2204–2215.

Uetz, P. et al. (2000). A comprehensive analysis of protein-protein interactions in *S. cerevisiae*. *Nature* 403, 623–630.

Venter, J. C. et al. (2001). The sequence of the human genome. *Science* 291, 1304–1350.

5.5

El genoma humano tiene menos genes de los esperados

Artículos de investigación

Clark, A. G. et al. (2003). Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. *Science* 302, 1960–1963.

Hogenesch, J. B., Ching, K. A., Batalov, S., Su, A. I., Walker, J. R., Zhou, Y., Kay, S. A., Schultz, P. G., and Cooke, M. P. (2001). A comparison of the Celera and Ensembl predicted gene sets reveals little overlap in novel genes. *Cell* 106, 413–415.

International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.

International Human Genome Sequencing Consortium (2004). Finishing the euchromatic sequence of the human genome. *Nature* 431, 931–945.

Venter, J. C. et al. (2001). The sequence of the human genome. *Science* 291, 1304–1350.

Waterston et al. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature* 420, 520–562.

5.6 ¿Cómo se distribuyen los genes y otras secuencias en el genoma?

Referencias

Nisbaum, C., Cody, M. C., Borowsky, M. L., Kamal, M., Kodira, C. D., Taylor, T. D., Whittaker, C. A., Chang, J. L., Cuomo, C. A., Dewar, K., et al. (2005). DNA sequence and analysis of human chromosome 18. *Nature* 437, 551–555.

5.7 El cromosoma Y tiene numerosos genes específicos de la masculinidad

Artículos de investigación

Skaletsky, H. et al. (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 423, 825–837.

5.8 Las especies más complejas evolucionan al agregar nuevas funciones génicas

Referencia

The Chimpanzee Sequencing and Analysis Consortium (2005). Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437, 69–87.

5.9 ¿Cuántos genes son esenciales?

Artículos de investigación

Gaever et al. (2002). Functional profiling of the *S. cerevisiae* genome. *Nature* 418, 387–391.
Goehl, M. G. and Petes, T. D. (1986). Most of the yeast genomic sequences are not essential for cell growth and division. *Cell* 46, 983–992.
Hutchison, C. A. et al. (1999). Global transposon mutagenesis and a minimal mycoplasma genome. *Science* 286, 2165–2169.
Kamath, R. S., Fraser, A. G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno,

S., Sohrmann, M., Welchman, D. P., Zipperlen, P., and Ahringer, J. (2003). Systematic functional analysis of the *C. elegans* genome using RNAi. *Nature* 421, 231–237.

Tong, A. H. et al. (2004). Global mapping of the yeast genetic interaction network. *Science* 303, 808–813.

5.10 Los genes se expresan en niveles muy diferentes

Artículos de investigación

Ilastic, N. B. and Bishop, J. O. (1976). The expression of three abundance classes of mRNA in mouse tissues. *Cell* 9, 761–774.

5.12 El número de genes expresados puede medirse en masa

Artículos de revisión

Mikos, G. L. G. and Rubin, G. M. (1996). The role of the genome project in determining gene function: insights from model organisms. *Cell* 86, 521–529.
Young, R. A. (2000). Biomedical discovery with DNA arrays. *Cell* 102, 9–15.

Artículos de investigación

Holstege, F. C. P. et al. (1998). Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* 95, 717–728.
Hughes, T. R., Marton, M. J., Jones, A. R., Roberts, C. J., Stoughton, R., Armour, C. D., Bennett, H. A., Coffey, E., Dai, H., He, Y. D., Kidd, M. J., King, A. M., Meyer, M. R., Slade, D., Lum, P. Y., Stepaniants, S. B., Shoemaker, D. D. et al. (2000). Functional discovery via a compendium of expression profiles. *Cell* 102, 109–126.
Stolc, V. et al. (2004). A gene expression map for the euchromatic genome of *Drosophila melanogaster*. *Science* 306, 655–660.
Velculescu, V. E. et al. (1997). Characterization of the yeast transcriptome. *Cell* 88, 243–251.

