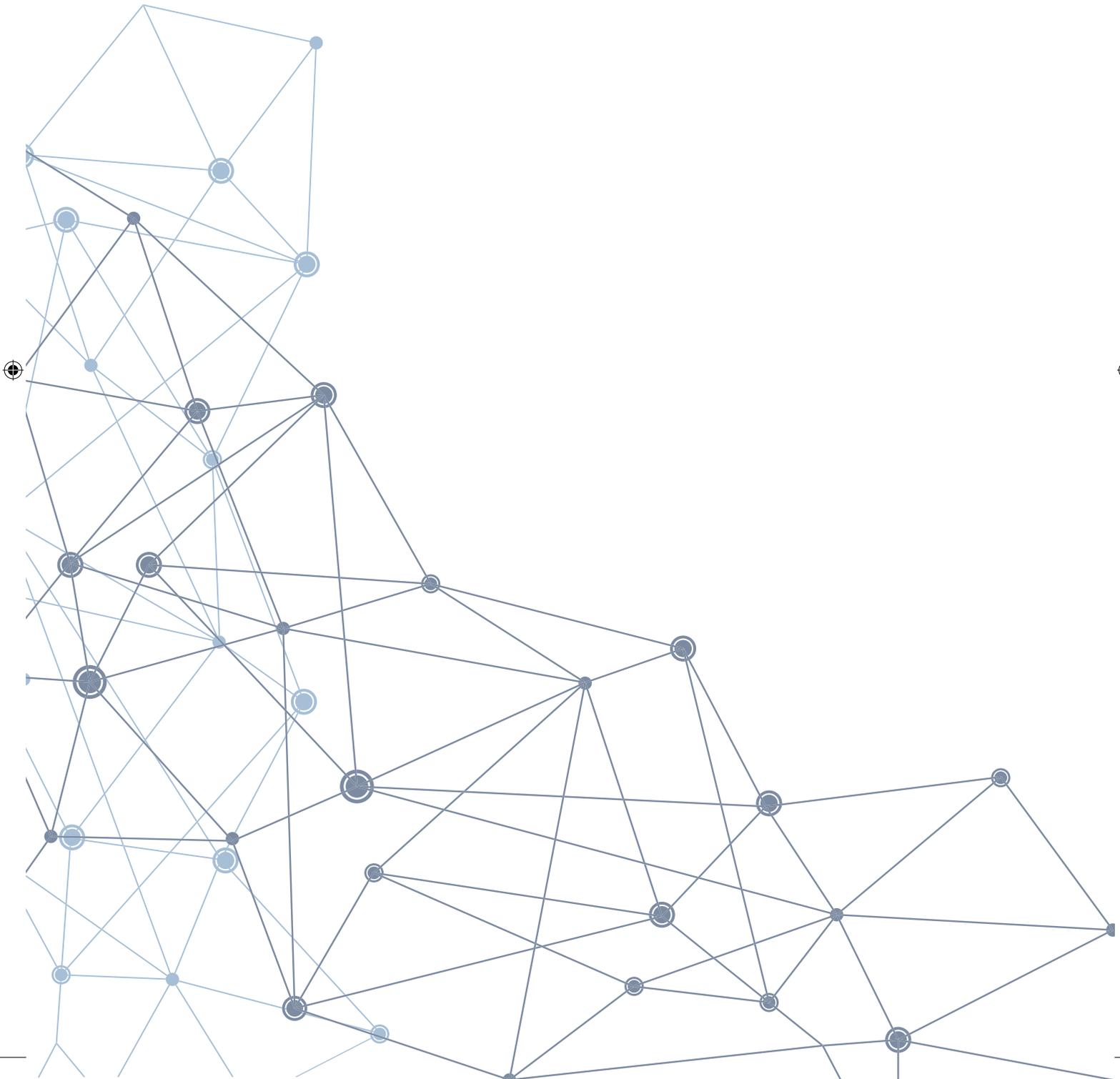


Parte 6

Evolução do Gene



25

Evolução de Genes e Genomas

Diogo Meyer e Tatiana Teixeira Torres

O objetivo deste capítulo é oferecer uma introdução à genômica evolutiva, área da evolução que busca compreender como processos evolutivos moldaram a variação genética observada dentro de espécies e entre elas. Após uma revisão de conceitos básicos de genética das populações, mostra como inferir se a seleção natural atuou sobre uma região do genoma. A seleção que atua sobre um gene também pode moldar a variação em regiões vizinhas, o que demonstra a importância de analisar a evolução do genoma como um todo. Alguns exemplos de seleção sobre fenótipos morfológicos e fisiológicos são apresentados. Em particular, discutem-se estudos que envolvem um importante fenótipo molecular, a expressão gênica, e são apresentadas interpretações sobre seleção em perfis de expressão gênica. Por fim, discute-se a importância de estudar evolução sob uma ótica “não adaptativa”, isto é, sem o pressuposto de que a maior parte das mudanças evolutivas resulta da seleção natural.

Introdução

Uma maneira de investigar o funcionamento dos seres vivos é realizar experimentos para entender como as moléculas atuam, o modo como interagem umas com as outras e as consequências desses processos. Uma segunda abordagem – distinta, mas não mutuamente exclusiva – consiste em analisar o funcionamento dos seres vivos com um olhar histórico, perguntando como os traços vistos hoje se originaram ao longo do tempo. Várias perguntas desse tipo podem ser formuladas. Por exemplo, a seleção natural é o principal processo capaz de explicar as mudanças evolutivas que ocorrem ao longo do tempo? As mudanças que resultam em alterações na morfologia envolvem alterações em proteínas ou no modo como elas são reguladas? O que torna alguns genes tão diferentes entre espécies e populações próximas, enquanto outros são mantidos conservados? Quais genes foram importantes no processo de adaptação das espécies ao seu ambiente? Responder a questões como essas requer olhar para a diversidade genética com uma perspectiva evolutiva, refletindo sobre a história dos genes e das espécies, e testar hipóteses sobre os processos evolutivos que nelas operaram. A genética evolutiva, ao longo das últimas décadas, consolidou-se como uma área de pesquisa com forte elemento quantitativo, capaz de nortear a realização de testes estatísticos, que possibilitam responder a questões complexas.

Este capítulo apresenta a lógica do pensamento evolutivo que possibilita responder às questões acima. Para tanto, é necessário discutir sobre conceitos de genética populacional, biologia molecular e genética. Também será mostrado que os grandes bancos de dados, descrevendo variação em escala genômica, mostraram-se cruciais para tornar possível uma compreensão dos processos evolutivos que moldam os genomas.

Quais processos microevolutivos moldam a diversidade genética?

Mutação, demografia e seleção

Um geneticista evolutivo busca compreender como a composição genética das populações muda ao longo do tempo. Nesse sentido, uma perspectiva bastante reducionista, porém bastante poderosa, consiste em contemplar a variação genética em um único *locus* e perguntar: como esse *locus* varia dentro da espécie que estou estudando? Como ele varia entre espécies? Essa simples descrição da variabilidade está na base das análises evolutivas e possibilita a classificação de um *locus* nas seguintes categorias:

- Polimórfico: se ele varia dentro de uma espécie, ou seja, existem diferenças entre indivíduos
- Monomórfico (ou fixo): quando o *locus* não apresenta variação e é idêntico em indivíduos da mesma espécie
- Divergente: quando o *locus* difere entre duas espécies, não sendo polimórfico em nenhuma.

Todos esses padrões são encontrados nos genomas. Por exemplo, estudos na nossa espécie mostram que, a cada 800 nucleotídeos sequenciados em um indivíduo, um deles é variável.¹ Quando se fazem comparações entre o genoma humano e o do chimpanzé, nota-se que cerca de 1/100 dos nucleotídeos são divergentes.²

Diversos processos ajudam a explicar a origem dos padrões de polimorfismo e divergência. Em um nível, há os processos genéticos, como a mutação e a duplicação gênica, que produzem variação genética, mas a variação populacional também é moldada por processos demográficos, como a deriva genética e a migração. A deriva genética (Figura 25.1) é um processo de mudança na composição genética de uma população que

ocorre na ausência de seleção natural e resulta do fato de cada nova geração ser o resultado de uma espécie de “sorteio” de gametas, que são transmitidos para os filhos (ver boxe “Deriva genética e teoria neutra”).

Esse sorteio introduz mudanças nas frequências alélicas. Já a migração é capaz de introduzir variação nova em uma população pelo influxo de imigrantes, trazendo variantes genéticas que não estavam presentes. É importante notar que, por meio da deriva e da migração, as populações mudam sua composição genética – isto é, evoluem –, mesmo que não ocorra a seleção natural.

A composição genética da população também pode ser moldada pela seleção natural, na medida em que as probabilidades de um alelo aumentar ou diminuir de frequência dependem do modo como ele contribui para as chances de sobrevivência e reprodução dos indivíduos que o carregam. Por exemplo, mutações fortemente deletérias são geralmente removidas da população, pois os indivíduos que as apresentam têm menor chance de sobreviver e, portanto, de passá-la adiante. Refere-se à seleção que remove mutações deletérias como seleção negativa ou purificadora. Já as mutações vantajosas tendem a se tornar mais comuns, pois sua presença aumenta as chances de sobrevivência e de reprodução de seu portador. A seleção mais comum que torna uma mutação vantajosa é a seleção positiva. Por outro lado, as mutações que não aumentam nem diminuem as chances de sobrevivência e reprodução de seus portadores terão a sua dinâmica evolutiva governada apenas pelos processos demográficos. Essas são as chamadas mutações neutras. A ocorrência de uma mutação

neutra em um gene não significa que ele não seja importante: significa apenas que aquela mutação não alterou as chances de sobrevivência do indivíduo em relação às dos indivíduos com a forma não mutante. A Figura 25.2 resume os processos microevolutivos que moldam a variação genética em populações.

Entre evolucionistas, não há dúvida de que algumas mudanças genéticas são consequência da seleção natural e outras ocorrem por deriva genética e outros processos demográficos. Mas o quão importante é o papel da seleção natural no processo de mudança de populações ao longo do tempo? Como se comportam populações quando experimentam, ao mesmo tempo, os efeitos tanto da deriva quanto da seleção?

Interação entre deriva genética e seleção natural

As contribuições de mutações, os eventos demográficos e a seleção natural precisam ser considerados de modo conjunto, a fim de compreender como as populações evoluem ao longo do tempo. Estudos teóricos foram capazes de definir de modo preciso como esses processos interagem e, a seguir, serão apresentados de modo qualitativo os principais achados.

Há uma espécie de cabo de guerra entre a natureza aleatória da deriva genética e a ação da seleção natural. A deriva genética resulta em mudanças na composição genética da população entre gerações, e essas mudanças ocorrem para todos os alelos, que são, essencialmente, sorteados no processo de formação e união dos gametas, supondo que a maior parte dos genes não influencia o desempenho do gameta. Porém, além de ser sorteado, um alelo também pode ser selecionado, na medida em que influencia a capacidade de sobrevivência de seu portador. Como esses dois processos – deriva genética e seleção natural – interagem? O padrão geral que emerge é o seguinte:

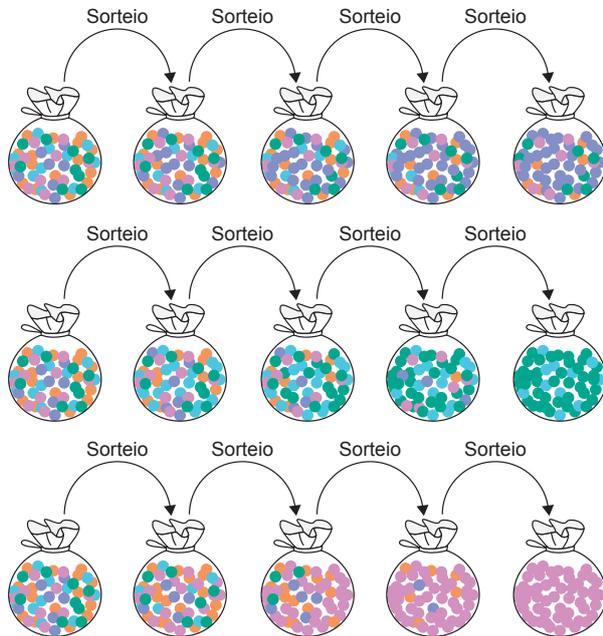


Figura 25.1 Esquema ilustrando a deriva genética. Cada bolinha representa uma cópia gênica, e o conjunto total de bolinhas em um saco corresponde à população. As cores representam diferentes alelos desse gene. Mesmo sem seleção natural, a transmissão das bolas de uma geração para outra envolve um sorteio, que, em organismos vivos, corresponde à formação dos gametas e à definição de quais gametas originaram o novo organismo. Assim, a composição genética da população muda ao longo do tempo. Reparar que, nas três instâncias do processo de deriva, diferentes alelos se tornaram comuns.