Agrupamientos y repeticiones

ESQUEMA DEL CAPÍTULO

- 6.1** Introducción
- 6.2** La duplicación de los genes es una fuerza importante en la evolución
- Los genes duplicados suelen separarse para dar lugar a genes diferentes o bien una de las copias se torna inactiva
- 6.3** Los agrupamientos de globinas se forman por duplicación y por divergencia
- Todos los genes de las globinas descienden por duplicación y mutación de un gen ancestral que tenía tres exones.
 - El gen ancestral dio origen a la mioglobina, la leghemoglobina y las globinas α y β
 - Los genes de las globinas α y β se separaron al principio del periodo de evolución de los vertebrados; después, las duplicaciones generaron agrupamientos individuales de genes similares a los genes α y β .
 - Una vez que un gen ha sido desactivado por una mutación, puede acumular más mutaciones y convertirse en un pseudogen homólogo al gen activo o los genes activos, pero que no tiene actividad funcional.
- 6.4** La divergencia secuencial es la base del reloj evolutivo
- Las secuencias de genes homólogos en especies diferentes varían en los sitios de remplazo (donde las mutaciones causan sustituciones de aminoácidos) y en los silenciosos (donde una mutación no afecta a la secuencia proteínica).
 - Las mutaciones se acumulan en los sitios silenciosos ~10 veces más rápido que en los de remplazo.
 - La divergencia evolutiva entre dos proteínas es medida por el porcentaje de posiciones en que difieren los aminoácidos correspondientes.
 - Las mutaciones se acumulan más o menos a una misma velocidad después de que los genes se separan, de manera que la divergencia entre cualquier par de secuencias de globina es proporcional al tiempo transcurrido desde la separación de sus genes.
- 6.5** La velocidad de sustitución neutral puede ser medida a partir de la divergencia de secuencias repetidas
- La velocidad de sustitución por año en sitios neutrales es mayor en el genoma del ratón que en el humano.
- 6.6** Los pseudogenes son callejones sin salida de la evolución
- Los pseudogenes carecen de función codificadora, sin embargo, pueden ser reconocidos por similitudes secuenciales con los genes funcionales existentes. Surgen de la acumulación de mutaciones en genes (anteriormente) funcionales.
- 6.7** Mediante entrecruzamiento desigual se reestructuran los agrupamientos de genes
- Cuando un genoma contiene un agrupamiento de genes con secuencias relacionadas, el apareamiento erróneo entre genes no alélicos puede causar un entrecruzamiento desigual, de modo que se produce una delección
- en un cromosoma recombinante y una duplicación correspondiente en el otro.
- Las talasemias diferentes son causadas por varias delecciones en que se elimina el gen de la globina α o el de la β . La gravedad de la enfermedad depende de cada delección.
- 6.8** Los genes del RNAr forman repeticiones en tándem
- El RNA ribosómico es codificado por un número considerable de genes idénticos que se repiten en tándem para formar uno o más agrupamientos.
 - Los agrupamientos de DNAr se organizan de manera tal, que las unidades de transcripción que resultan en un precursor unido a los RNAr principales alternan con espaciadores no transcritos.
- 6.9** Los genes repetidos de RNAr mantienen una secuencia constante
- Todos los genes de un agrupamiento de DNAr tienen secuencias idénticas.
 - Los espaciadores no transcritos consisten en unidades repetidoras más cortas cuyo número varía, de modo que la longitud de cada espaciador es diferente.
- 6.10** La fijación entrecruzada podría mantener repeticiones idénticas
- Un entrecruzamiento desigual cambia el tamaño de un agrupamiento de repeticiones en tándem.
 - Las unidades de repetición pueden eliminarse o repartirse entre los agrupamientos.
- 6.11** Los DNA satélite a menudo están en la heterocromatina
- La secuencia de repetición del DNA altamente repetitivo es muy corta y no tiene función codificadora.
 - Se presenta en grandes bloques que pueden tener propiedades físicas diferentes.
 - Suele ser el constituyente principal de la heterocromatina centromérica.
- 6.12** Los satélites de los artrópodos tienen repeticiones idénticas muy cortas
- La longitud de las unidades de repetición de los DNA satélite de los artrópodos es de sólo unos cuantos nucleótidos. La mayoría de las copias de la secuencia son idénticas.
- 6.13** Los satélites de los mamíferos constan de repeticiones jerárquicas
- El DNA satélite del ratón ha evolucionado por duplicación y mutación de una unidad de repetición corta para dar origen a una unidad de repetición básica de 234 bp en la que pueden ser reconocidas la mitad, la cuarta parte y la octava parte de la repetición original.
- 6.14** Los minisatélites facilitan el mapeo genético
- La variación entre los microsatélites o minisatélites de los genomas permite identificar la herencia inequívocamente, pues demuestra que el 50% de las bandas de un individuo derivan de un progenitor específico.
- 6.15** Resumen

6.1 Introducción

La familia de genes es un agrupamiento constituido por descendientes por duplicación y variación de algún gen ancestral cuyos miembros se agrupan o dispersan en diferentes cromosomas (o ambos). El análisis del genoma muestra que muchos genes pertenecen a familias, por ejemplo, los 25 000 del genoma humano identificados están en ~15 000 familias, así que el gen promedio tiene varios parientes en el genoma (véase la Fig. 5.7). Las familias de genes varían mucho en cuanto a grado de relación entre sus miembros, desde las que constan de múltiples miembros idénticos hasta aquellas en que la relación es bastante lejana. En general, los genes se relacionan únicamente por sus exones, pues sus intrones se han separado (véase la sección 3.5, Las secuencias de los exones se conservan, pero las de los intrones varían). Los genes también pueden estar relacionados sólo por algunos de sus exones, mientras que otros son únicos (véase la sección 3.9, Algunos exones pueden equipararse con funciones proteínicas).

El evento inicial que permite que los exones o los genes relacionados se desarrollen es una duplicación, es decir, que de alguna secuencia del genoma se genera una copia. La duplicación en tándem (cuando las duplicaciones permanecen juntas) puede provocar algunos errores de duplicación o de recombinación. La separación de los duplicados puede tener lugar por una **translocación** que transfiere material de un cromosoma a otro. Un duplicado en una ubicación nueva también puede ser producido directamente por un evento de transposición asociado con la copia de una región de DNA de los alrededores del transposón. La duplicación se aplica tanto a genes intactos como a colecciones de exones, incluso a exones individuales. Cuando se trata de un gen intacto, la duplicación genera dos copias cuyas actividades son indistinguibles, en cuyo caso, normalmente las copias difieren a medida que van acumulando mutaciones diferentes.

Los miembros de una familia de genes estructurales bien relacionados suelen tener funciones relacionadas, incluso idénticas, a pesar de que pueden ser expresadas en diferentes momentos o en diferentes tipos de células. Así pues, en los eritrocitos de embriones y en los eritrocitos adultos se expresan proteínas diferentes de globina, mientras que en las células musculares y en las no musculares se utilizan diferentes actinas. Cuando los genes se separan significativamente, o cuando sólo algunos de los exones están relacionados, las proteínas pueden tener diferentes funciones.

Algunas familias de genes están formadas por miembros idénticos, y el agrupamiento es un requisito previo para que lo sigan siendo, a pesar de

que los genes agrupados no son necesariamente idénticos. Los **agrupamientos de genes** van del caso en que una duplicación ha generado dos genes relacionados adyacentes hasta casos en que cientos de genes idénticos yacen en una matriz en tándem; esta última puede ser frecuentísima si se necesitan cantidades inusualmente grandes del producto, por ejemplo, los genes de los RNAr o de las histonas. Este fenómeno da lugar a una situación especial respecto del mantenimiento de la identidad y los efectos de la presión selectiva.

Los agrupamientos de genes representan una oportunidad para examinar a las fuerzas implicadas en la evolución del genoma en regiones más amplias que un gen. Las secuencias duplicadas, especialmente las que permanecen en la misma área, proporcionan el sustrato para una evolución posterior por recombinación. Una población evoluciona mediante la recombinación clásica ilustrada en las **FIGURAS 6.1 y 6.2**, en las cuales tiene lugar un entrecruzamiento exacto. Los cromosomas recombinantes están organizados de la misma manera que el cromosoma progenitor, contienen exactamente los mismos loci, en el mismo orden, pero las combinaciones de alelos difieren y proporcionan la materia prima para la selección natural. De cualquier forma, la existencia de secuencias duplicadas permite eventos aberrantes ocasionales que modifican el contenido de los genes y no sólo la combinación de los alelos.

El **entrecruzamiento desigual** (o **recombinación no recíproca**) describe un evento de recombinación entre dos sitios no homólogos. La característica que hace posible estos eventos es la existencia de secuencias repetidas. En la **FIGURA 6.3** se observa que esto permite que, en un cromosoma, la copia de una



FIGURA 6.1 La formación del quiasma representa la generación de recombinantes.