Deriva genética e teoria neutra

Deriva genética

Como é possível que a composição genética de uma população mude de uma geração para outra, mesmo sem seleção natural? Na década de 1930, Sewall Wright (1889-1988) introduziu o conceito de deriva genética e desenvolveu suas bases matemáticas. Wright calculou a probabilidade de a frequência alélica mudar simplesmente em virtude do “sorteio” de alelos que produz os gametas. Uma analogia ajuda a compreender esse processo: quando se faz uma pesquisa eleitoral, há uma margem de erro, que resulta no fato de ter uma amostra reduzida da grande população de eleitores. A formação de uma nova geração é também uma amostragem: mesmo que o conjunto dos milhões de gametas disponíveis carreguem os alelos “A” e “a” em frequências de 50% cada, apenas alguns desses gametas irão de fato originar organismos. Assim, a frequência dos alelos “A” e “a” oscila em relação à original, e essa oscilação será maior quando a amostra for menor. Consequentemente, quanto menor a população, maior será a mudança de frequências alélicas de uma geração para outra, causada por deriva genética.

Teoria neutra

O quão importante é a deriva genética para a mudança evolutiva? Na década de 1960, o geneticista Motoo Kimura desenvolveu a teoria neutra da evolução molecular, de acordo com a qual uma grande proporção das mudanças evolutivas que ocorrem em populações naturais resulta da deriva. Segundo Kimura, uma grande quantidade de mutações está constantemente surgindo em populações, e elas são de dois tipos principais: as deletérias, que são removidas pela seleção natural, e as neutras, que não contribuem nem prejudicam o seu portador. O que definirá o destino de uma mutação neutra será a deriva genética. Muitas delas serão perdidas, mas algumas aumentarão de frequência, por meio do processo de “sorteio” que caracteriza a deriva, até se fixarem (alcançarem frequências de 100%) na população. Nesse caso, diz-se que, por meio da deriva, ocorreu uma “substituição” naquela espécie. Para Kimura, a maior parte das substituições que distinguem duas espécies surgiu por deriva, e não porque eram vantajosas.

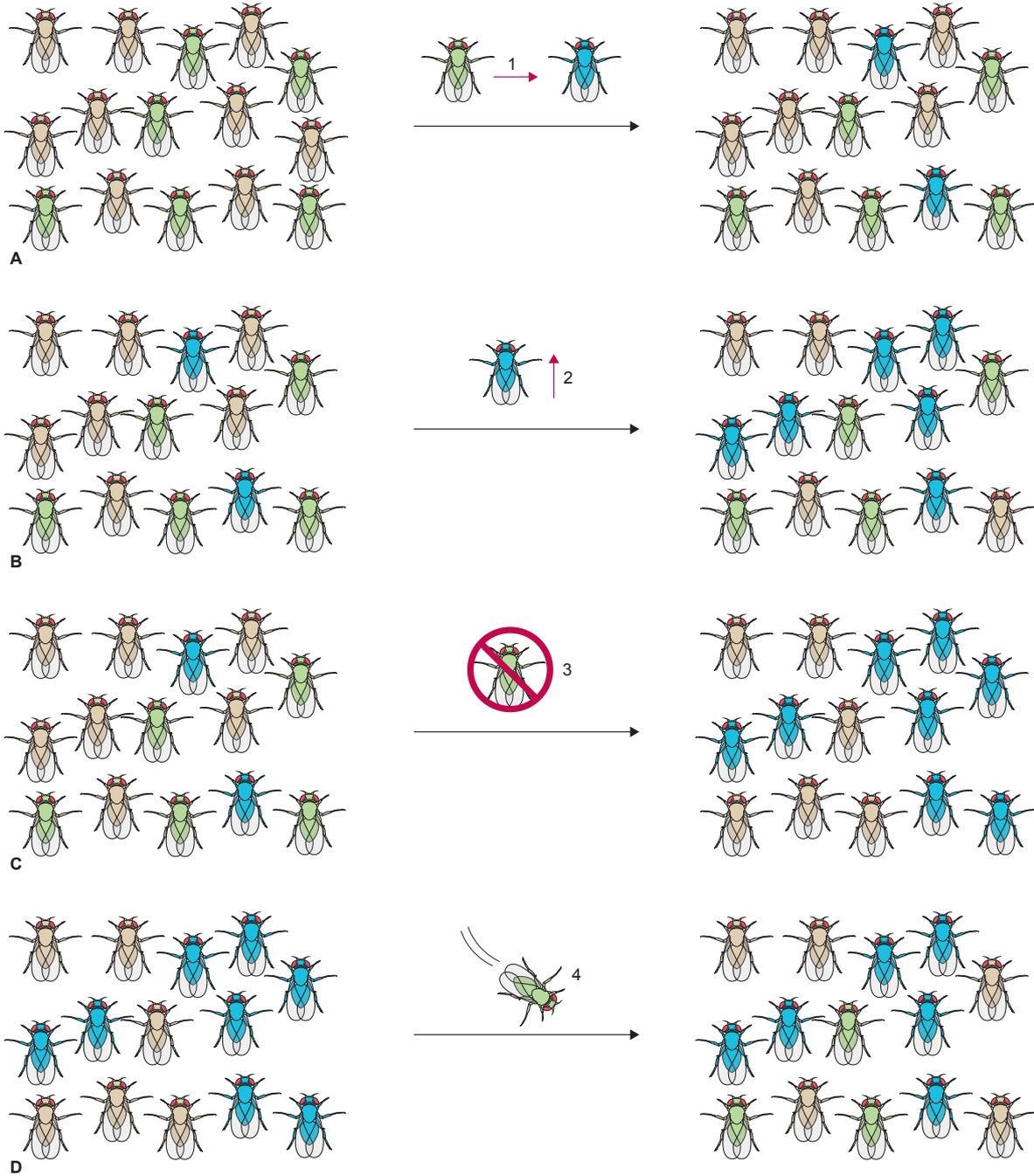


Figura 25.2 Esquema de processos microevolutivos que atuam sobre populações. **A.** Uma população sofre mutações (1) que criam nova variação, modificando a nova população. **B.** Uma população pode mudar sua constituição por deriva genética (2). Por exemplo, os indivíduos com pigmentação azul foram transmitidos mais frequentemente e aumentaram de frequência. **C.** Em outra população, os indivíduos com a pigmentação verde podem sofrer seleção natural (3) e são eliminados da população, privando-a de indivíduos com esse traço. **D.** A composição da população pode ser alterada pelo influxo de indivíduos, ou seja, houve migração (4), levando a um aumento dos indivíduos representados pela pigmentação verde.

- Quando há muita deriva genética (p. ex., em uma população com tamanho reduzido), mesmo mutações que não são neutras poderão ter mudanças entre gerações predominantemente influenciadas pela deriva genética, e não pela seleção. Assim, uma mutação deletéria pode se tornar comum ou uma mutação vantajosa pode ser perdida (Figura 25.3)
- Quando há pouca deriva genética (p. ex., no caso de espécies com grandes tamanhos populacionais), as frequências alélicas mudam pouco entre gerações, como consequência do sorteio gamético (afinal, como muitos gametas estão sendo sorteados, as sucessivas gerações são semelhantes entre si). Isso implica que mudanças nas chances de sobrevivência e reprodução, conferidas por mutações, podem ter mais influência sobre as chances de um alelo se tornar comum ou ser eliminado.

Em síntese, nas populações em que há muita deriva genética, a seleção torna-se uma força mais fraca, enquanto, em populações maiores – com menos deriva genética –, a seleção torna-se mais importante e é capaz de criar mudanças. Fica claro que a chance de uma mutação se tornar comum ou ser eliminada não depende apenas do seu efeito sobre o fenótipo, mas do tamanho da população em que ela ocorre.

Busca por evidências de seleção natural

As sessões anteriores apresentaram os elementos teóricos do pensamento evolutivo. Esta demonstra como investigar, na prática, se os genes tiveram sua trajetória evolutiva moldada por processos seletivos e quais os tipos de seleção em que atuaram.

Como demonstrado a seguir, a abordagem de um evolucionista é complementar à de um biólogo molecular, que está interessado em questões funcionais. Em vez de medir empiricamente a importância de um gene para um fenótipo, o evolucionista tira um retrato da variação observada e se pergunta: a origem da variação observada depende da ação da seleção? Ou poderia surgir pelo acúmulo de mutações em um cenário neutro, em que a seleção não favorece mutações vantajosas?

Testes baseados na razão dN/dS

Uma das maneiras de fazer inferências sobre a ação de seleção natural consiste em comparar sequências de duas espécies diferentes (isto é, examinar a sua divergência). A análise de regiões que codificam proteínas possibilita classificar diferenças entre as sequências de DNA de duas espécies em sinônimas e não sinônimas. Mutações sinônimas são aquelas que resultam de uma alteração de nucleotídeo em um códon que não altera o aminoácido especificado. Mutações nucleotídicas não sinônimas alteram o códon de um modo que resulta em mudança no aminoácido (Figura 25.4). É razoável supor que as mutações sinônimas são neutras, pois não alteram a proteína, sendo observadas apenas na sequência de DNA. As não sinônimas são alvos da seleção natural, uma vez que mudanças na sequência de uma proteína frequentemente alteram o fenótipo do organismo e, consequentemente, suas chances de sobrevivência e reprodução. Tipicamente, o número de substituições não sinônimas por sítio (taxa não sinônima, dN) é menor que o número de substituições sinônimas por sítio (taxa sinônima, dS), pois a seleção natural remove mutações não sinônimas da população. Isso é esperado, uma vez que a maior parte das mudanças introduzidas em uma proteína funcional tende a perturbar sua função, diminuindo as chances de sobrevivência do organismo. A Figura 25.4 explica como taxas dN/dS são estimadas.

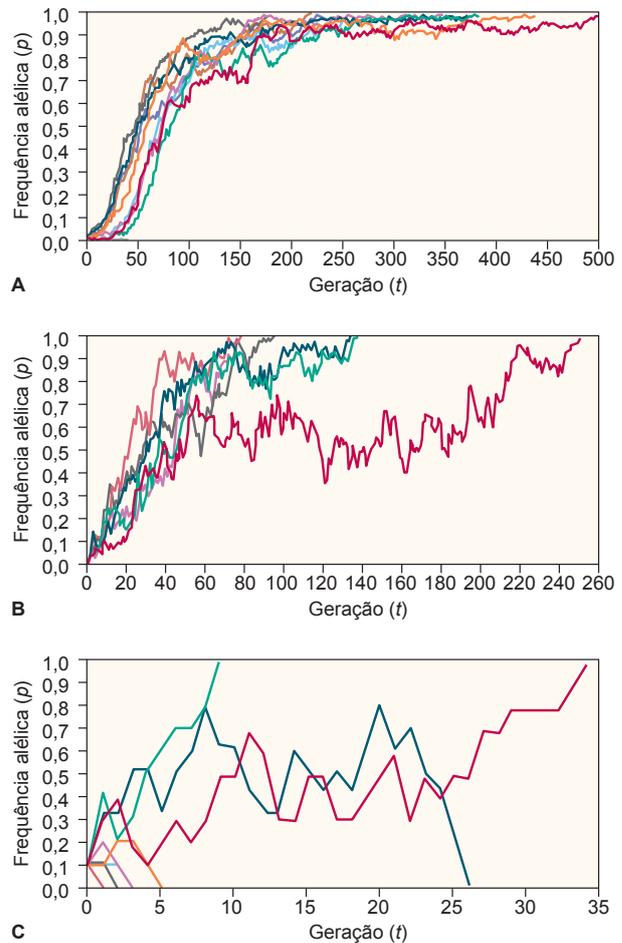


Figura 25.3 Interação da deriva genética e da seleção natural. Simulações em computador possibilitam recriar o processo evolutivo. Em cada painel, é usada uma combinação de valores para o tamanho populacional N , o coeficiente de seleção s e a frequência inicial do alelo sob seleção p . As linhas indicam como a frequência do alelo muda ao longo do tempo, e cada cor representa uma simulação diferente. **A.** Seleção sobre um alelo vantajoso em uma população grande. Uma população grande ($N = 500$) tem um alelo inicialmente raro ($p = 0,01$) favorecido por seleção natural (quem tem o alelo vantajoso tem uma probabilidade 10% maior de sobreviver e se reproduzir). Entre as dez simulações, em nove delas o alelo vantajoso se fixou, e em apenas uma ele foi perdido. **B.** Seleção sobre um alelo vantajoso em uma população pequena. A intensidade da seleção e a frequência inicial do alelo permanecem iguais às do cenário anterior, mas o tamanho populacional é apenas $N = 50$. Nesse cenário houve cinco casos em que o alelo vantajoso foi perdido (as simulações em que a frequência cai para 0). **C.** Seleção contra uma mutação fracamente deletéria em uma população pequena. As simulações mostram o que acontece com uma mutação inicialmente presente em uma única cópia e que é prejudicial ao seu portador (tanto homocigotos quanto heterocigotos com a mutação têm chances de sobrevivência e reprodução reduzidas em 1% em relação a quem não apresenta a mutação). A população simulada é muito pequena ($N = 5$) e a mutação foi fixada em dois casos. Em vários casos, ela demorou algumas gerações até ser eliminada, ilustrando que, em uma população pequena, mesmo mutações deletérias podem contribuir para a variação. Se essa mutação tivesse surgido em uma população grande (p. ex., $N = 500$), ela teria sido sempre removida, geralmente em poucas gerações.

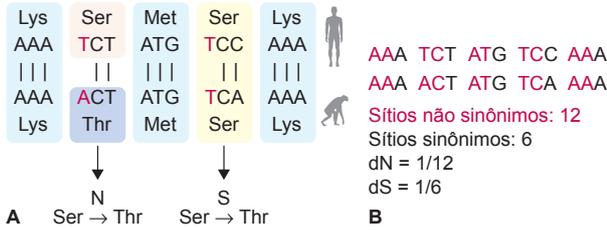


Figura 25.4 Exemplo simplificado de como são estimadas as taxas de substituição sinônimas e não sinônimas. **A.** Os sítios divergentes entre espécies são classificados como sinônimos (S) e não sinônimos (N). **B.** O número de sítios sinônimos e não sinônimos corresponde à fração de sítios que, quando mutados, originariam mudanças sinônimas e não sinônimas, respectivamente. Uma simplificação razoável é supor que mudanças nas primeiras duas posições do códon resultam em mudanças não sinônimas, enquanto mudanças na terceira posição resultam em mudanças sinônimas, refletindo a natureza do código genético. Assim, há mais sítios “disponíveis” para mudanças não sinônimas. Levando esse fator em conta, nota-se que há 12 sítios não sinônimos e seis sítios sinônimos. Podemos então estimar as taxas de substituição por sítio, evidenciando uma taxa sinônima maior.

Se não houvesse nenhuma seleção sobre as mutações não sinônimas, dN seria igual a dS (pois parte-se do pressuposto de que as mudanças sinônimas são neutras), resultando em uma razão de $dN/dS = 1$. Quando a seleção natural elimina mudanças não sinônimas, pois geralmente são deletérias, o valor de dN cai, e a razão dN/dS assume valores baixos. Por exemplo, comparando humanos e chimpanzés para cerca de 15 mil genes, Kosiol *et al.*³ encontraram um dN/dS médio de 0,25. Isso significa que, do total de mudanças não sinônimas que surgiram, apenas 25% eram neutras (ou seja, não sofreram ação da seleção natural negativa), e as 75% restantes foram removidas pela seleção natural. Comparando os genomas de ratos e camundongos, foi encontrado um $dN/dS = 0,12$, o que indica que as mutações não sinônimas são removidas pela seleção natural de modo mais frequente entre essas duas espécies do que entre humanos e chimpanzés. Uma explicação para a diferença observada no valor de dN/dS de primatas e roedores reside no tamanho populacional dos últimos, que é muito maior do que o de primatas. Isso resulta em menor taxa de variação por deriva, o que leva a maior eficiência da seleção natural, que remove as mutações deletérias. Sob essa perspectiva, a divergência de primatas e humanos carrega uma fração considerável de diferenças não sinônimas, mas que não foram removidas pela seleção natural porque há muita deriva na história dessas linhagens.

A razão de dN/dS também pode ser informativa sobre seleção positiva. Quando a seleção natural atua sobre um gene, favorecendo sucessivas mudanças não sinônimas, a taxa não sinônima pode superar a sinônima (resultando em $dN/dS > 1$). Nesse caso, há evidência de que a seleção positiva está atuando sobre o gene em questão. Em um estudo envolvendo genomas de oito espécies de mamíferos, Kosiol *et al.*³ identificaram genes com $dN/dS > 1$ e investigaram quais são as suas funções. Genes com funções ligadas à imunidade e defesa (como os genes para citocinas e envolvidos na ativação do sistema complemento, genes envolvidos na imunidade mediada por linfócitos, entre outros) eram a classe que mais comumente apresentava evidência de seleção positiva, seguidos por genes envolvidos na percepção sensorial (como os genes de receptores olfatórios) e genes que atuam nos processos ligados à reprodução

e fertilização. Essas são classes para as quais se espera que a seleção positiva seja de fato importante, pois envolvem proteínas que estão sob seleção, vindo de um fator ambiental que está constantemente mudando (patógenos, estímulos ambientais, receptividade reprodutiva, respectivamente). A seleção está constantemente favorecendo novas mudanças nas proteínas por ele codificadas, elevando a taxa não sinônima.

A análise de razões de dN/dS tem sido imensamente influente na genética evolutiva. Ela tem sido frequentemente usada para documentar a ação da seleção natural sobre proteínas de superfície de vírus como o da *influenza*, que evoluem sob seleção natural, favorecendo mudanças que possibilitam que eles burlam o reconhecimento pelo sistema imune.⁴

Comparação da variação dentro e entre espécies

Uma fonte adicional de informação sobre processos evolutivos vem da comparação de variação genética em dois níveis: aquela que existe dentro das espécies (o polimorfismo) e aquela entre espécies (a divergência). Uma predição da teoria neutra é que a variação que existe dentro de espécies é proporcional à observada entre espécies. A lógica é simples: mudanças que são deletérias em uma população são eliminadas pela seleção natural imediatamente, de modo que não são vistas nem entre espécies nem nas populações. Segundo essa lógica, a variação que se vê dentro de uma espécie corresponde à classe de mutações neutras, e são essas também que se tornam diferenças entre espécies. Como testar a hipótese de que a evolução de fato ocorre dessa maneira?

Assim como as diferenças entre espécies, a variação genética que existe dentro de nossa espécie pode ser classificada como sinônima ou não sinônima. Analisando sequências de todos os exons humanos, Bustamante *et al.*⁵ notaram que, em 48% dos sítios polimórficos, as diferenças eram não sinônimas. Entretanto, comparando o mesmo conjunto de exons entre humanos e chimpanzés, viram que só 37% das diferenças eram não sinônimas, ou seja, as diferenças genéticas entre dois humanos têm maior chance de ser do tipo não sinônimo que as diferenças entre um humano e um chimpanzé. Esse padrão pode ser explicado ao partir do pressuposto de que muitas das mutações não sinônimas são fracamente deletérias. Mutações fracamente deletérias não são tão prejudiciais a ponto de serem instantaneamente removidas da população, mas também não são neutras: elas diminuem as chances de sobrevivência e reprodução de seus portadores o suficiente para que, em grandes escalas de tempo (como o tempo de divergência entre o homem e o chimpanzé), eventualmente sejam removidas. Consequentemente, observam-se mais mudanças não sinônimas fracamente deletérias dentro das espécies que na comparação entre elas.

Esse padrão de variação genética tem sido confirmado em estudos subsequentes, como o feito pelo projeto dos 1.000 genomas.¹ Esse trabalho evidenciou que a espécie humana carrega muitas mutações não sinônimas (incluindo muitas fracamente deletérias) e documentou também que essas mutações ocorrem em baixa frequência. Isso significa que é muito provável que cada indivíduo carregue um conjunto de mutações fracamente deletérias praticamente exclusivo dele. Esse padrão tem consequências importantes para a compreensão da base genética das doenças humanas, pois é possível que muitas mutações diferentes estejam associadas a fenótipos relativamente comuns, trazendo um desafio para estudos que buscam localizar mutações partilhadas por muitos indivíduos como causa de doenças humanas.

Ao investigar simultaneamente divergência e polimorfismo dentro e entre espécies filogeneticamente próximas, podem-se identificar processos que moldaram a variação ao longo do genoma. Analisando o genoma completo de seis linhagens endocruzadas de *Drosophila simulans* em um contexto populacional e comparando com os genomas de duas espécies próximas, *D. melanogaster* e *D. yakuba*, Begun *et al.*⁶ observaram que, como esperado sob o modelo de evolução neutra, o polimorfismo encontrado dentro de *D. simulans* tinha alta correlação com a divergência em relação às outras espécies (ou seja, genes que variam mais dentro da espécie também diferem mais na comparação entre espécies distintas). No entanto, algumas regiões apresentavam um padrão distinto. Nelas, o polimorfismo parecia ser proporcionalmente menor do que a divergência (Figura 25.5). Ao analisar as funções biológicas dessas regiões com uma grande redução do polimorfismo, os pesquisadores notaram uma alta representação de genes envolvidos em reprodução e imunidade, evidenciando a importância dos processos de seleção sexual e interação parasita-hospedeiro na evolução do genoma de *D. simulans*. Uma das regiões que apresentaram redução significativa do polimorfismo contém os genes *scpr-A*, *scpr-B* e *scpr-C*. Esses três genes apresentam alta expressão em tecidos germinativos masculinos e estão, provavelmente, relacionados à reprodução. Os autores concluíram que o polimorfismo reduzido nessas regiões é, provavelmente, resultado da ocorrência de mutações benéficas, que aumentaram de frequência por meio de seleção positiva recente e recorrente em *D. simulans*. Outra observação interessante nessas análises é que as regiões vizinhas às mutações selecionadas positivamente também apresentam polimorfismo reduzido. Essa observação será discutida a seguir.

Busca por evidências de seleção em genomas usando variação em populações

Genes muito diferenciados

A seleção natural também pode deixar assinaturas na variação genética *entre* populações. Por exemplo, a seleção positiva aumenta a frequência do alelo na população em que ele é vantajoso. Caso aquele mesmo alelo não seja vantajoso para outras populações, a diferença da frequência alélica entre as

populações será aumentada pela seleção natural, em relação ao esperado na ausência de pressões seletivas. Consequentemente, a busca por genes com frequências alélicas muito diferentes entre populações constitui uma estratégia para encontrar genes sob seleção positiva.⁷

Um exemplo recente ilustra a aplicação dessa abordagem. Um conjunto de sequências de exons humanos (abrangendo 92% dos genes conhecidos) foi gerado para 50 tibetanos, e a frequência dos polimorfismos foi comparada com a de chineses. Os tibetanos são geneticamente muito semelhantes aos chineses; entretanto, há um gene específico que apresenta uma mutação que alcança uma frequência de quase 90% em tibetanos e tem frequência de apenas 10% em chineses. Diferenças de frequência alélica dessa magnitude são raríssimas na comparação de populações humanas e sugerem que a mutação foi, de algum modo, vantajosa para os tibetanos. Nesse caso, há bons argumentos mecanísticos para explicar o achado: o gene em questão é a proteína EPAS1, um fator de transcrição que regula a resposta à hipoxia. Aparentemente, a seleção favoreceu mudanças nesse gene na população tibetana, que habita elevadas altitudes e tem o grande desafio de sobreviver em um ambiente com baixas concentrações de oxigênio (Figura 25.6).⁸

O uso da diferenciação genética para encontrar regiões do genoma sob seleção tornou-se uma ferramenta valiosa para análises evolutivas em várias espécies. Em um estudo para compreender as consequências genéticas da seleção artificial, Akey *et al.*⁹ analisaram o genoma de várias raças de cachorro, buscando regiões altamente diferenciadas. Entre os achados estavam genes que regulam a cor e a textura da pelagem, a morfologia do esqueleto e o tamanho. A forte seleção artificial, exercida por criadores, resultou em diferenças marcantes justamente nos genes que contribuem para os fenótipos escolhidos. Tais estudos fazem uma aproximação entre a nossa compreensão de evolução vista pela ótica populacional – a mudança de frequências alélicas em populações – e a evolução morfológica.