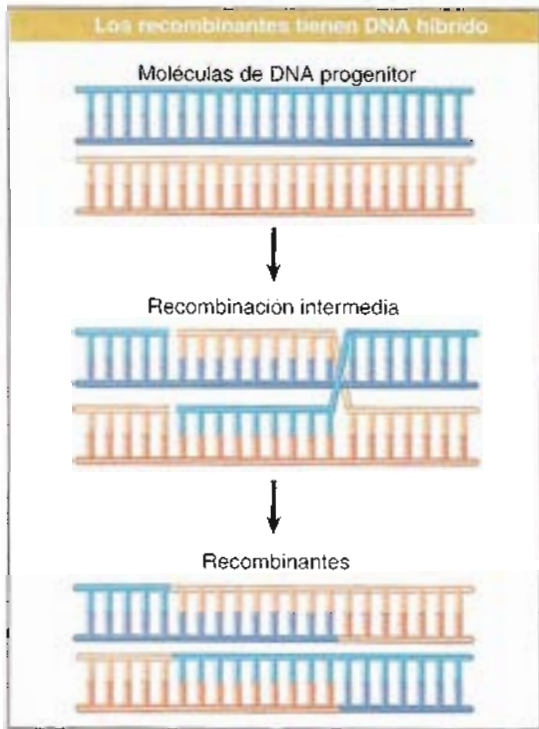


FIGURA 6.2 La recombinación implica apareamientos entre las cadenas complementarias de los dos DNA dúplex progenitores.



FIGURA 6.3 El entrecruzamiento desigual resulta del apareamiento entre repeticiones no equivalentes en regiones de DNA formadas por unidades de repetición. Aquí la unidad de repetición es la secuencia ABC, y la tercera repetición del cromosoma azul se ha alineado con la primera repetición del cromosoma negro. En toda la región de apareamiento, las unidades ABC de un cromosoma están alineadas con las del otro. El entrecruzamiento genera cromosomas con dieciséis repeticiones cada uno, y no ocho de cada progenitor.

repetición se desalinee y se recombine con una copia diferente de la repetición en el cromosoma homólogo y no con la copia correspondiente. Cuando sucede la recombinación, aumenta el número de repeticiones en un cromosoma y disminuye en el otro. De hecho, un cromosoma recombinante sufre una deleción y el otro una inserción. Este mecanismo es responsable de la evolución de los agrupamientos de secuencias relacionadas, y es posible rastrear su operación contrayendo o expandiendo el tamaño de una matriz

en los agrupamientos de genes y en las regiones de DNA muy repetido.

La fracción del genoma que se repite mucho está formada por múltiples copias en tándem de unidades de repetición muy cortas, que suelen tener propiedades inusuales, una de las cuales es que en un análisis de gradiente de densidad de DNA pueden ser identificadas como un pico independiente, lo cual da lugar al nombre de **DNA satélite**. A menudo se relacionan con regiones inertes de los cromosomas, en particular los centrómeros (que contienen los puntos de unión para la segregación en un huso mitótico o meiótico). Como resultado de su organización repetitiva, presentan parte del mismo comportamiento respecto de la evolución que los agrupamientos de genes en tándem. Además de las secuencias satélites, fragmentos más pequeños de DNA, llamados **minisatélites**, muestran un comportamiento similar, y permiten mostrar un alto grado de divergencia entre genomas específicos que pueden ser usados con fines de mapeo.

Estos eventos que modifican la constitución del genoma son raros, pero importantes, durante la evolución.

6.2 La duplicación de los genes es una fuerza importante en la evolución

Concepto principal

- Los genes duplicados suelen separarse para dar lugar a genes diferentes o bien una de las copias se torna inactiva.

Los exones se comportan como módulos para construir genes probados en varias combinaciones en el curso de la evolución. En un extremo, un exón individual de un gen puede ser copiado y utilizado en otro; en el otro extremo, un gen completo, incluidos exones e intrones, puede ser duplicado. En tal caso, las mutaciones suelen acumularse en una copia sin atraer la atención adversa de la selección natural, de modo que esta copia puede evolucionar a una nueva función, expresarse en un lugar o tiempo distintos a los de la primera copia, o bien, asumir actividades diferentes.

En la **FIGURA 6.4** se resume la visión actual del ritmo a que ocurren estos procesos. La probabilidad de que determinado gen sea incluido en una duplicación en un periodo de un millón de años es de ~1%, y después de que el gen se ha duplicado, las diferencias desarrolladas son resultado de las mutaciones ocurridas en cada copia, las cuales se acumulan a una velocidad de ~0.1% en un millón de



FIGURA 6.4 Después de que un gen ha sido duplicado, las diferencias pueden acumularse entre las copias. Los genes pueden adquirir funciones diferentes, o una de las copias tornarse inactiva.

años (véase la sección 6.4, La divergencia secuencial es la base del reloj evolutivo).

No es probable que el organismo necesite retener dos copias idénticas del gen, de modo que a medida que se desarrollan diferencias entre los genes duplicados, es probable que suceda uno de estos dos eventos:

- Ambos genes se vuelven necesarios porque las diferencias entre ellos generan proteínas con diferentes funciones o porque se expresan específicamente en lugares y momentos distintos.
- De no ser así, es probable que uno de los genes sea eliminado porque haya adquirido una mutación deletérea y no habrá una selección adversa para eliminar dicha copia, fenómeno que habitualmente toma ~4 millones de años. En dicha situación, es meramente una cuestión de probabilidad cuál de las copias se desactiva (lo cual puede contribuir a la incompatibilidad entre individuos, y, en última instancia, a la evolución de las especies, si diferentes copias se tornan inactivas en distintas poblaciones).