Genes com pouca variação dentro de populações

Ao mesmo tempo em que leva a um aumento na diferenciação entre populações, a seleção positiva também pode diminuir a variação dentro das populações. Quando um alelo vantajoso se torna muito frequente na população, ele resulta em uma

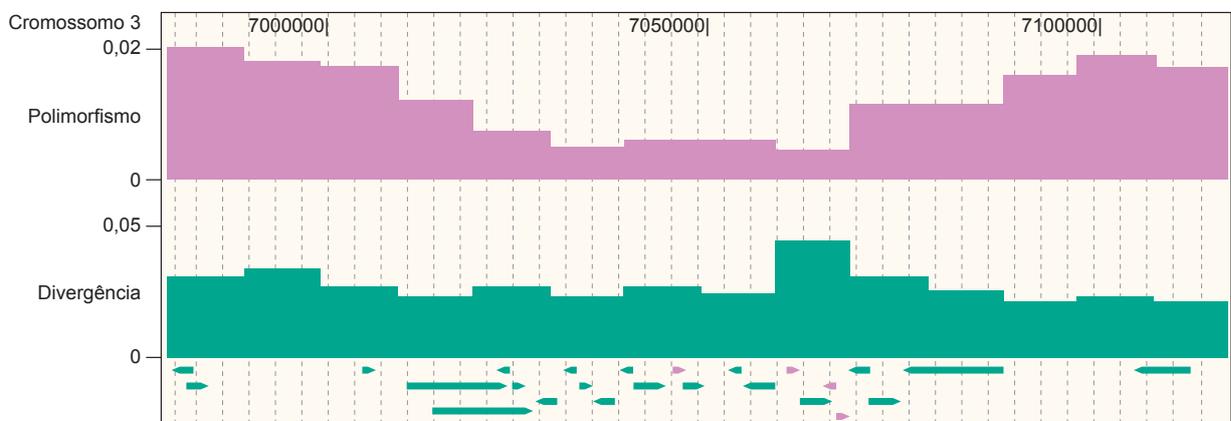


Figura 25.5 Região do cromossomo 3 de *D. simulans* mostrando a redução do polimorfismo em relação à divergência. O polimorfismo (verde) e a divergência (rosa) foram determinados em janelas não sobrepostas de 10 mil pares de base. A região com polimorfismo reduzido é rica em genes, e três deles estão destacados em vermelho, os genes *scpr-A*, *scpr-B* e *scpr-C*.

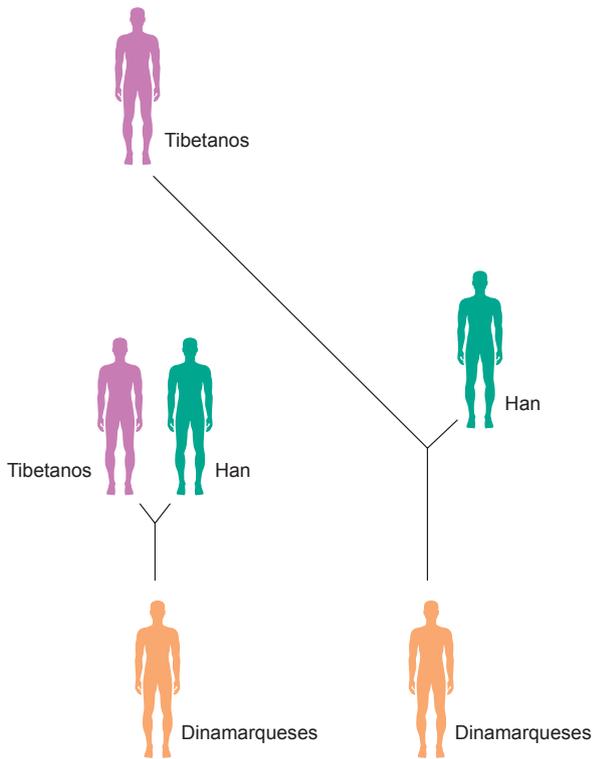


Figura 25.6 Árvore menor: divergência média, ao longo da porção codificadora do genoma, entre populações tibetanas, chinesas (Han) e dinamarquesas. Árvore maior: mesma medida de diferenciação, porém para o gene *EPAS1*. O tamanho dos ramos representa a divergência observada entre as populações. Na árvore menor, a divergência média pode ser tomada como uma medida de evolução neutra e é proporcional ao tempo de divergência entre as populações. Adaptada de Yi *et al.*, 2008.⁸

redução na variação genética para aquele gene (afinal, quase todos os indivíduos, por terem sido selecionados, apresentarão a mesma variante). Essa perda de variação não afeta apenas o *locus* selecionado, mas toda a região do genoma ao redor dele. A taxa de recombinação entre os *loci* depende da distância deles no cromossomo. Assim, o indivíduo que herda o alelo vantajoso recebe também todo o segmento cromossômico com o qual estava ligado, reduzindo, dessa maneira, a variação nessa região (Figura 25.7).

Dessa maneira, a busca por regiões de baixa variação populacional constitui outra abordagem para detectar eventos de seleção positiva recente. Um exemplo marcante vem do estudo do gene da lactase, a enzima que é responsável pela digestão do açúcar presente no leite, a lactose. O gene da lactase (*LCT*) e toda a região ao redor dele apresentam-se virtualmente desprovidos de variação em populações do norte da Europa. Já em populações de outras regiões do mundo, há considerável variação no gene e nas regiões ao seu redor.¹⁰ Há uma explicação baseada em seleção natural para esse padrão: evidências arqueológicas indicam que o gado foi domesticado originalmente no norte da Europa e usado para produzir leite para o consumo humano. Dessa maneira, teria sido nessa região que a seleção natural teria atuado, favorecendo as mutações que possibilitam a digestão do leite na idade adulta, o que representa uma fonte adicional de nutrientes e eletrólitos.

Consequências da seleção natural sobre regiões vizinhas

A discussão anterior mostrou que há uma redução da variação genética no gene que é alvo da seleção natural e que essa redução também pode se prolongar às regiões vizinhas. Mas será que esse padrão é comum e ocorre em diversas regiões do genoma?

Uma importante evidência de que esse fenômeno é comum vem da observação de que, em duas espécies extensivamente estudadas – humanos e *Drosophila* –, há correlação positiva entre a taxa de recombinação e a diversidade genética. A Figura 25.8 mostra que há mais variação em genes localizados nas regiões de maior recombinação, e menos nas de menor recombinação. Por quê? Para responder a essa pergunta, considere os efeitos da seleção positiva atuando em um gene: quanto mais forte for a ligação física do gene com as regiões vizinhas, mais a perda de variação no gene (causada por seleção) resultará em perda de variação nas regiões ligadas. No entanto, se há alta recombinação ao redor do gene selecionado, o alelo favorecido aumentará de frequência, mas não necessariamente carregará consigo uma região única, pois a recombinação embaralhará as associações entre o gene selecionado e as regiões vizinhas.¹¹

Há uma segunda interpretação para esse padrão, que difere ao invocar outro tipo de seleção. Charlesworth *et al.*¹³ propuseram que o tipo de seleção mais comum no genoma não é o positivo, mas o purificador, que remove mutações deletérias. Supondo que haja uma constante “chuva” de mutações deletérias em nossos genomas, é razoável imaginar que as alterações nos genes por ela atingidos resultam em indivíduos menos aptos, que não as passam para gerações subsequentes. Se o gene afetado por mutações deletérias estiver em uma região de baixa recombinação, a remoção da região deletéria pela seleção natural também resultará na perda de variação nessa região. Entretanto, se a recombinação for alta, as regiões vizinhas poderão ser transmitidas mesmo que o alelo deletério não o seja.

Essa interpretação é consistente com uma das observações que emergiu da análise feita pelo projeto 1.000 genomas, que observou que regiões do genoma humano fisicamente próximas a genes são menos variáveis do que as mais distantes (Figura 25.9). Isso resultaria da seleção negativa em regiões gênicas, a qual, ao remover mutações deletérias, causa perda de variação em regiões vizinhas.

Independentemente de qual interpretação está correta (aquela baseada em seleção positiva ou negativa), ambas as explicações partilham um ponto importante em comum: a ação da seleção natural explica a variação no nível de polimorfismo ao longo do genoma. A variação ao longo do genoma depende não só das funções daquela região, mas da vizinhança em que ela se encontra. Essa visão integrada do genoma, em que padrões de variação dependem dos efeitos da seleção natural sobre regiões vizinhas, é uma importante contribuição das análises na era genômica.

Expressão gênica como fenótipo | Bases genéticas e evolução

Foi visto anteriormente como os padrões de polimorfismo e divergência nas sequências de DNA podem ser moldados por processos demográficos e seletivos. Mas qual o tipo de variação genética subjacente à evolução fenotípica?

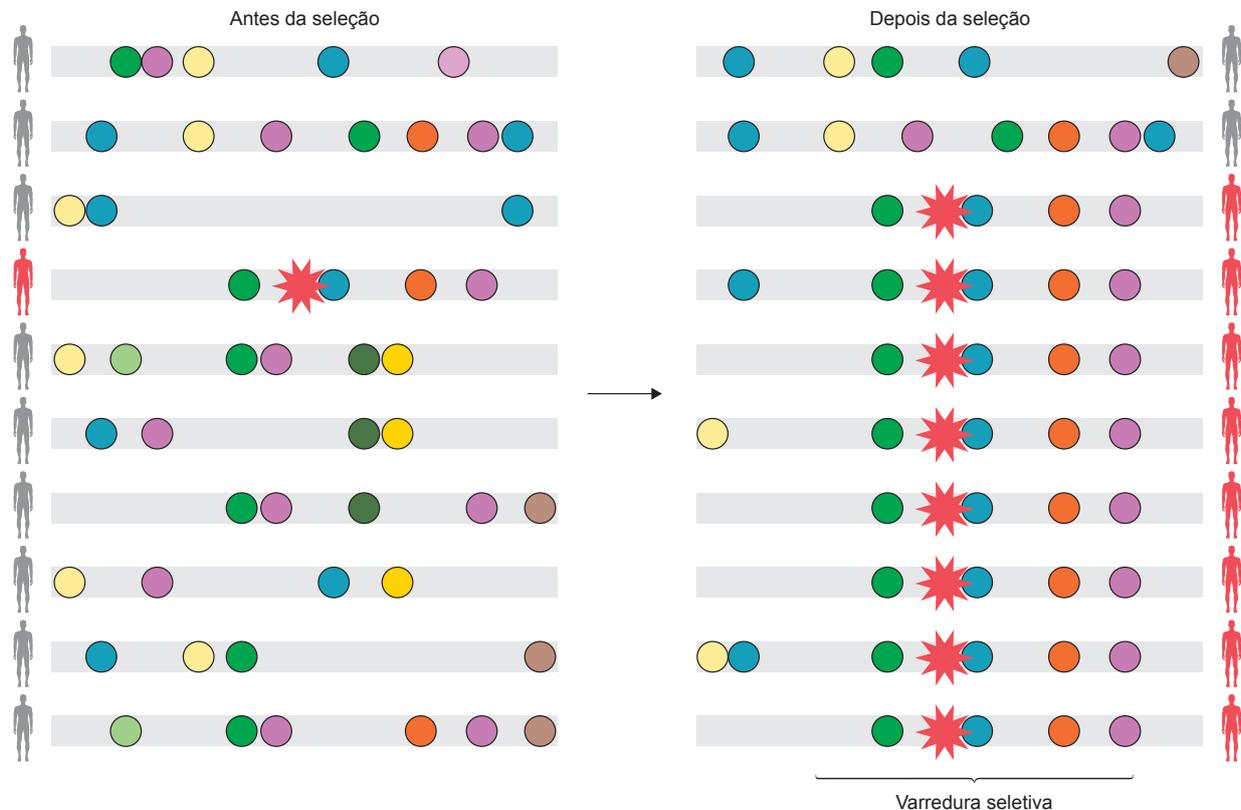


Figura 25.7 A seleção positiva altera a frequência de alelos em uma população. Nesse caso, o alelo representado pela estrela vermelha era vantajoso ao seu portador, de modo que aumentou de frequência em gerações subsequentes. Entretanto, o aumento de frequência desse alelo arrastou com ele variantes genéticas que estavam fisicamente ligadas (simplesmente porque, a não ser que haja recombinação, o aumento de frequência de um alelo provoca o aumento dos outros alelos que estão no mesmo cromossomo). A consequência é que, ao redor do *locus* favorecido, há uma região de homogeneidade genética.