El análisis de la secuencia del genoma humano muestra que ~5% comprende duplicaciones de segmentos identificables cuya longitud oscila entre 10 y 300 kb. El número de dichas duplicaciones se ha elevado recientemente porque no ha habido suficiente tiempo como para que la divergencia entre ellas elimine su relación; incluyen una parte proporcional (de ~6%) de los exones expresados, lo cual

demuestra que las duplicaciones están sucediendo más o menos a pesar del contenido genético. Los genes de estas duplicaciones pueden ser especialmente interesantes por la implicación de que su evolución es reciente y, por lo tanto, pueden ser importantes para desarrollos evolutivos no demasiado lejanos (como la separación del hombre del mono).

6.3 Los agrupamientos de las globinas son formados por duplicación y por divergencia

Conceptos principales

- Todos los genes de las globinas descienden por duplicación y mutación de un gen ancestral que tenía tres exones.
- El gen ancestral dio origen a la mioglobina, la leghemoglobina y las globinas α y β .
- Los genes de las globinas α y β se separaron al principio del periodo de evolución de los vertebrados; después, las duplicaciones generaron agrupamientos individuales de genes similares a los genes α y β .
- Una vez que un gen ha sido desactivado por una mutación, puede acumular más mutaciones y convertirse en un pseudogén homólogo al gen activo o los genes activos, pero que no tiene actividad funcional.

El tipo más común de duplicación genera una segunda copia del gen cerca de la primera. En algunos casos, ambas copias permanecen asociadas y la duplicación posterior puede generar un agrupamiento de genes relacionados. El ejemplo mejor caracterizado de un agrupamiento de genes es el de los genes de globina, que constituyen una familia de genes antigua implicada en una función clave para el reino animal, el transporte de oxígeno por el torrente sanguíneo.

El constituyente mayor de los eritrocitos es el tetrámero de globina que se asocia con su grupo hem (de unión al hierro) en forma de hemoglobina. Los genes funcionales de globina tienen la misma estructura general en todas las especies, están divididos en tres exones ya ilustrados en la Figura 3.7. Se llega a la conclusión de que todos los genes de las globinas derivan de un solo gen ancestral, y siguiendo el desarrollo de los genes individuales de globina en una especie, o en especies diferentes, se llegan a conocer los mecanismos involucrados en la evolución de las familias de genes.

En las células adultas, el tetrámero de globina consiste en dos cadenas α idénticas y dos cadenas β idénticas. Las células sanguíneas de los embriones contienen tetrámeros de hemoglobina distintos de

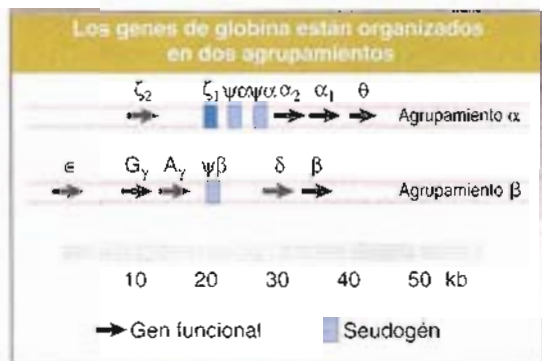


FIGURA 6.5 Cada una de las familias de genes de globina tipo α y tipo β , está organizada en un solo agrupamiento que incluye genes funcionales y pseudogenes (ψ).

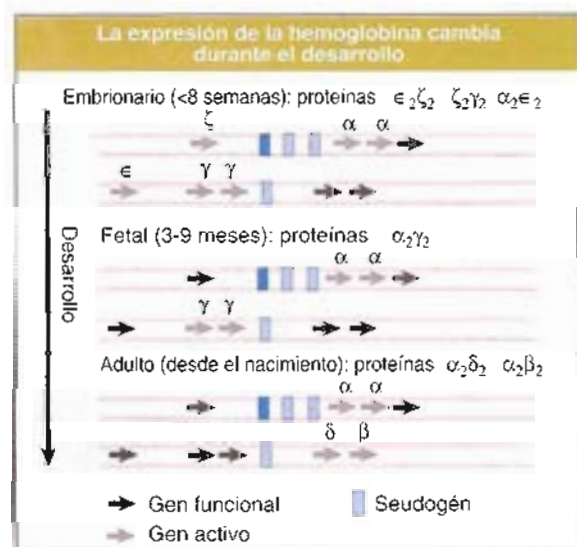


FIGURA 6.6 Durante el periodo embrionario, el fetal y el adulto del desarrollo humano se expresan genes de hemoglobina diferentes.

los de la forma adulta, pues cada tetrámero consta de dos cadenas idénticas tipo α y dos cadenas idénticas tipo β , cada una de las cuales está relacionada con el polipéptido adulto y posteriormente es remplazada por él. Éste es un ejemplo del control del desarrollo, en el cual diferentes genes se activan y desactivan sucesivamente para proporcionar productos alternos que realicen las mismas funciones en momentos diferentes.

La división de las cadenas de globina en tipo α y tipo β refleja la organización de los genes; cada tipo de globina es codificado por genes organizados en un solo agrupamiento. Las estructuras de los dos agrupamientos del genoma de los primates superiores se ilustran en la **FIGURA 6.5**.

Con una extensión de más de 50 kb, el agrupamiento β contiene cinco genes funcionales (ϵ , dos γ , δ y β) y un gen no funcional ($\psi\beta$). La secuencia co-

dificadora de los dos genes y difiere en sólo un aminoácido, la variante G tiene glicina en la posición 136, mientras que la variante A tiene alanina.

La longitud del agrupamiento α más compacto tiene más de 28 kb e incluye un gen activo ζ , un gen ζ no funcional, dos genes α , dos genes α no funcionales y el gen θ , cuya función se desconoce. Los dos genes α codifican a la misma proteína. Dos (o más) genes idénticos en el mismo cromosoma se describen como copias **no alélicas**.

Los detalles de la relación entre la hemoglobina embrionaria y la adulta varían en función del organismo. En el humano, la ruta tiene tres etapas, embrionaria, fetal y adulta. La distinción entre embrionaria y adulta es común en todos los mamíferos, pero varía el número de etapas previas a la adultez. En el hombre, ζ y α son dos cadenas tipo α , en tanto que ϵ , γ , δ y β son tipo β . En la **FIGURA 6.6** se muestra cómo se expresan las cadenas en diferentes etapas del desarrollo.

En la ruta del humano, ζ es la primera cadena tipo α en ser expresada, pero es rápidamente remplazada por α . En la ruta β , ϵ y γ se expresan primero y δ y β los remplazan después. En el adulto, la forma $\alpha_2\beta_2$ proporciona 97% de la hemoglobina y $\alpha_2\delta_2$, ~2%, en tanto que ~1% es aportado por la persistencia de la forma fetal $\alpha_2\gamma_2$.

¿Qué importancia tienen las diferencias entre la globina embrionaria y del adulto? Las formas embrionaria y fetal tienen mayor afinidad con el oxígeno para obtenerlo de la sangre de la madre, lo cual explica que no haya un equivalente en el pollo, por ejemplo, dado que las etapas embrionarias tienen lugar fuera del cuerpo (es decir, en el huevo).

Los genes funcionales se definen por su expresión en RNA y, en última instancia, por las proteínas a las cuales codifican. Los genes no funcionales se definen como tales por su incapacidad para codificar proteínas por diferentes razones; las deficiencias pueden estar en la transcripción o la traducción (o en ambas). A estos genes se les denomina **seudogenes** y se indican mediante el símbolo ψ . Una organización general similar se encuentra en otros agrupamientos de genes de globina de vertebrados, pero los detalles en cuanto a tipo, número y orden de los genes varía como se ilustra en la **FIGURA 6.7**. Cada agrupamiento contiene genes tanto embrionarios como adultos, y su longitud total varía ampliamente. El más largo se encuentra en la cabra, en donde un agrupamiento básico de 4 genes se ha duplicado 2 veces. La distribución de los genes activos y los pseudogenes difiere en cada caso, lo cual ilustra la naturaleza fortuita de la conversión de una copia de un gen duplicado al estado de inactividad.