Algumas mutações alteram a sequência de aminoácidos das proteínas e podem alterar sua função, enquanto mutações em sequências regulatórias podem alterar o padrão de expressão das proteínas. Como visto no Capítulo 4, mutações que alteram a sequência de proteínas são prontamente identificadas a partir da sequência de DNA, devido à simplicidade e universalidade do código genético.

Porém, se as variações funcionais em proteínas fossem a principal causa da variação fenotípica entre as espécies, como explicar a grande diversidade morfológica, fisiológica e comportamental encontrada em espécies filogeneticamente próximas com regiões codificadoras tão conservadas? Essa mesma pergunta foi feita há mais de 35 anos por Allan Charles Wilson e sua então aluna de doutorado, Mary-Claire King. Ao comparar proteínas isoladas de chimpanzés (*Pan troglodytes*) e de humanos (*Homo sapiens*), King e Wilson¹⁴ demonstraram que a divergência observada entre as duas espécies era mínima (hoje se sabe que a sequência de aminoácidos é idêntica em 99% de sua extensão). Uma diferenciação tão baixa entre as proteínas seria insuficiente para explicar a extensa diferença morfofisiológica entre as espécies, sugerindo que grande parte da diversidade teria sua origem em mudanças na expressão gênica, e não em variações na função gênica.¹⁴ O paradoxo da conservação de estrutura e função de genes diante da diversidade morfológica ficou ainda mais evidente

com a surpreendente descoberta de que a maioria dos animais compartilhavam várias famílias de genes envolvidos no estabelecimento do padrão corporal. Se os animais compartilham genes com função e estrutura tão parecidas, como explicar a emergência de padrões tão diferentes em sua morfologia, fisiologia e comportamento?

Evolução da expressão gênica | Bases moleculares

Vários fatores, como quando, onde e quanto um gene é expresso, são tão importantes quanto a própria função bioquímica do produto gênico final e, por isso, alterações no padrão de expressão podem causar importantes diferenças fenotípicas. Como visto no Capítulo 6, o controle dos níveis de expressão de cada gene ocorre por meio da ligação de proteínas (fatores de transcrição, ativadores e inibidores) a motivos específicos na sua região regulatória. Basicamente, a evolução da expressão gênica ocorre de duas maneiras distintas:

- Podem ocorrer mudanças em regiões regulatórias em *trans*, como quando há mutação no próprio fator de transcrição, alterando sua ligação ao motivo específico
- Podem ocorrer mudanças regulatórias em *cis*, como quando a mutação ocorre no próprio motivo regulatório, destruindo ou criando novos sítios de reconhecimento (Figura 25.10).

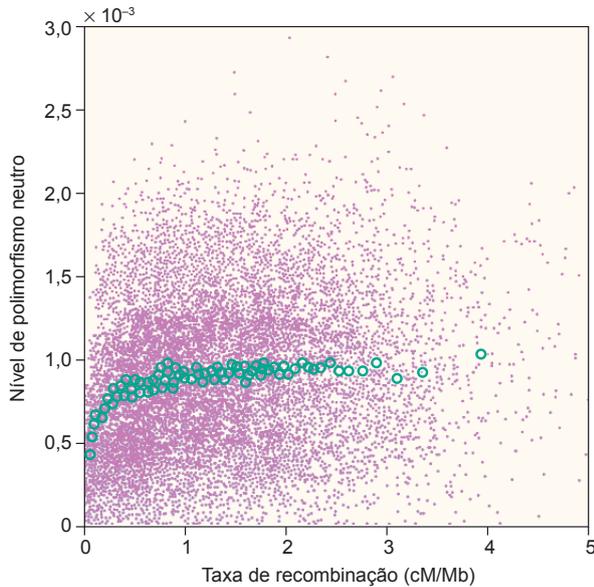


Figura 25.8 No genoma humano há um aumento da variação genética (eixo Y) associado ao aumento da taxa de recombinação. Esse padrão empírico já foi documentado em *Drosophila* na década de 1990 e, nas duas espécies (*Drosophila* e ser humano), pode ser interpretado como consequência dos efeitos da seleção (positiva ou negativa) que reduzem a variação ao seu redor. Adaptada de Cai *et al.*, 2009.¹²

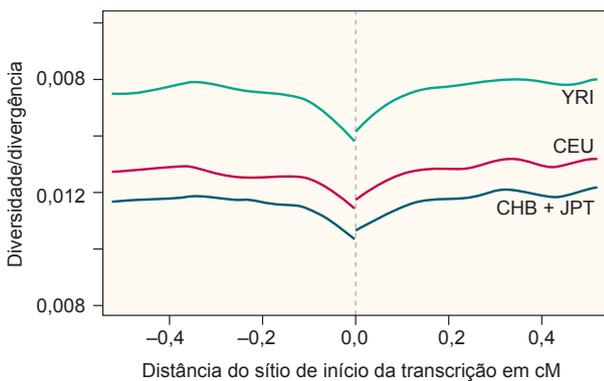


Figura 25.9 Quanto mais distantes do início ou do fim da transcrição de um gene, maior a variação genética. Esse achado mostra que estar perto de um gene provoca redução da variação daquela região, provavelmente como consequência dos efeitos da seleção natural negativa, que remove a variação deletéria nas regiões codificadoras e, consequentemente, nas regiões vizinhas.¹

Nos últimos anos, a contribuição relativa de cada um dos tipos de mutação regulatória na evolução da expressão gênica vem sendo analisada em escala genômica. De modo geral, fatores de transcrição de espécies filogeneticamente próximas apresentam pouca variação em seus níveis de expressão e em suas regiões codificadoras, o que sugere que a maioria das mudanças nos níveis de expressão globais de espécies próximas resulta de mudanças em *cis*, e não em *trans*.

As regiões regulatórias em *cis* são sequências curtas (de 6 a 20 nucleotídeos), encontradas na vizinhança de regiões codificadoras, onde se ligam fatores de transcrição que regulam a produção de RNA mensageiro (mRNA) e, consequentemente, a produção de proteínas.

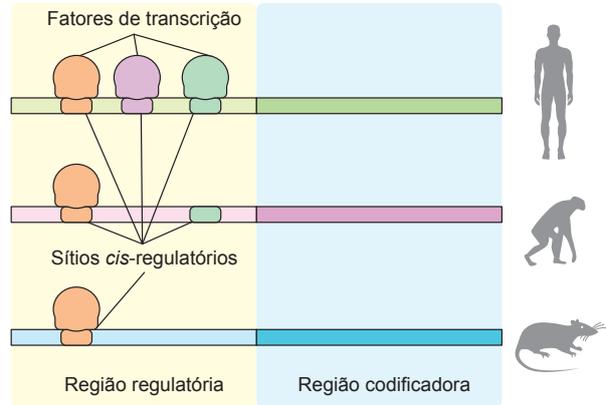


Figura 25.10 Organização dos módulos *cis*-regulatórios em diferentes espécies. Mutações na sequência do motivo regulatório podem provocar a destruição do sítio de ligação de fatores de transcrição, alterando os níveis de expressão do gene. A presença do fator de transcrição também é essencial para a manutenção dos níveis de expressão. Mutações em *trans* no gene que codifica o próprio fator de transcrição podem provocar uma alteração em sua função e impossibilitar sua ligação no sítio *cis*-regulatório.

Outra distinção importante entre variações em regiões codificadoras e regulatórias é que a primeira delas, em geral, altera a função do produto gênico, enquanto a segunda altera seu padrão espacial e temporal de expressão. Como muitos genes têm uma ação pleiotrópica, ou seja, podem atuar em múltiplos fenótipos, a alteração de sua função por meio de mutações na região codificadora pode afetar simultaneamente várias vias metabólicas, podendo incluir funções vitais. Esse efeito pleiotrópico é minimizado quando a mutação ocorre em regiões *cis*-regulatórias, que atuam como interruptores, ligando e desligando genes de modo contexto-dependente. Cada gene pode conter vários módulos *cis*-regulatórios, e essa modularidade possibilita que um mesmo gene seja regulado por diferentes fatores de transcrição (Figura 25.10). Assim, uma mutação em um módulo pode passar despercebida em um tecido e causar grandes alterações em outros tecidos do mesmo organismo. Para ilustrar essa ideia, está descrito a seguir o caso da evolução do espinho pélvico em esgana-gata.

O esgana-gata, *Gasterosteus aculeatus*, é um peixe pequeno que habita lagos de água doce, águas salobras e o litoral marinho, com ampla distribuição no norte e centro europeu, no norte da Ásia e na América do Norte. Ao longo dos últimos 10 a 20 mil anos, diferentes populações do *G. aculeatus*, isoladas em diversos lagos, perderam os espinhos pélvicos. Cruzando indivíduos de *G. aculeatus* de duas populações com distintas morfologias na região pélvica, Shapiro *et al.*¹⁵ mapearam a região genômica envolvida na diferença morfológica. O locus de maior efeito explicava metade da variação nas estruturas pélvicas e correspondia à sequência regulatória do gene fator de transcrição 1 homeodomínio tipo pareado (*pitx1*, do inglês, *paired-like homeodomain transcription factor 1*). O curioso é que, apesar de esse gene ter papel importante no desenvolvimento embrionário de vertebrados, a mutação identificada em *G. aculeatus* afetava apenas o desenvolvimento da região pélvica. A explicação para esse fato é que a perda do espinho foi o resultado de deleções de um módulo *cis*-regulatório que controla a expressão de *pitx1* apenas na região pélvica no alelino (Figura 25.11).

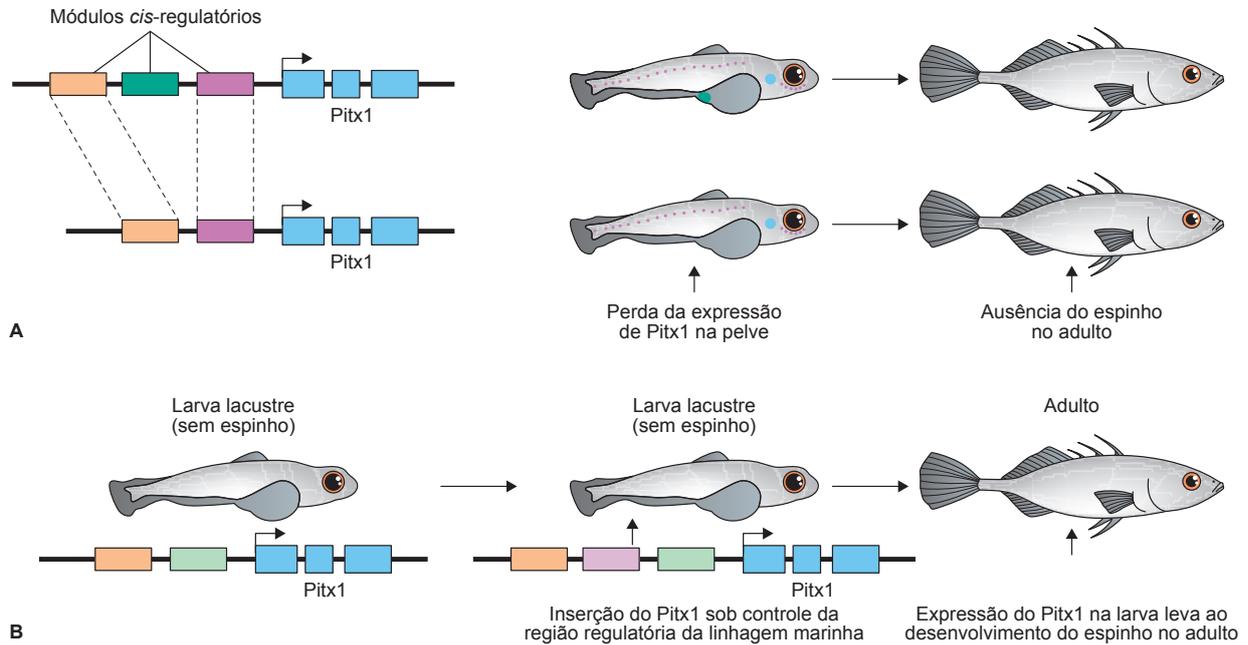


Figura 25.11 Evolução do espinho pélvico em esgana-gata, *G. aculeatus*. A redução ou perda do espinho pélvico é resultado de mutações que levaram a perdas recorrentes de um módulo *cis*-regulatório que controla a expressão do gene *pitx1* (A). Sem o elemento regulatório, não há expressão de *pitx1* na pele e, conseqüentemente, não há desenvolvimento do espinho. A introdução do transgene com a região regulatória ancestral (sem a deleção do módulo regulatório) na linhagem sem espinhos restaura a expressão de *pitx1* e o desenvolvimento do espinho pélvico (B).

Ao introduzir o gene *pitx1* com sua região regulatória ancestral no alevino de *G. aculeatus* da linhagem sem espinho (Figura 25.11), a expressão do gene foi restaurada na pele, levando à formação do espinho no peixe adulto. Essa observação comprova a função da região *cis*-regulatória na regulação espacial da expressão do gene e é um bom exemplo de que a presença de vários sítios regulatórios possibilita a flexibilidade na expressão do gene em um tecido (nesse caso, na pele) sem afetar suas demais funções (nesse caso, as relevantes à osteogênese).¹⁶

O caso da evolução do espinho pélvico do esgana-gata também demonstra como os achados da genética podem ser conectados com os de história natural e biogeografia. O espinho está presente em todas as populações marinhas de *G. aculeatus*, enquanto, nas populações lacustres, que ocuparam os lagos apenas nos últimos 10 mil anos, houve redução ou perda completa do espinho. Esse cenário sugere que a redução do espinho pode ser uma resposta adaptativa ao novo ambiente. Em ambientes marinhos, o esgana-gata é predado por peixes maiores, e o espinho oferece um modo de defesa diante de predadores, enquanto, nos lagos, os principais predadores são os insetos, que agarram o esgana-gata pelo espinho.

Análises genéticas também ajudam a explicar o surpreendente fato de os eventos de perda do espinho terem ocorrido independentemente em diferentes lagos. O módulo regulatório está localizado em uma região repetitiva, próxima ao telômero, o que explica as deleções recorrentes que ocorreram nessa região instável do genoma de *G. aculeatus* em diferentes populações. Nesse caso, a estrutura genômica da região em que se localiza o gene favorece o processo que origina a variação genética que é favorecida pela seleção natural.

Evolução da expressão gênica | Seleção

A expressão diferencial pode estar envolvida no aparecimento de fenótipos com diferentes valores adaptativos. Como visto na seção anterior, uma variante do gene da lactase foi selecionada em populações europeias. A lactase normalmente deixa de ser produzida após o desmame, e uma mutação que ocorreu no norte da Europa teria possibilitado que os humanos usassem esse valioso recurso alimentar. Hoje se sabe, especificamente, qual é a mutação vantajosa: uma única alteração de C para T em uma posição localizada mais de 10 mil bases antes do gene da lactase produz essa mudança fenotípica. Essa mudança ocorre em uma região fora do gene da lactase, mas que é capaz de modular sua expressão. A seguir são apresentados os tipos de variações (como essa no gene da lactase) que podem levar a uma alteração no nível de expressão do produto gênico, as chamadas variações regulatórias.

Outro caso bem documentado de seleção natural vem do estudo da evolução de diferentes espécies de tentilhões nas ilhas Galápagos. Os tentilhões constituem um grupo de diversas espécies de pássaros de pequeno porte que serviram como inspiração para Charles Darwin. Essas espécies são muito parecidas em tamanho, forma e coloração, mas duas características mostram uma variação substancial entre as diferentes espécies: a forma e o tamanho dos bicos (Figura 25.12).

A diversidade de forma dos bicos reflete diretamente a variedade de alimentos que essas aves podem explorar, desde sementes a pequenos insetos. Por exemplo, no gênero *Geospiza*, de modo geral, aves com bico pequeno alimentam-se de sementes pequenas. Aves com bico grande, por sua vez, alimentam-se de sementes maiores. Essa relação é observada tanto em indivíduos da mesma espécie quanto de diferentes espécies (Figura 25.12). Abzhanov *et al.*^{18,19} mostraram que a

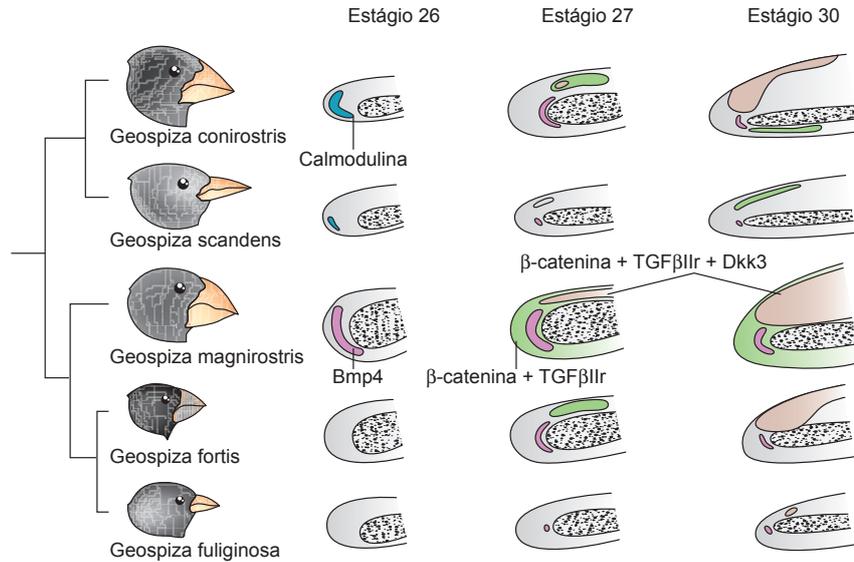


Figura 25.12 As diferentes morfologias do bico em *Geospiza* são resultantes de variações na expressão espacial e temporal de diferentes genes.¹⁷ Espécies com maior altura do bico, como *G. magnirostris* e *G. fortis*, apresentam expressão precoce e espacialmente distribuída de *Bmp4*, *TGFβIIr*, β -catenina e *Dkk3*, enquanto, em espécies com bico mais longo, como *G. conirostris* e *G. scandens*, a expressão de calmodulina, *TGFβIIr*, β -catenina e *Dkk3* é localizada na porção distal do bico.

alteração no nível de expressão de duas proteínas – proteína morfogenética óssea 4 (*bmp4*, do inglês, *bone morphogenetic protein 4*) e calmodulina – levava, respectivamente, a uma alteração na forma e no tamanho do bico de algumas espécies (Figura 25.12). Em seis espécies de tentilhões do gênero *Geospiza*, a proteína *bmp4* é expressa em níveis mais altos nos bicos mais robustos (com altura maior). Já a calmodulina é expressa em altos níveis em bicos longos. A variação na expressão desses genes tem efeito direto na diversidade morfológica do bico, fornecendo o material bruto para seleção natural, como descrito a seguir.

Peter Grant, Rosemary Grant *et al.*^{20,21} estudaram diversas espécies de tentilhões em várias ilhas de Galápagos por mais de 30 anos. Eles notaram que o tamanho do bico está diretamente relacionado com o tamanho das sementes que o animal utiliza na alimentação.

Para entender a evolução do tamanho do bico por seleção natural, será abordado um exemplo na espécie *Geospiza fortis*, um tentilhão de solo que se alimenta principalmente de sementes e utiliza o bico para quebrá-las. A partir de 1973, os pesquisadores iniciaram o monitoramento de vários indivíduos de *G. fortis*, registrando, para cada indivíduo, peso, comprimento da asa e da cauda e três medidas do bico: altura, largura e comprimento. Em 1977, um período de seca causou o desaparecimento de 84% de *G. fortis* da ilha Daphne Maior. Após a seca, os tipos de planta disponíveis para alimentação mudou drasticamente e frutos mais duros e maiores passaram a ser a principal fonte de alimentação. Após a seca, tentilhões com bico mais forte exploravam esse alimento com maior sucesso, e as aves sobreviventes tinham, em média, bico com altura maior em comparação com o das aves do período anterior à seca (Figura 25.13). A evolução dos bicos foi acompanhada durante os anos subsequentes, mostrando como a seleção sobre uma característica diretamente influenciada por variações na expressão gênica moldou a população de *G. fortis*

(Figura 25.13). Esses estudos ilustram a integração de estudos de história natural e inferência de seleção com estudo da base genética da variação morfológica, justamente um grande desafio para a biologia.

Outros exemplos de variação adaptativa em módulos *cis*-regulatórios já foram observados, envolvendo desde genes de pigmentação em *D. melanogaster* e em camundongos a genes que participam da resposta a inseticidas em *D. melanogaster* e outros insetos.

Apesar de enfatizar variações adaptativas em módulos *cis*-regulatórios, a noção geral de evolução regulatória diz respeito não só ao controle da transcrição, mas de outras etapas do processo, que culmina com a produção de uma proteína. Entre as etapas do processo que influenciam a abundância de proteínas ativas estão a regulação pós-transcricional, o controle do processamento, distribuição e degradação do mRNA e a regulação pós-traducional. Todas as variações que influenciam essas etapas são variações regulatórias. Existem aspectos ainda pouco esclarecidos sobre a evolução desses mecanismos, mas as abordagens que utilizam dados genômicos vêm possibilitando identificar regiões que desempenham papéis regulatórios importantes, como discutido a seguir.

O pouco que se sabe de padrões gerais de evolução de regiões não codificadoras vem, em grande parte, de estudos em *Drosophila*. Comparações de regiões do genoma de espécies de *Drosophila* indicam que a maior parte das regiões não codificadoras já conhecidas, como regiões intergênicas, introns e regiões transcritas não traduzidas (UTR), apresenta índices de polimorfismo e divergência inferiores àqueles de sítios sinônimos em regiões codificadoras (Figura 25.14).^{22,23} Essa observação sugere que mutações em regiões não codificadoras têm um efeito deletério e, como consequência, são frequentemente eliminadas pela seleção natural, resultando nos baixos níveis de polimorfismo. No entanto, algumas dessas regiões não codificadoras apresentam excesso de divergência em

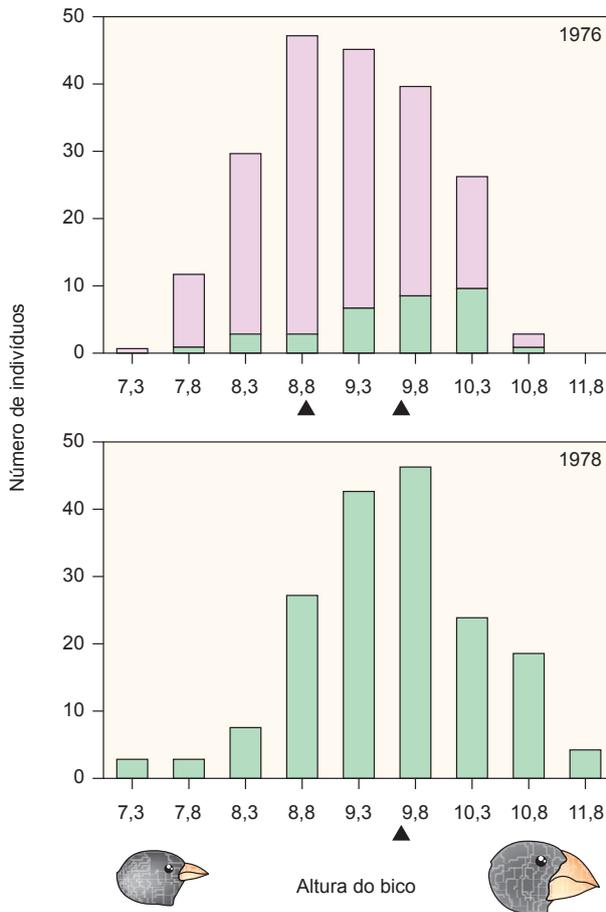


Figura 25.13 Evolução do bico na população de *G. fortis* na ilha Daphne Maior.²¹ O gráfico de cima mostra a distribuição na altura dos bicos na população em 1976 (rosa), com os sobreviventes da seca que se reproduziram em 1978 (verde). O gráfico de baixo mostra a distribuição da altura dos bicos dos tentilhões adultos nascidos em 1978. A evolução no tamanho dos bicos entre as gerações pode ser expressa pela diferença das médias em 1976 (antes da seleção) e em 1978 (após a seleção).

comparação com os níveis de polimorfismo, o que pode ser interpretado como assinatura de seleção positiva, como visto anteriormente (ver Figura 25.5). Essas observações reforçam a ideia de que regiões não codificadoras, que constituem uma imensa proporção dos genomas eucarióticos, são importantes alvos tanto da seleção natural negativa quanto da positiva.

Interação entre genes

Os exemplos vistos anteriormente ilustram muito bem os efeitos das variações regulatórias na divergência morfológica. No entanto, um ponto importante a considerar é que a grande maioria dos fenótipos complexos não é diretamente relacionada a um único gene ou via metabólica. Não se pode considerar a morfologia, a fisiologia ou o comportamento de um organismo como simples resultado da presença ou ausência de genes e proteínas. Para entender a complexidade do sistema e por que os padrões observados emergem, é necessário observar como genes e proteínas interagem entre si em módulos funcionais (grupos de genes que atuam conjuntamente para a determinação de um fenótipo).

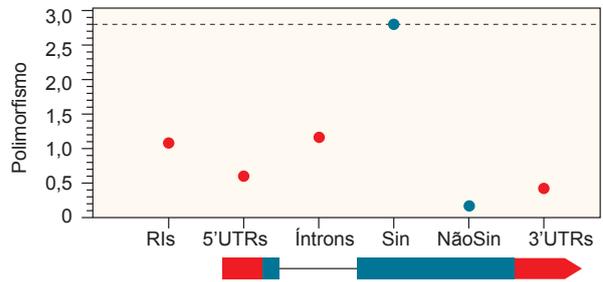


Figura 25.14 Média dos níveis de polimorfismo observados em regiões específicas do genoma de *D. melanogaster* (as barras indicam dois erros-padrão). O nível de polimorfismo esperado em *loci* evoluindo sob neutralidade é mostrado como a linha vermelha. A média do polimorfismo nas regiões observadas é menor que o esperado sob neutralidade, indicando a ação de seleção purificadora nos *loci* funcionalmente importantes. NãoSin: sítios não sinônimos; RiS: regiões intergênicas; Sin: sítios sinônimos.

Ao observar conjuntos de genes funcionalmente relacionados, é possível contextualizar as variações na expressão, observando o papel desses genes em vias metabólicas ou outros tipos de interações gênicas. Por exemplo, uma maior fração dos genes diferencialmente expressos entre cérebros de humanos e chimpanzés é envolvida no metabolismo energético e no enovelamento de proteínas.²⁴

Outro modo de olhar a complexidade da regulação gênica é por meio de redes de coregulação. Agrupamentos de genes coexpressos definem grupos de genes que são funcionalmente relacionados. Mudanças no número e na identidade dos genes em uma rede podem indicar mudanças adaptativas em módulos funcionais. A comparação do nível de expressão global em cérebro humano e de chimpanzé, por meio de redes de coregulação, possibilitou a identificação de módulos de genes coexpressos que correspondem a distintas regiões funcionais do cérebro, refletindo a organização funcional observada no cérebro.²⁵ A maior parte das diferenças entre as redes de coexpressão das duas espécies foi observada no córtex cerebral, região onde foi observado um grande crescimento na linhagem humana.²⁵

Um dos exemplos mais claros da dimensão das interações entre genes (chamadas de interações epistáticas) vem do estudo da levedura *Saccharomyces cerevisiae*. Constanzo *et al.*²⁶ valeram-se da possibilidade de introduzir mutações de perda de função em genes específicos desse organismo e examinaram as consequências fenotípicas do desligamento de pares específicos de genes. No total, 5,4 milhões de diferentes combinações de duplos mutantes foram analisadas e, em cada um desses casos, foi feita uma comparação entre a aptidão (estimada com base em taxas de crescimento e sobrevivência) dos duplos mutantes e as linhagens com apenas uma das mutações. Quando a aptidão do duplo mutante era melhor ou pior do que a prevista pela soma dos efeitos das mutações individuais, havia evidência de que os genes estavam interagindo. É importante notar que, para que a interação epistática ocorra, não é necessária uma interação física entre os produtos gênicos, basta que os dois genes estejam envolvidos na expressão de um único fenótipo. Dessa maneira, construiu-se um mapa das interações gênicas na levedura.

A análise das interações mostrou que genes envolvidos em processos biológicos similares compartilham perfis similares de interação. Esse resultado é esperado ao considerar que genes que operam conjuntamente no exercício de uma função

complexa devem interagir mais frequentemente que genes que não estão tão intimamente relacionados. A maioria dos genes interage com poucos outros, enquanto alguns poucos genes interagem com vários. Na comparação de sequências de 23 espécies diferentes de fungos do filo Ascomycota, foi verificado que o nível de conectividade do gene (o número de interações) tinha uma correlação positiva com a conservação deste. Genes com muitas interações evoluem mais lentamente que genes com poucas interações, uma evidência da ação de seleção purificadora. Com esse exemplo, fica claro que os fenótipos apresentados por todos os organismos dependem da coordenação espacial e temporal da expressão de conjuntos de genes (e de sua interação com o ambiente). Entender os processos subjacentes à evolução desses genes requer uma visão global da arquitetura das redes constituídas por eles.

Evolução da expressão gênica | Evolução neutra

Assim como as variações morfológicas, a diferença nos níveis de expressão gênica pode ser adaptativa ou neutra. Pode-se identificar a contribuição relativa de eventos demográficos e da seleção natural invocando conceitos vistos no início deste capítulo. Se as mudanças evolutivas observadas são causadas por processos estocásticos, e não por seleção natural, elas acumulam-se proporcionalmente ao tempo de divergência dos organismos. Partindo dessa lógica, Khaitovich *et al.*²⁷ compararam níveis globais de expressão gênica em tecido cerebral e hepático de primatas e em embriões de diferentes espécies de *Drosophila*.²⁸ Eles viram que a divergência de expressão entre as espécies aumenta linearmente com o tempo de divergência entre elas (Figura 25.15).

Sob neutralidade, espera-se que a ação das forças evolutivas seja igual na determinação das taxas de evolução dentro de entre espécies e entre elas. Assim, no caso da evolução da expressão gênica, espera-se que genes com maior variação nos níveis de expressão dentro de uma espécie sejam também aqueles com maior divergência entre espécies. Essa correlação entre

variação dentro de espécies e entre elas em níveis de expressão foi observada também em camundongos e peixes teleosteos, sugerindo que a maioria das diferenças no nível de expressão evolui neutramente.²⁹ Combinadas, essas informações possibilitam utilizar o modelo neutro de evolução da expressão gênica como hipótese nula para identificação de mudanças adaptativas em níveis de expressão.

A perturbação de redes ou vias de regulação pode, em vários casos, provocar uma alteração no valor adaptativo do indivíduo. No entanto, há diferença importante entre os mecanismos que deram origem à arquitetura das redes e os processos adaptativos que atuam sobre seus produtos finais. Como visto no exemplo do peixe esgana-gata, a presença ou ausência de espinhos pélvicos pode ser adaptativa, mas a evolução da arquitetura genômica, como a modularidade em regiões *cis*-regulatórias, pode ser explicada por meio de processos não adaptativos.

Evolução não adaptativa da organização genômica

Neste capítulo, viu-se que a seleção natural pode tornar comum uma mutação vantajosa. Discutiu-se também que, em populações pequenas, as mudanças podem ocorrer por deriva genética e, muitas vezes, em uma direção oposta ao previsto pela seleção natural (p. ex., mutações favoráveis podem ser perdidas ou mutações deletérias podem se tornar comuns; ver Figura 25.3). Contudo, além de estudar a dinâmica evolutiva de genes individuais, a teoria evolutiva possibilita inferências sobre características mais gerais, como o tamanho e a composição do genoma. Por exemplo, sabe-se que o tamanho do genoma varia em ordens de grandeza entre todos os seres vivos (procariotos comumente têm genomas de tamanho entre 1 e 10 mega pares de bases (Mpb), eucariotos unicelulares entre 10 e 100 Mpb e vertebrados entre 1.000 e 10.000 Mpb). Sabe-se também que introns são praticamente inexistentes nos genomas de procariotos, estão presentes, mas são raros, em eucariotos unicelulares e são abundantes em metazoários. De modo similar, elementos transponíveis só se tornaram muito abundantes nos metazoários, chegando a representar cerca de 45% do genoma da nossa espécie.³⁰

Encontrar uma teoria capaz de explicar essas características da organização genômica é um grande desafio. Ao longo dos anos foram feitas diversas propostas, muitas delas alvo de controvérsia. Algumas são baseadas no argumento de que a seleção natural favoreceu a evolução dos traços genômicos observáveis. Por exemplo, há argumentos de que o aumento do genoma tenha sido um pré-requisito para que houvesse um aumento do volume nuclear e, conseqüentemente, do volume celular e da complexidade de associações de células em tecidos.³¹ De modo similar, há argumentos de que o aumento no tamanho e na abundância de introns foi selecionado, pois a presença deles aumenta a chance de ocorrer recombinação entre dois genes, algo que pode ser vantajoso, pois “libera” o gene de associações com outros (associações que, como visto, podem levar ao aumento de frequência de uma variante deletéria, caso ela esteja associada a uma vantajosa).

Uma perspectiva radicalmente diferente consiste em ver a evolução da complexidade genômica como um processo que ocorre por deriva genética. Em vez de considerar que a seleção natural teria favorecido o ganho de complexidade, Lynch³⁰ sugere que as mudanças genômicas que levam ao aumento de complexidade são, na realidade, prejudiciais. O seu argumento é que o ganho de introns, o aumento de regiões intergênicas e a inserção de elementos transponíveis tornam o genoma

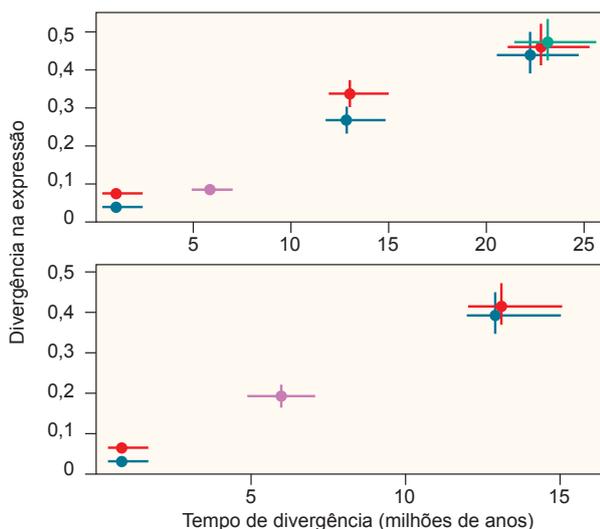


Figura 25.15 Variação na expressão gênica em primatas (homem, chimpanzé e *rhesus*). Média das diferenças em expressão gênica em cérebro (superior) e fígado (inferior) dentro de espécies de primatas e entre elas. A diferença na expressão varia linearmente com o tempo de divergência. As barras indicam os intervalos de confiança (95%).

mais suscetível a sofrer alterações que resultam em uma perda da função normal. Por exemplo, com a presença de introns, mutações que afetam o processamento do mRNA poderiam resultar em uma proteína truncada; a introdução de elementos transponíveis pode desligar genes funcionais e aumentar o potencial de novos eventos de transposição; a expansão de regiões intergênicas representa um custo energético para a célula e também aumenta a chance de ocorrer mutações regulatórias deletérias. Se o ganho de complexidade genômica é prejudicial, então por que ocorre?

Para responder a essa pergunta, o primeiro passo é investigar uma propriedade básica dos grandes grupos dos seres vivos: seu tamanho populacional. Tamanhos populacionais podem ser estimados a partir de dados sobre a diversidade genética de populações naturais, pois há uma proporcionalidade entre o tamanho da população e o tamanho populacional. Há uma grande variação entre os principais grupos de seres vivos (Figura 25.16), e tamanhos populacionais são maiores entre procariontos, que são seguidos por eucariotos uni ou oligocelulares, invertebrados, plantas e, finalmente, vertebrados, que têm os menores tamanhos populacionais. O tamanho populacional reduzido em vertebrados teria como consequência um aumento da deriva genética nesse grupo em relação aos demais. Qual seria a consequência desse aumento na deriva para a organização genômica? Mutações que causam um aumento de tamanho genômico são geralmente prejudiciais e, quando há muita deriva, elas têm maiores chances de se tornarem comuns e escaparem do crivo da seleção natural negativa. Assim, mesmo sendo eventos deletérios, as mudanças que tornam os genomas maiores ocorrem em populações de organismos que têm tamanhos populacionais menores. A Figura 25.17 ilustra como genomas maiores estão associados a tamanhos populacionais menores, o que é consistente com esse argumento.

O argumento exposto acima sustenta que o aumento da complexidade genômica foi um processo que ocorreu mesmo

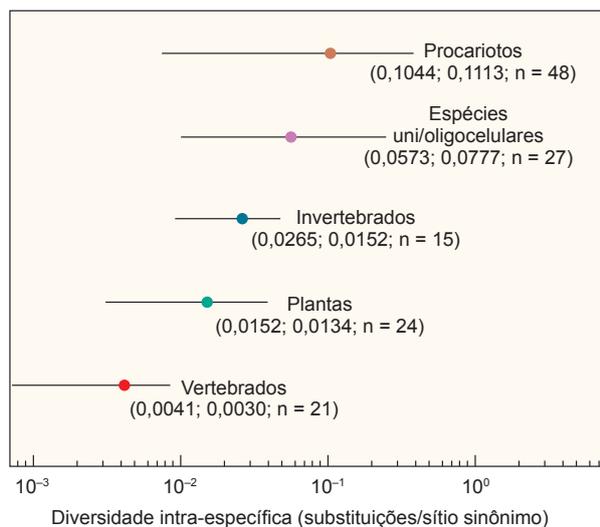


Figura 25.16 Estimativa da diversidade genética em diferentes grupos de seres vivos, com base no nível de polimorfismo intraespecífico. Os três números associados a cada grupo correspondem à média, ao desvio padrão e ao número de gêneros estudados. As linhas horizontais compreendem os 90% centrais da distribuição, reduzindo, assim, o efeito de *outliers*.

sem ser vantajoso. Entretanto, isso não exclui a possibilidade de certas mudanças genômicas, originalmente não adaptativas, terem secundariamente se tornado úteis e favorecidas pela seleção natural. Por exemplo, o ganho de introns pode ter sido originalmente prejudicial, mas, uma vez instalada uma arquitetura genômica envolvendo o processamento de mRNA, a possibilidade de produzir múltiplas isoformas a partir de um mesmo trecho genômico provavelmente representou uma vantagem. De modo similar, há casos de elementos transponíveis que foram “cooptados” para realizar funções que beneficiam seus hospedeiros, como o conhecido caso das enzimas RAG, que atuam na recombinação gênica que cria a diversidade em genes ligados a receptores de células T.

A conclusão desta seção é de que o aumento da complexidade não deve ser visto como um processo que resultou apenas da ação da seleção natural. O elemento aleatório, representado pela deriva genética, deve ter sido fundamental para possibilitar certas mudanças que, mesmo sendo originalmente prejudiciais, secundariamente se tornaram vantajosas.

Considerações finais

Ao longo deste capítulo, foram examinadas diversas ferramentas analíticas que possibilitam aprofundar o conhecimento acerca de vários aspectos em genética e evolução. É possível descobrir quais genes estão sob seleção, investigar como a seleção natural interage com outros processos – como a deriva genética e a recombinação –, estudar a importância da evolução regulatória e discutir a importância de mudanças não adaptativas para grandes mudanças genômicas. O que se prevê para os próximos anos no campo das ciências genômicas? Os avanços devem ocorrer em várias frentes. Do ponto de vista empírico, novos genomas serão sequenciados, e o pesquisador que souber

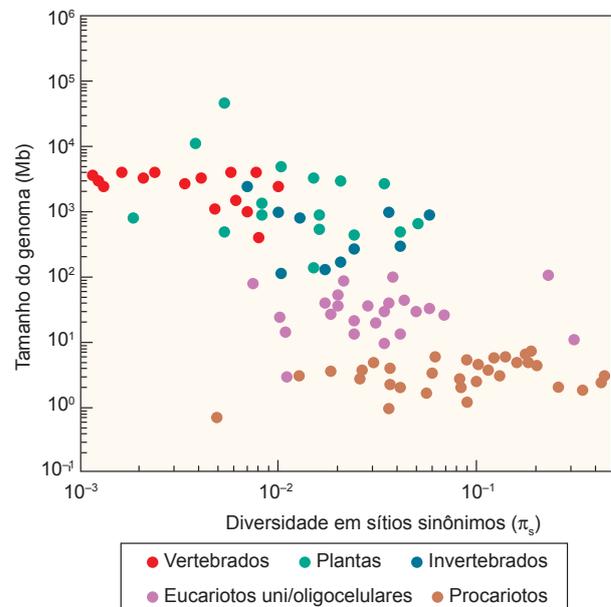


Figura 25.17 Relação entre a diversidade genética, que oferece uma estimativa do tamanho populacional, e o tamanho do genoma para vários seres vivos. Esta figura ilustra de modo claro que genomas maiores ocorrem em espécies menos diversas e, portanto, com menores tamanhos populacionais.

explorar essa rica base de dados com perguntas pertinentes estará em posição privilegiada. Uma pergunta interessante pode envolver a investigação de como determinadas moléculas funcionam (sozinhas ou em associações com outras moléculas em vias metabólicas complexas), para que se possa estudar como fenótipos complexos emergem da interação dos genomas com o ambiente, bem como a atuação da seleção sobre eles. Outra questão biológica interessante envolve o estudo das características da história das espécies sob estudo (histórias migratórias e demográficas) para suscitar testes de hipótese e contextualização apropriada dos dados genômicos. Para ter sucesso nessa investigação, o pesquisador também deve estar preparado para lidar com dados em grande escala, e isso envolve preparação em métodos estatísticos e computacionais. Compreender como se relacionam variáveis reveladas pela análise de milhares de sítios (como taxas de substituição e de polimorfismo e estimativas de recombinação, para citar um exemplo visto neste capítulo) requer o uso de métodos analíticos que levam em conta o poder que tal riqueza de dados cria. Essa tarefa não é trivial, mas indica que, cada vez mais, o pesquisador do futuro não será “de bancada” ou “de computador”, mas um investigador capaz de transitar entre o universo experimental – compreendendo as características do processo que originou os dados e as características moleculares do sistema sob estudo – e o analítico – dominando as ferramentas computacionais e estatísticas, essenciais para um projeto de pesquisa. Com essas ferramentas, será possível associar informações importantes, como o funcionamento de genes e genomas e a história demográfica das espécies sob estudo, para identificar padrões que emergem da comparação de diferentes populações e/ou espécies e inferir os processos evolutivos subjacentes aos padrões observados.

Referências bibliográficas

- Altshuler D, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, Clark AG, et al. A map of human genome variation from population-scale sequencing. *Nature*. 2010;467(7319):1061-73.
- The Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*. 2005;437(7055):69-87.
- Kosiol C, Vinar T, da Fonseca RR, Hubisz MJ, Bustamante CD, Nielsen R, et al. Patterns of positive selection in six mammalian genomes. *PLoS Genet*. 2008;4:e1000144.
- Suzuki, Y. Natural selection on the influenza virus genome. *Mol Biol Evol*. 2006;23(10):1902-11.
- Bustamante CD, Fledel-Alon A, Williamson S, Nielsen R, Hubisz MT, Gnanowski S, et al. Natural selection on protein-coding genes in the human genome. *Nature*. 2005;437:1153-7.
- Begun DJ, Holloway AK, Stevens K, Hillier LW, Poh Y-P, Hahn MW, et al. Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biol*. 2007;5(11):e310.
- Akey JM, Zhang G, Zhang K, Jin L, Shriver MD. Interrogating a high-density SNP map for signatures of natural selection. *Genome Res*. 2002;12:1805-14.
- Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZXP, Pool JE, et al. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science*. 2010;329:75-8.
- Akey JM, Ruhe AL, Akey DT, Wong AK, Connelly CF, Madeoy J, et al. Tracking footprints of artificial selection in the dog genome. *Proc Natl Acad Sci USA*. 2010;107:1160-5.
- Coelho M, Luiselli D, Bertorelle G, Lopes AI, Seixas S, Destro-Bisol G, et al. Microsatellite variation and evolution of human lactase persistence. *Hum Genet*. 2005;117:329-39.
- Begun DJ, Aquadro CF. Levels of naturally occurring DNA polymorphism correlate with recombination rate in *D. melanogaster*. *Nature*. 1992;356:519-20.
- Cai JJ, Macpherson JM, Sella G, Petrov DA. Pervasive hitchhiking at coding and regulatory sites in humans. *PLoS Genet*. 2009;5:e1000336.
- Charlesworth B, Morgan MT, Charlesworth D. The effect of deleterious mutations on neutral molecular variation. *Genetics*. 1993;134:1289-303.
- King MC, Wilson AC. Evolution at two levels in humans and chimpanzees. *Science*. 1975;188:107-16.
- Shapiro MD, Marks ME, Peichel CL, Blackman BK, Nereng KS, Jónsson B, et al. Genetic and developmental basis of evolutionary pelvic reduction in threespine stickleback. *Nature*. 2004;428:717-23.
- Chan YF, Marks ME, Jones C, Villarreal G, Shapiro MD, Brady SD, et al. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *Science*. 2010;327:302-5.
- Mallarino R, Grant PR, Grant BR, Herrel A, Kuo W, Abzhanov A. Two developmental modules establish 3D beak-shape variation in Darwin's finches. *Proc Natl Acad Sci USA*. 2011;108:4057-62.
- Abzhanov A, Kuo WP, Hartmann C, Grant R, Grant PR, Tabin CJ. The Calmodulin pathway and the evolution of elongated beak morphology in Darwin's Finches. *Nature*. 2006;442:563-7.
- Abzhanov A, Protas MB, Grant R, Grant PR, Tabin CJ. *Bmp4* and morphological variation of beaks in Darwin's finches. *Science*. 2004;305:1462-5.
- Weiner J. *The beak of the finch: a story of evolution in our time*. New York: Vintage, 1995.
- Grant BR, Grant PR. What Darwin's finches can teach us about the evolutionary origin and regulation of biodiversity. *BioScience*. 2003;53:965-75.
- Andolfatto P. Adaptive evolution of non-coding DNA in *Drosophila*. *Nature*. 2005;437(7062):1149-52.
- Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, et al. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature*. 2007;450(7167):203-18.
- Nowick K, Gernat T, Almaas E, Stubbs L. Differences in human and chimpanzee gene expression patterns define an evolving network of transcription factors in brain. *Proc Natl Acad Sci*. 2009;106:25358-63.
- Oldham MC, Horvath S, Geschwind DH. Conservation and evolution of gene coexpression networks in human and chimpanzee brains. *Proc Natl Acad Sci USA*. 2006;103:17973-8.
- Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, et al. The genetic landscape of a cell. *Science*. 2010;327:425-31.
- Khaitovich P, Weiss G, Lachmann M, Hellmann I, Enard W, Muetzel B, et al. A neutral model of transcriptome evolution. *PLoS Biol*. 2004;2:E132.
- Rifkin SA, Kim J, White KP. Evolution of gene expression in the *Drosophila melanogaster* subgroup. *Nature Genet*. 2003;33:138-44.
- Oleksiak MF, Churchill GA, Crawford DL. Variation in gene expression within and among natural populations. *Nat Genet*. 2002;32:261-6.
- Lynch M. The frailty of adaptive hypotheses for the origins of organizational complexity. *Proc Natl Acad Sci USA*. 2007;104(suppl. 1):8597-604.
- Cavalier-Smith T. Nuclear volume control by nucleoskeletal DNA, selection for cell volume and cell growth rate, and the solution of the DNA C-value paradox. *J Cell Sci*. 1978;34:247-78.