La caracterización de estos agrupamientos de genes conduce a un punto general importante: *una*

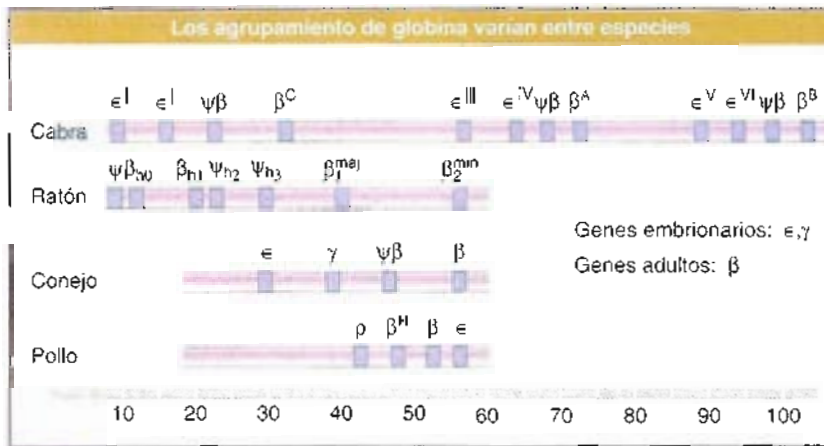


FIGURA 6.7 En los vertebrados se encuentran agrupamientos de genes de globina β y pseudogenes. Siete genes de ratón incluyen dos genes embrionarios tempranos, un gen embrionario tardío, dos genes adultos y dos pseudogenes. El conejo y el pollo tienen cuatro genes cada uno.

familia de genes puede tener más miembros, tanto funcionales como no funcionales, de lo que sospecharíamos con base en el análisis proteínico. Los genes funcionales adicionales pueden representar duplicaciones que codifican a polipéptidos idénticos o estar relacionados con las proteínas conocidas, pero de diferente forma (y presumiblemente se expresan sólo de forma breve o en cantidades bajas).

Respecto de la interrogante sobre cuánto DNA se necesita para codificar a una función en particular, se observa que la codificación de las globinas tipo β requiere de un rango de 20 a 120 kb en distintos mamíferos, mucho más de lo que se esperaría sólo de analizar las proteínas conocidas de la globina β , incluso del análisis de cada gen. De cualquier forma, los agrupamientos de este tipo no son comunes; la mayoría de los genes se encuentran como loci individuales.

A partir de la organización de los genes de la globina en una variedad de especies, debería ser posible rastrear la evolución de los actuales agrupamientos de genes de globina partir de un solo gen ancestral de globina. La perspectiva actual de la ascendencia evolutiva se ilustra en la **FIGURA 6.8**.

El gen de leghemoglobina de las plantas, relacionado con los genes de la globina, podría representar a la forma ancestral. Lo más que es posible remontarse en cuanto a un gen de globina en su forma moderna depende de la secuencia de la cadena individual de la mioglobina de los mamíferos, la cual se separó de la línea de ascendencia de la globina hace ~800 millones de años. El gen de la mioglobina tiene la misma organización que los genes de globina, de modo que es posible tomar la estructura de tres exones para representar a su ancestro común.

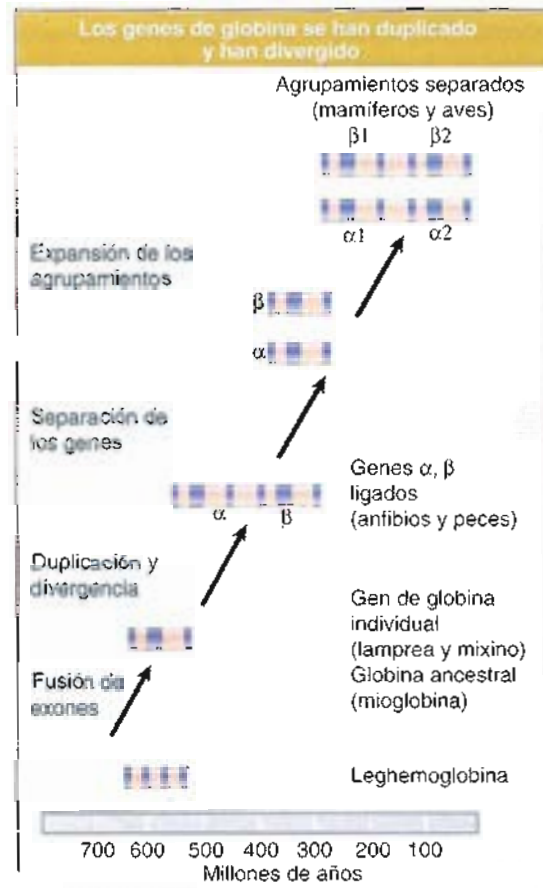


FIGURA 6.8 Todos los genes de globina han evolucionado por una serie de duplicaciones, transposiciones y mutaciones de un solo gen ancestral.

Algunos "peces primitivos" tienen sólo un tipo de cadena de globina, así que deben haberse separado de la línea de la evolución antes de que el gen

de globina ancestral se duplicara para dar lugar a las variantes α y β , lo cual aparentemente ocurrió hace ~500 millones de años, durante la evolución de los osteicnios.

La siguiente etapa de la evolución está representada por el estado de los genes de globina de la rana *Xenopus laevis*, que tiene dos agrupamientos de genes de globina. No obstante, cada agrupamiento tiene *ambos* genes, α y β , tanto de tipo larvario como adulto, de modo que el agrupamiento debe haber evolucionado por duplicación de un par $\alpha\beta$ ligado, seguida de la divergencia entre cada copia. Más tarde, el agrupamiento entero se duplicó.

Los anfibios se separaron de la línea de mamíferos/aves hace ~350 millones de años, así que la separación de los genes de globina α y β debe haber resultado de una transposición en la línea predecesora de mamíferos/aves después de dicho periodo, lo cual probablemente ocurrió en las primeras etapas de la evolución de los vertebrados. Tanto en aves como en mamíferos hay agrupamientos independientes de globinas α y β , de modo que los genes alfa y beta deben haberse separado físicamente antes de que mamíferos y aves divergieran de su ancestro común, evento ocurrido probablemente hace ~270 millones de años.

Los cambios sucedidos en los agrupamientos α y β son más recientes, como se verá a partir de la descripción de la divergencia de cada gen en la siguiente sección.

6.4

La divergencia secuencial es la base del reloj evolutivo

Conceptos principales

- Las secuencias de genes homólogos en especies diferentes varían en los sitios de remplazo (donde las mutaciones causan sustituciones de aminoácidos) y en los silenciosos (donde una mutación no afecta a la secuencia proteínica).
- Las mutaciones se acumulan en los sitios silenciosos ~10 veces más rápido que en los de remplazo.
- La divergencia evolutiva entre dos proteínas es medida por el porcentaje de posiciones en que difieren los aminoácidos correspondientes.
- Las mutaciones se acumulan más o menos a una misma velocidad después de que los genes se separan, de manera que la divergencia entre cualquier par de secuencias de globina es proporcional al tiempo transcurrido desde la separación de sus genes.

La mayor parte de los cambios de las secuencias proteínicas se debe a mutaciones pequeñas que se

acumulan lentamente con el tiempo. Las mutaciones puntuales y las inserciones y deleciones pequeñas son aleatorias, quizá más o menos con la misma probabilidad en todas las regiones del genoma. Las excepciones son los puntos calientes, en los que las mutaciones suceden con mucha mayor frecuencia. Casi todas las mutaciones que modifican la secuencia de aminoácidos son deletéreas y serán eliminadas por selección natural.

Pocas mutaciones son provechosas, y cuando ocurre una inusual, es muy probable que se extienda entre la población, reemplazando de modo eventual a la secuencia anterior. Cuando una nueva variante reemplaza a la versión previa del gen, se dice que se ha *fijado* en la población.

Un aspecto controvertido es qué proporción de mutaciones de una secuencia de aminoácidos es **neutral**, es decir, sin efecto en la función de las proteínas y, por lo tanto, susceptible de acumularse como resultado del **flujo aleatorio** y de la **fijación** .

La velocidad a la cual se acumulan los cambios derivados de mutaciones es característica de cada proteína, y presumiblemente depende, por lo menos en parte, de su flexibilidad respecto del cambio. En una especie, una proteína evoluciona por sustitución de mutaciones, seguida de delección o de fijación en el acervo de reproducción individual. Cabe recordar que cuando se escudriña el acervo genético de una especie, se ven sólo las variantes que han sobrevivido. Cuando múltiples variantes están presentes, puede que sean estables (porque ninguna tiene una ventaja selectiva), o de hecho, una podría ser transitoria porque está en proceso de ser desplazada.

Cuando una especie se divide en dos nuevas especies, cada una de las resultantes constituye un acervo independiente para la evolución. Al compararse las proteínas correspondientes de dos especies, se ven las diferencias acumuladas *desde el momento en que sus ancestros dejaron de hibridarse* . Algunas están muy conservadas y muestran pocos cambios, o ninguno, de especie a especie, lo cual indica que casi cualquier cambio es deletéreo y, por lo tanto, no seleccionado.

La diferencia entre dos proteínas se expresa como su **divergencia**, el porcentaje de posiciones en que difieren los aminoácidos. La divergencia entre proteínas puede ser diferente de la divergencia entre las secuencias de ácido nucleico correspondientes, debido a la representación de cada aminoácido en un codón formado por tres bases, en el cual, a menudo la tercera no incide en el significado.

La secuencia nucleotídica de una región codificadora puede dividirse en **sitios de remplazo** y **sitios silenciosos** potenciales: