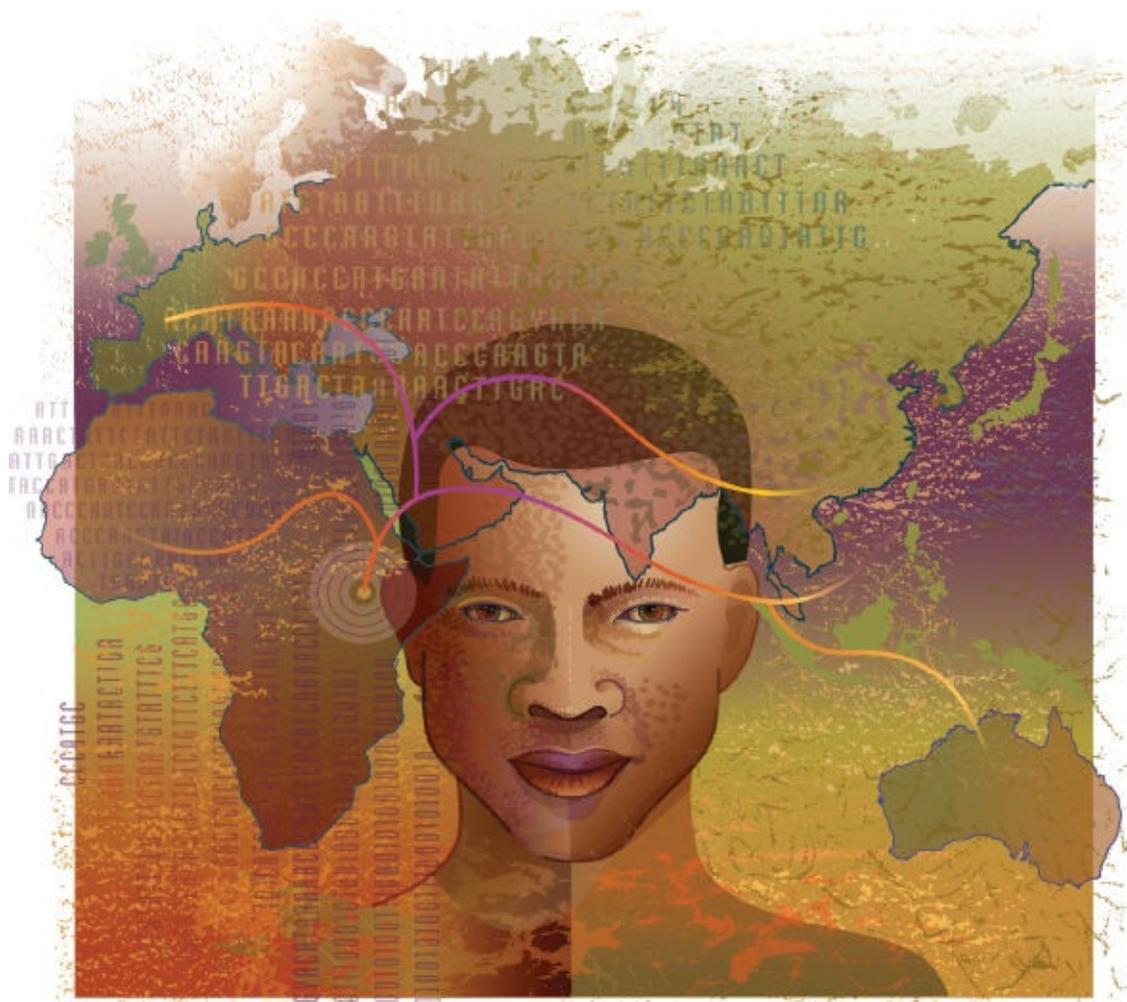


Genética de Populações

CAPÍTULO

18



Concepção da artista Lynn Fellman do “Adão Eurasiano”, um homem africano com um cromossomo Y pertencente a um grupo de haplótipos que foi o ancestral de todos os cromossomos Y de homens fora da África e que surgiu na África há aproximadamente 70.000 anos. (Lynn Fellman, www.Fellmanstudio.com.)

TÓPICOS

- 18.1 Detecção da variação genética
- 18.2 Conceito de pool gênico e lei de Hardy-Weinberg
- 18.3 Sistemas de acasalamento
- 18.4 Variação genética e sua medida
- 18.5 Modulação da variação genética
- 18.6 Aplicações biológicas e sociais

RESULTADOS DE APRENDIZAGEM

Após ler este capítulo, você será capaz de:

- Analisar dados para determinar quanta variação genética existe nas populações
- Desenhar um experimento para testar se uma população de organismos está de acordo com as expectativas de Hardy-Weinberg
- Explicar como novos alelos entram em uma população
- Compreender os impactos negativos do endocruzamento em uma população
- Descrever os tipos de seleção em relação à genética de populações
- Prever de que maneira forças como a seleção, a mutação e a deriva genética alteram a quantidade de variação nas populações.

Em 2009, Sean Hodgson foi libertado de uma prisão britânica após cumprir 27 anos pelo assassinato de Teresa De Simone, balconista e garçonne de um bar. Hodgson, que sofre de doença mental, inicialmente confessou o crime, mas retirou a sua confissão durante o processo. Durante seus anos na prisão, ele afirmou a sua inocência. Mais de duas décadas após o crime, a justiça analisou o DNA do assaltante, encontrado na cena do crime, e determinou que ele não era de Hodgson. A sua condenação foi anulada e a polícia reabriu a investigação do assassinato da Sra. De Simone. Conforme você aprenderá neste capítulo, a análise com base no DNA utilizada para absolver o Sr. Hodgson e centenas de outros prisioneiros erroneamente condenados depende da análise

genética de populações.

Os princípios da genética de populações são o cerne de muitas questões que a sociedade enfrenta atualmente. Quais são os riscos de que um casal tenha um filho com uma doença genética? As práticas de cultivo de plantas e criação de animais causaram uma perda da diversidade genética na agroindústria que imponha riscos para o nosso suprimento alimentar? Na medida em que a população humana continua a se expandir e a vida selvagem recua para partes cada vez menores da Terra, as espécies selvagens serão capazes de evitar o endocruzamento e sobreviver? Os princípios da genética de populações também são fundamentais para a compreensão de muitas questões históricas e evolutivas. Como as populações humanas de diferentes regiões do mundo estão relacionadas umas às outras? Como o genoma humano reagiu à medida que os seres humanos se espalharam por todo o globo e se tornaram adaptados a diferentes ambientes e estilos de vida? Como as populações e as espécies evoluem ao longo do tempo?

Uma **população** é um grupo de indivíduos da mesma espécie. A **genética de populações** analisa a quantidade e a distribuição da variação genética nas populações e as forças que controlam essa variação. Ela tem suas origens no início do século 20, quando geneticistas começaram a estudar como as leis de Mendel poderiam ser estendidas para compreender a variação genética dentro de populações inteiras de organismos. Embora as leis de Mendel expliquem como os genes são transmitidos dos genitores para a descendência nos casos de cruzamentos controlados e heredogramas conhecidos, essas leis são insuficientes para a compreensão da transmissão de genes de uma geração para a próxima em populações naturais, nas quais nem todos os indivíduos produzem descendência e nem toda a descendência sobrevive. Geneticistas iniciaram o desenvolvimento dos princípios da genética de populações no início do século 20, mas naquela época eles tinham ferramentas um tanto quanto limitadas para realmente medir a variação genética. Com o desenvolvimento das tecnologias com base no DNA ao longo das últimas três décadas, os geneticistas atualmente têm a capacidade de observar diretamente as diferenças entre as sequências de DNA dos indivíduos em todo o seu genoma e podem medir essas diferenças em grandes amostras de indivíduos em muitas espécies. O resultado tem sido uma revolução na nossa

compreensão sobre a variação genética nas populações.

Neste capítulo, consideraremos o conceito de *pool* gênico e como os geneticistas estimam as frequências de alelos e de genótipos nas populações. Em seguida, examinaremos o impacto que os sistemas de cruzamento apresentam sobre as frequências de genótipos em uma população. Também discutiremos como os geneticistas medem a variação com a utilização de tecnologias baseadas no DNA. Em seguida, discutiremos as forças que modulam os níveis de variação genética nas populações. Finalmente, observaremos alguns estudos de caso que envolvem a aplicação da genética de populações para questões de interesse para a sociedade.

18.1 Detecção da variação genética

Os métodos da genética de populações podem ser utilizados para analisar qualquer variável ou *locus* polimórfico nas sequências de DNA de uma população de organismos. Historicamente, os geneticistas não tinham as ferramentas moleculares necessárias para observar diretamente as diferenças nas sequências de DNA entre os indivíduos e, assim, a maior parte das análises da genética de populações observava as diferenças nas proteínas ou nos fenótipos. Por exemplo, as diferenças na proteína glicosiltransferase codificada pelo gene *ABO*, que controla o grupo sanguíneo ABO em seres humanos, podem ser detectadas com a utilização de sondas de anticorpos. A partir dessas diferenças na proteína, os investigadores podem inferir diferenças na sequência de DNA desse gene entre os indivíduos. Ao longo das últimas três décadas, foram desenvolvidas novas tecnologias, tais como sequenciamento do DNA, microarranjos de DNA e PCR (ver [Capítulos 10 e 14](#)), possibilitando que os geneticistas observem as diferenças nas sequências de DNA diretamente. Como resultado, as análises da genética de populações deixaram de estar confinadas a um pequeno conjunto de genes, tais como *ABO*, tendo sido expandidas para incluir todos os nucleotídeos no genoma.

Na genética de populações, um *locus* é simplesmente um local no genoma; ele

pode ser um sítio de nucleotídeo único ou um segmento de muitos nucleotídeos. O tipo mais simples de variação que se pode observar entre os indivíduos em um *locus* é uma diferença no nucleotídeo presente em um sítio de nucleotídeo único, seja adenina, citosina, guanina ou timina. Esses tipos de variantes são denominados **polimorfismos de nucleotídeo único (SNP)** e são as variantes mais amplamente estudadas na genética de populações humanas (Figura 18.1; ver também Capítulo 4). A genética de populações também faz uso extensivo dos *loci* de **microssatélites** (ver Capítulo 4). Esses *loci* apresentam um *motif* de sequência curta, com comprimento de dois a seis pares de bases, que é repetido múltiplas vezes com diferentes alelos e números diferentes de repetições. Por exemplo, o *motif* de sequência AG de 2 pb em um *locus* pode ser repetido em *tandem* cinco vezes em um alelo (AGAGAGAGAG), mas três vezes em outro (AGAGAG) (ver Figura 18.1).

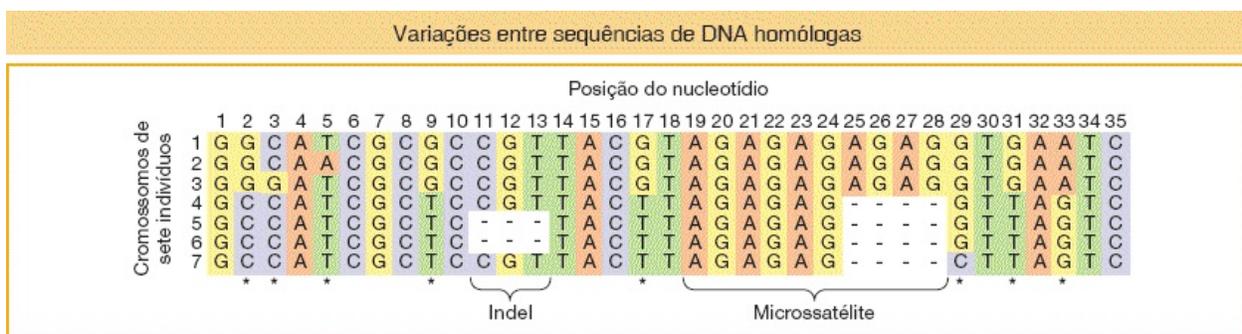


FIGURA 18.1 Variação nas sequências de DNA alinhadas de sete cromossomos de diferentes pessoas. Os asteriscos demonstram a localização dos SNP. A localização de uma indel (inserção/deleção de uma série de pares de nucleotídeos) e de um microssatélite também está indicada.

Polimorfismos de nucleotídeo único

Os polimorfismos de nucleotídeo único (SNP) são os tipos de polimorfismo mais prevalentes na maior parte dos genomas. A maioria dos SNP apresenta apenas dois alelos — por exemplo, A e C. Os SNP normalmente são considerados **SNP comuns** em uma população se o alelo menos comum ocorre a uma frequência de aproximadamente 5% ou superior. Os SNP em relação aos quais o alelo menos comum ocorre a uma frequência inferior a 5% são considerados **SNP raros**. Em relação aos seres humanos, existe um SNP comum aproximadamente a cada 300 a

1.000 pb no genoma. É claro que existe um número muito maior de SNP raros.

Os SNP ocorrem dentro de genes, incluindo éxons, íntrons e regiões reguladoras. Os SNP em regiões codificadoras de proteínas podem ser classificados em um de três grupos: *sinônimos*, se os diferentes alelos codificam o mesmo aminoácido; *não sinônimos*, se os dois alelos codificam aminoácido diferentes; e *sem sentido*, se um alelo codifica um códon de fim e o outro, um aminoácido. Portanto, por vezes é possível associar um SNP a uma variação funcional nas proteínas e a uma alteração associada no fenótipo. Os SNP localizados fora de sequências codificadoras são denominados SNP *não codificadores* (ncSNP). Se os ncSNP não apresentam efeitos sobre a função gênica e o fenótipo, eles são denominados *silenciosos*. Os ncSNP silenciosos podem ser muito úteis na genética de populações, tendo em vista que podem ser utilizados como marcadores para abordar questões a respeito de processos da genética de populações, tais como o fluxo gênico entre populações.

Para estudar a variação de SNP em uma população, primeiramente precisamos determinar quais sítios de nucleotídios no genoma são variáveis — ou seja, constituem um SNP. Essa primeira etapa é denominada descoberta de SNP. Os SNP com frequência são descobertos por meio do sequenciamento de genomas de uma pequena amostra de indivíduos de uma espécie e em seguida com a comparação dessas sequências. Por exemplo, a descoberta de SNP em seres humanos teve início por meio do sequenciamento parcial de genomas de um **painel de descoberta** de 48 indivíduos de todo o mundo. Os sítios de nucleotídios variáveis foram descobertos por meio da comparação das sequências genômicas parciais desses 48 indivíduos entre si. Esse esforço inicial levou à descoberta de mais de 1 milhão de SNP.

Após a descoberta dos SNP, pode ser determinado o genótipo (a composição alélica) de diferentes indivíduos na população para cada SNP. Os microarranjos de DNA são uma tecnologia amplamente utilizada para essa finalidade ([Figura 18.2](#)). Os microarranjos utilizados para análises de SNP podem conter milhares de sondas correspondentes aos SNP conhecidos. Biotecnólogos desenvolveram diversos métodos diferentes para detectar variantes de SNP com a utilização de microarranjos. Em um método, o DNA de um indivíduo é marcado com corantes

fluorescentes e hibridizado com o microarranjo. Cada ponto (SNP) no microarranjo fluorescerá em vermelho para uma classe homozigota, em verde para a outra homozigota e em amarelo para o heterozigoto (ver [Figura 18.2](#)). O procedimento inteiro tem sido intensificado com a robótica para possibilitar a rápida *genotipagem* ou atribuição de genótipos (p. ex., *A/A versus A/C*) em grande escala.

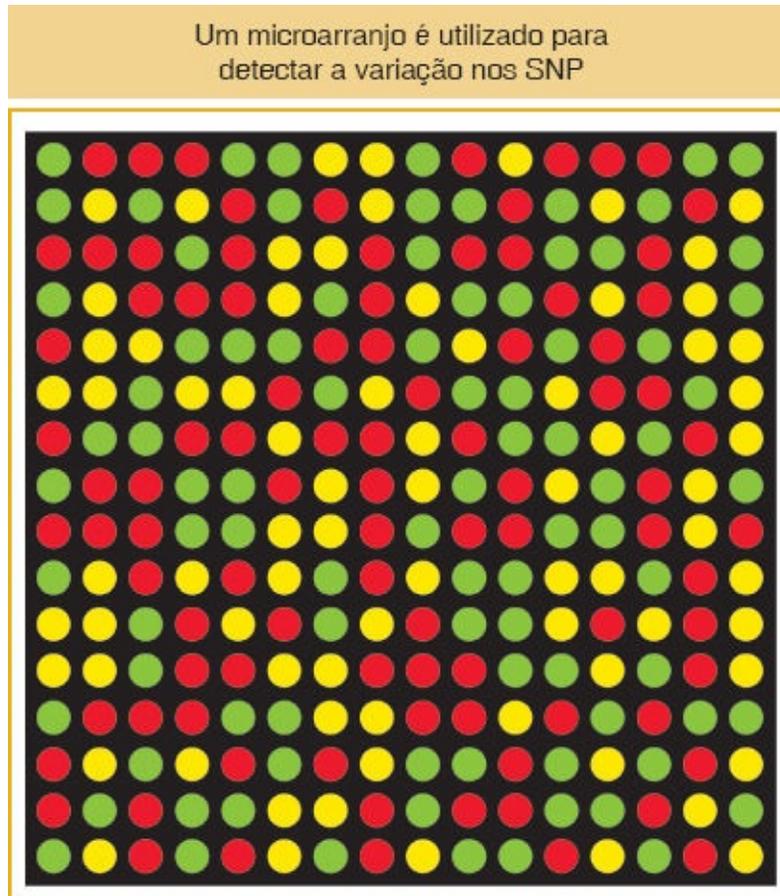


FIGURA 18.2 Detecção da variação no DNA: os SNP. Visualização de uma pequena parte de um microarranjo utilizado para realizar a varredura do genoma de um único indivíduo. Cada ponto representa um SNP, com *vermelho* e *verde* para as classes homozigotas e *amarelo* para os heterozigotos.

Microssatélites

Os microssatélites são *loci* poderosos para a análise da genética de populações por diversos motivos. Primeiramente, ao contrário dos SNP, que tipicamente apresentam apenas dois alelos por *locus* e que nunca podem apresentar mais de

quatro alelos, o número de alelos em um microsatélite com frequência é muito grande (20 ou mais). Em segundo lugar, eles apresentam uma alta taxa de mutação, tipicamente na faixa de 10^{-3} a 10^{-4} mutações por *locus* por geração, em comparação a 10^{-8} a 10^{-9} mutações por sítio por geração em relação aos SNP. A alta taxa de mutação significa que os níveis de variação são mais altos: mais alelos por *locus* e maior chance de que quaisquer dois indivíduos apresentem genótipos diferentes. Em terceiro lugar, os microsatélites são muito abundantes na maior parte dos genomas. Os seres humanos apresentam mais de um milhão de microsatélites.

Os microsatélites são observados por todo o genoma da maior parte dos organismos e podem estar presentes em éxons, íntrons, regiões reguladoras e sequências não funcionais de DNA. Os microsatélites com repetições de trinucleotídeos são observados nas sequências codificadoras de alguns genes; esses codificam séries de um único aminoácido. O gene da doença de Huntington (*HD*) (ver [Capítulo 16](#)) contém uma repetição de CAG, que codifica uma série de glutaminas. Os indivíduos que carregam alelos com mais de 30 glutaminas são predispostos ao desenvolvimento da doença. Entretanto, em geral, a maior parte dos microsatélites está localizada fora das sequências codificadoras e a variação no número de repetições não está associada a diferenças no fenótipo.

São utilizados dois métodos principais para descobrir os *loci* de microsatélites no genoma. Se uma sequência genômica completa estiver disponível em relação a um organismo, pode-se simplesmente conduzir uma pesquisa para encontrá-los com a utilização de um computador. Em relação às espécies sem sequências genômicas (a maior parte dos organismos não modelo), é necessário um trabalho laboratorial considerável para descobrir microsatélites. Tipicamente, cria-se uma biblioteca genômica, realiza-se a varredura da biblioteca com uma sonda para o *motif* de interesse (p. ex., repetições AG) e determina-se a sequência de DNA dos clones selecionados para identificar os microsatélites e as sequências que os flanqueiam. Os métodos moleculares para a realização desse tipo de trabalho foram discutidos no [Capítulo 10](#).

Após a identificação de um microsatélite e suas sequências flanqueadoras,

amostras de DNA de um conjunto de indivíduos na população podem ser analisadas para determinar o número de repetições que estão presentes em cada indivíduo. Para realizar a análise, são projetados *primers* de oligonucleotídeos que correspondam às sequências flanqueadoras para a utilização na PCR. Se os *primers* forem marcados com um corante fluorescente, então os tamanhos dos produtos da PCR podem ser determinados no mesmo aparato utilizado para determinar a sequência de moléculas de DNA (Figura 18.3). Esses tamanhos revelam o número de repetições em um alelo de microssatélite. Por exemplo, o produto da PCR de um alelo de microssatélite que contém sete repetições AG será 8 pb mais longo do que um alelo que contém três repetições AG. Os indivíduos heterozigotos possuirão produtos de dois tamanhos diferentes. Tendo em vista que a PCR, a determinação do tamanho dos produtos da PCR e a classificação dos alelos podem ser, todas, automatizadas, é possível determinar os genótipos de grandes amostras de indivíduos em relação a grandes números de microssatélites de modo relativamente rápido.

Haplótipos

Para algumas questões na genética de populações, é importante considerar os genótipos de *loci* ligados como um grupo, em vez de individualmente. Os geneticistas utilizam o termo **haplótipo** para fazer referência à combinação de alelos em múltiplos *loci* no mesmo cromossomo homólogo. Dois cromossomos homólogos que compartilham o mesmo alelo em cada um dos *loci* em consideração têm o mesmo haplótipo. Se dois cromossomos apresentam genótipos diferentes, até mesmo em um dos *loci* em questão, eles então apresentam haplótipos diferentes. Se o *locus A* com alelos *A* e *a* estiver ligado ao *locus B* com alelos *B* e *b*, então existem quatro haplótipos possíveis em relação ao segmento cromossômico no qual esses *loci* estão localizados:

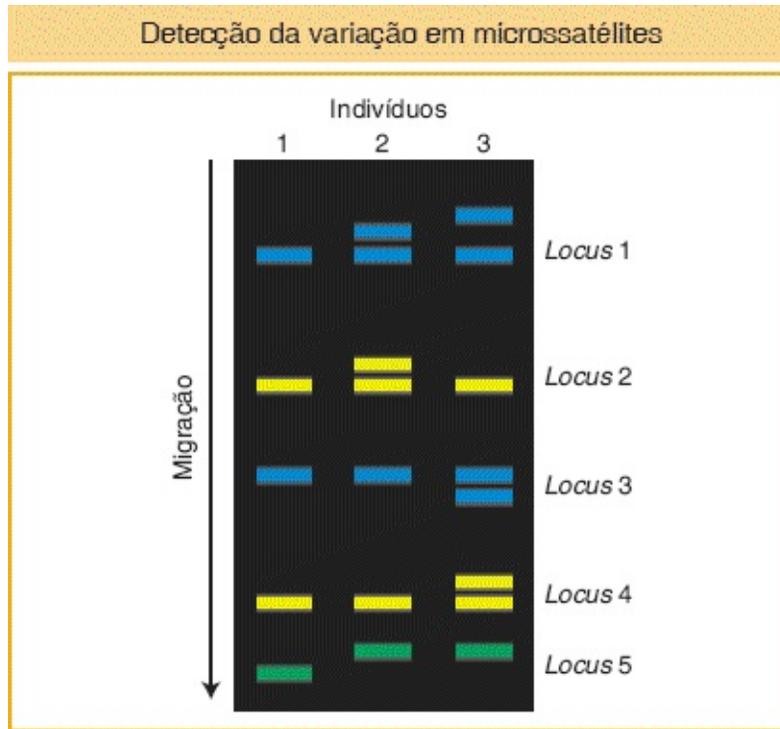


FIGURA 18.3 Deteção da variação no DNA: microssatélites. Desenho esquemático de uma imagem de gel dos *loci* em relação a cinco microssatélites avaliados simultaneamente. As três fileiras verticais correspondem a três indivíduos. Observe que existem três alelos presentes em relação ao *locus* 1 e que os indivíduos 2 e 3 são, ambos, heterozigotos em relação a este *locus*.

<i>A</i>	<i>B</i>

Um exemplo mais complexo, porém mais realista, está demonstrado na [Figura 18.4](#). Na [Figura 18.4 A](#), existem sete segmentos cromossômicos, mas apenas seis haplótipos, tendo em vista que os segmentos cromossômicos 5 e 6 apresentam o mesmo haplótipo (E).

Os haplótipos são utilizados com mais frequência na genética de populações em relação aos *loci* que estão fisicamente próximos. Por exemplo, os sítios de nucleotídeos variáveis em um único gene podem ser utilizados para definir

haplótipos em relação àquele gene. Entretanto, o conceito de haplótipo funciona em relação a regiões maiores quando há pouca ou nenhuma recombinação ao longo da região. Ele pode até mesmo ser aplicado para um cromossomo inteiro, tal como o cromossomo Y humano. Finalmente, por vezes ele é útil para agrupar haplótipos em classes. Conforme demonstrado na [Figura 18.4 A](#), existem duas classes principais de haplótipos (I e II), que diferem em cinco sítios de nucleotídios mais um microssatélite. Entretanto, cada classe contém diversos subtipos (I-a, I-b, ...). A **rede de haplótipos** demonstra as relações entre os haplótipos, posicionando cada mutação em um dos ramos ([Figura 18.4 B](#)).

Quais percepções podemos obter a partir da análise de haplótipos? Os geneticistas de populações estudando o cromossomo Y humano entre homens asiáticos descobriram um haplótipo altamente prevalente, denominado haplótipo “aglomerado estelar” ([Figura 18.5 A](#)). Tipicamente, a maior parte dos homens apresenta um haplótipo de cromossomo Y raro, mas o haplótipo “aglomerado estelar” está presente em 8% dos homens asiáticos. Com a utilização da taxa de mutação conhecida, os pesquisadores estimaram que esse haplótipo comum surgiu entre 700 e 1.300 anos atrás. (Posteriormente neste capítulo, discutiremos as taxas de mutação e sua utilização na genética de populações.) Esse haplótipo é mais comum na Mongólia, sugerindo que ele surgiu ali. Os pesquisadores inferiram que o haplótipo “aglomerado estelar” remonta a um homem na Mongólia há aproximadamente 1.000 anos. Notavelmente, a distribuição atual desse haplótipo segue os limites geográficos do Império Mongol estabelecido por Genghis Khan há aproximadamente 1.200 anos ([Figura 18.5 B](#)). Aparentemente todos os homens contemporâneos com esse haplótipo são descendentes de Genghis Khan (ou de seus parentes da linhagem masculina).

Uma rede de haplótipos demonstra a relação entre os haplótipos

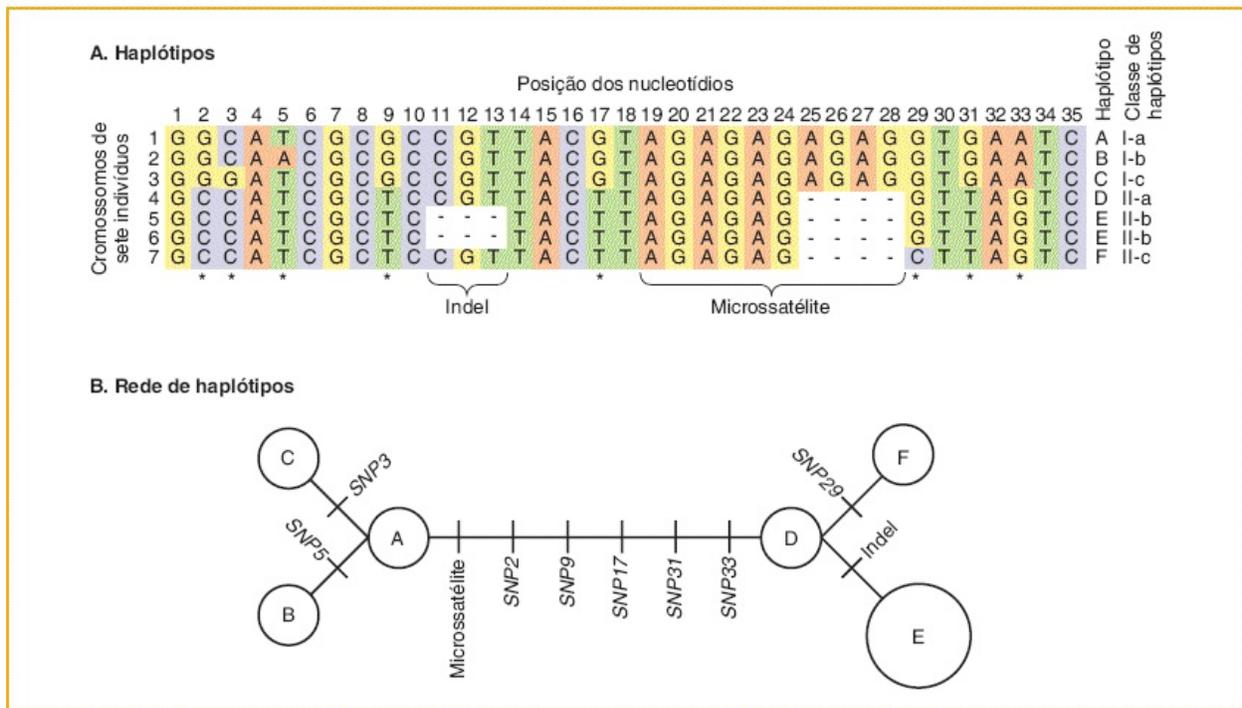


FIGURA 18.4 A. Existe um total de seis haplótipos (A a F) nas seqüências de DNA alinhadas de sete cromossomos individuais de diferentes pessoas. B. Estes seis haplótipos estão unidos em uma rede de haplótipos que demonstra as relações entre os haplótipos. Cada círculo representa um dos seis haplótipos. Quaisquer dois haplótipos diferem nos *loci* observados em todos os ramos que os conectam. Os asteriscos demonstram a localização dos SNP.

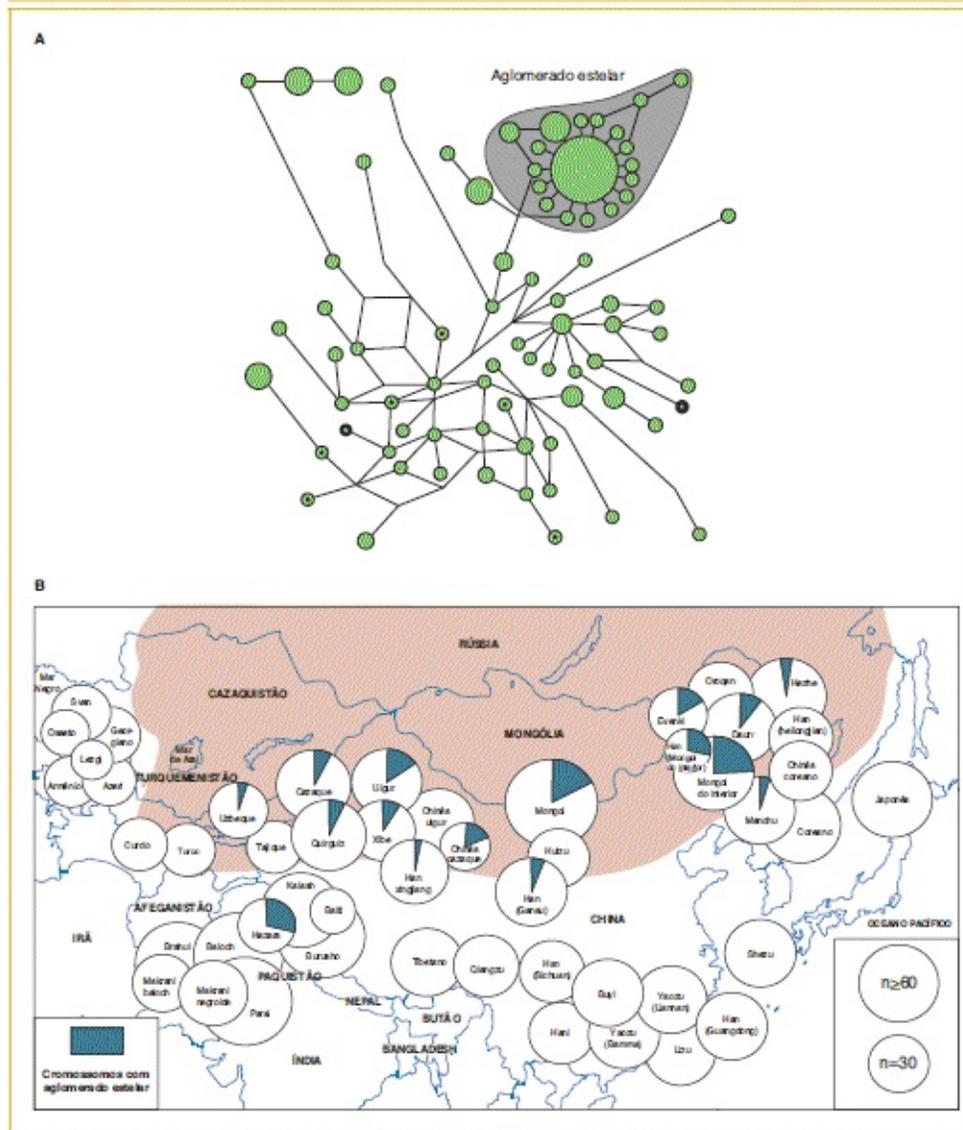


FIGURA 18.5 A. Rede de haplótipos em relação aos cromossomos Y de homens asiáticos, demonstrando a predominância do haplótipo aglomerado estelar que se acredita remontar a Genghis Khan. A área do *círculo* é proporcional ao número de indivíduos com o haplótipo específico que o círculo representa. **B.** Distribuição geográfica do haplótipo aglomerado estelar. As populações estão demonstradas como *círculos*, com uma área proporcional ao tamanho da amostra; a proporção de indivíduos na amostra que carregam cromossomos com aglomerado estelar está indicada pelos *setores verdes*. Não foram observados cromossomos com aglomerado estelar nas populações que não apresentam *setor verde no círculo*. A *área sombreada* representa a extensão do império de Genghis Khan. (Dados de T. Zerjal *et al.*, *Am. J. Hum. Genet.* 72, 2003, 717-721.)

Outras fontes e outros tipos de variação

Além dos SNP e dos microssatélites, qualquer variação na sequência de DNA dos

cromossomos em uma população é passível de análise genética de populações. As variações que podem ser analisadas incluem inversões, translocações, deleções ou duplicações, e a presença ou a ausência de um elemento de transposição em um *locus* em particular no genoma. Outro tipo comum de variação é o polimorfismo por inserção e deleção ou *indel*, abreviadamente (ver [Capítulo 16](#)). Esse tipo de polimorfismo envolve a presença ou a ausência de um ou mais nucleotídeos em um *locus* em um alelo em relação a outro. Na [Figura 18.1](#), os segmentos cromossômicos 5 e 6 diferem dos outros cinco segmentos por uma *indel* de 3 pb. Contrariamente aos microssatélites, as *indels* não contêm *motifs* repetidos, tais como AGAGAGAG.

Até agora, a nossa discussão sobre os SNP e os microssatélites enfocou o genoma nuclear. Entretanto, também pode ser observada uma interessante variação genética nos genomas mitocondrial (mtDNA) e de cloroplasto (cpDNA) de eucariotos. Tanto os SNP quanto os microssatélites são observados nesses genomas de organelas. Tendo em vista que o mtDNA e o cpDNA normalmente são herdados por via materna, sua análise pode ser utilizada para seguir a história das linhagens femininas. Em 1987, um proeminente estudo da linhagem mitocondrial humana traçou a história dos haplótipos do mtDNA humano e determinou que os genomas mitocondriais de todos os seres humanos modernos remontam a uma única mulher que viveu na África há aproximadamente 150.000 anos ([Figura 18.6](#)). Ela foi apelidada de “Eva Mitocondrial” pela imprensa popular. Esse estudo do mtDNA foi a primeira análise genética completa a sugerir que todos os seres humanos modernos vieram da África.

Projeto HapMap

Um importante avanço na genética de populações humanas ao longo da última década foi a criação de um mapa de haplótipos de todo o genoma, ou **HapMap**. Um consórcio de cientistas ao redor do mundo genotipou milhares de pessoas que representam a diversidade de nossa espécie em relação a centenas de milhares de SNP e microssatélites. O resultado é um quadro altamente detalhado da variação em nossa espécie. Os dados estão disponíveis para o público em diversos *web sites*, incluindo aquele do International HapMap Project (www.hapmap.org) e do

Human Genome Diversity Project (hgdp.uchicago.edu). Neste capítulo, utilizaremos esses dados para apresentar os princípios da genética de populações. Embora primeiramente desenvolvidos para seres humanos, os HapMaps têm sido, desde então, desenvolvidos para diversas outras espécies, incluindo *Drosophila*, camundongo, *Arabidopsis*, arroz e milho.

CONCEITO-CHAVE Os genomas estão repletos de diversos tipos de variação passíveis de análise por genética de populações. Os SNP e os microssatélites são os dois tipos de polimorfismos mais comumente estudados na genética de populações. Tecnologias de alta produção possibilitam que centenas de milhares de polimorfismos sejam classificados em dezenas de milhares de indivíduos.

Os haplótipos mitocondriais podem ser utilizados para traçar as origens humanas até a África

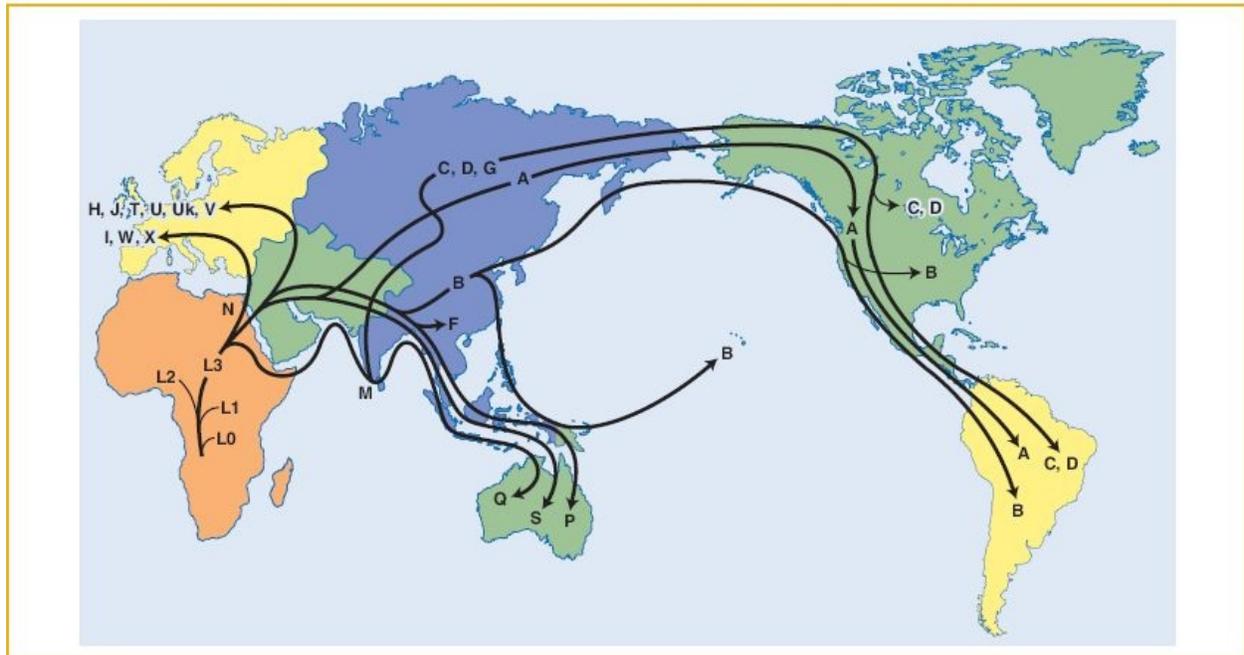


FIGURA 18.6 A rede de haplótipos para grupos de haplótipos do mtDNA desenhada em um mapa mundial. O grupo de haplótipo L ancestral aparece na África e os grupos derivados (A, B e assim por diante) estão dispersos por todo o mundo. (Dados de www.mitomap.org.)

18.2 Conceito de pool gênico e lei de Hardy-Weinberg

Talvez você tenha visto alguém realizando um truque mortal e pensado que a pessoa sofria o risco de eliminar a si própria do “*pool* gênico”. Em caso afirmativo, você estava utilizando um conceito, o *pool* gênico, que tem sua origem diretamente na genética de populações e que trilhou o seu caminho até a cultura popular. O conceito de *pool* gênico é uma ferramenta básica para se pensar a respeito da variação genética nas populações. Podemos definir o **pool gênico** como a soma total de todos os alelos nos membros reprodutivos de uma população em um determinado momento. Por exemplo, a [Figura 18.7](#) demonstra uma população de 16 rãs, cada uma das quais carregando dois alelos no *locus* autossômico *A*. Por meio da simples contagem, podemos determinar que existem cinco homozigotos *A/A*, oito heterozigotos *A/a* e três homozigotos *a/a*. O tamanho da população, normalmente simbolizado pela letra *N*, é 16, e existe um total de 32 alelos, ou $2N$, nessa população diploide. Com esse simples conjunto de números, descrevemos o *pool* gênico em relação ao *locus* *A*.

Tipicamente, os geneticistas de populações não se importam com contagens absolutas dos diferentes genótipos em uma população, mas sim com as **frequências genotípicas**. Podemos calcular a frequência do genótipo *A/A* simplesmente dividindo o número de indivíduos *A/A* pelo número total de indivíduos na população (*N*), para obter 0,31. A frequência de heterozigotos *A/a* é de 0,50, e a frequência de homozigotos *a/a* é de 0,19. Por serem frequências, elas somam 1,0. As frequências são uma medida mais prática do que as contagens absolutas, tendo em vista que raramente os geneticistas de populações são capazes de estudar todos os indivíduos em uma população. Em vez disso, os geneticistas de populações coletarão uma amostra aleatória ou não tendenciosa de indivíduos em uma população e utilizarão a amostra para inferir as frequências genotípicas na população inteira.

O *pool* gênico é a soma total de alelos em uma população

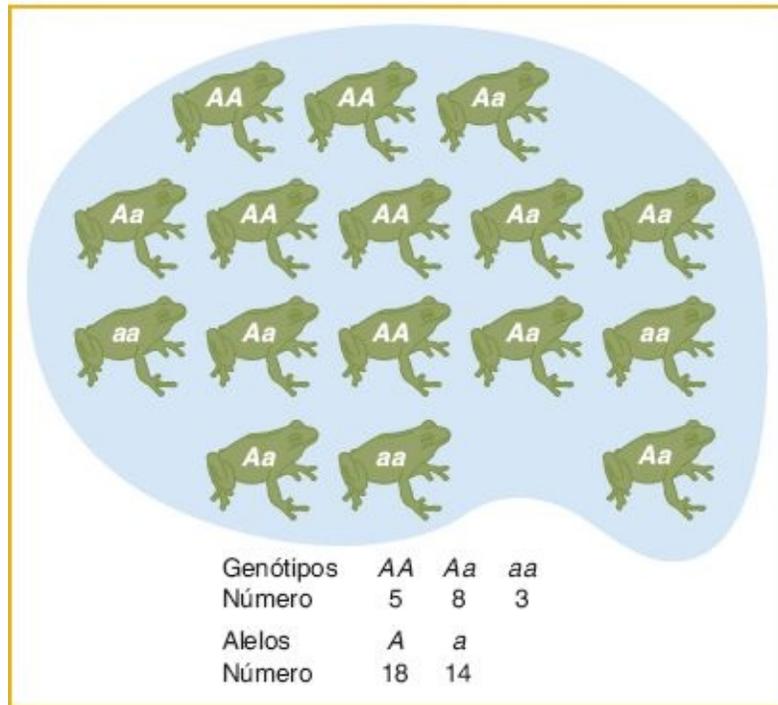


FIGURA 18.7 Um *pool* gênico de rãs.

Podemos fazer uma descrição mais simples desse *pool* gênico de rãs se calcularmos as **frequências alélicas** em vez das frequências genotípicas (Quadro 18.1). Na Figura 18.7, 18 dos 32 alelos são A e, assim, a frequência de A é $18/32 = 0,56$. A frequência do alelo A é tipicamente simbolizada pela letra p e, nesse caso, $p = 0,56$. A frequência do alelo a é simbolizada pela letra q e, nesse caso, $q = 14/32 = 0,44$. Novamente, tendo em vista que essas são frequências, a sua soma é 1,0: $p + q = 0,56 + 0,44 = 1,0$. Agora temos uma descrição do nosso *pool* gênico de rãs com a utilização de apenas dois números, p e q .

CONCEITO-CHAVE O *pool* gênico é um conceito fundamental para o estudo da variação genética nas populações: ele é a soma de todos os alelos nos membros reprodutivos de uma população em uma determinada ocasião. Podemos descrever a variação em uma população em termos das frequências genotípicas e alélicas.

Conforme mencionado anteriormente, um objetivo importante da genética de populações é compreender a transmissão de alelos de uma geração até a próxima nas populações naturais. Nesta seção, começaremos a ver como isso funciona. Veremos como podemos utilizar as frequências alélicas no *pool* gênico para fazer previsões a respeito das frequências genotípicas na próxima geração.

Quadro 18.1

Cálculo das frequências alélicas.

Em um *locus* com dois alelos A e a , definiremos as frequências dos três genótipos AA , A/a e a/a como f_{AA} , $f_{A/a}$ e $f_{a/a}$, respectivamente. Podemos utilizar essas frequências genotípicas para calcular as frequências alélicas: p é a frequência do alelo A e q é a frequência do alelo a . Tendo em vista que cada homocigoto AA é composto apenas por alelos A e tendo em vista que metade dos alelos de cada heterocigoto A/a é de alelos A , a frequência total p de alelos A na população é calculada como:

$$p = f_{AA} + \frac{1}{2}f_{A/a} = \text{Frequência de } A$$

De modo semelhante, a frequência q do alelo a é fornecida por:

$$q = f_{a/a} + \frac{1}{2}f_{A/a} = \text{Frequência de } a$$

Portanto:

$$p + q = f_{AA} + f_{A/a} + f_{a/a} = 1,0$$

e:

$$q = 1 - p$$

Se houver mais de duas formas alélicas diferentes, a frequência em relação a cada alelo é simplesmente a frequência de seu homocigoto mais metade da soma das frequências de todos os heterocigotos nos quais ele aparece.

A frequência de um alelo no *pool* gênico é igual à probabilidade de que o alelo seja escolhido na coleta aleatória de um alelo do *pool* gênico para formar um ovócito ou um espermatozoide. Sabendo disso, podemos calcular a probabilidade de que uma rã na próxima geração seja um homozigoto A/A . Se nos dirigirmos ao *pool* gênico das rãs (ver [Figura 18.7](#)) e coletarmos o primeiro alelo, a probabilidade de que ele seja um A é $p = 0,56$ e, de modo semelhante, a probabilidade de que o segundo alelo que coletarmos também seja um A é $p = 0,56$. O produto dessas duas probabilidades, ou $p^2 = 0,3136$, é a probabilidade de que uma rã na próxima geração seja A/A . A probabilidade de que uma rã na próxima geração seja a/a é $q^2 = 0,44 \times 0,44 = 0,1936$. Existem dois modos para produzir um heterozigoto. Podemos primeiramente coletar um A com probabilidade p e em seguida coletar um a com probabilidade q , ou podemos coletar o a primeiramente e o A em segundo lugar. Portanto, a probabilidade de que uma rã na próxima geração seja heterozigota A/a é $pq + qp = 2pq = 0,4928$. Em geral, as frequências (f) genótípicas são:

$$f_{A/A} = p^2$$

$$f_{a/a} = q^2$$

$$f_{A/a} = 2pq$$

Finalmente, conforme esperado, a soma da probabilidade de ser A/A , mais a probabilidade de ser A/a , mais a probabilidade de ser a/a é 1,0:

$$p^2 + 2pq + q^2 = 1,0$$

Essa equação simples é a **lei de Hardy-Weinberg**, um dos fundamentos da teoria da genética de populações.

O processo de acesso ao *pool* gênico para coletar um alelo é denominado *amostragem* do *pool* gênico. Tendo em vista que qualquer indivíduo que contribui para o *pool* gênico pode produzir muitos ovócitos ou espermatozoides que carregem exatamente a mesma cópia de um alelo, é possível coletar uma cópia em particular e em seguida voltar ao *pool* gênico e coletar exatamente a mesma cópia mais uma vez. Existe também um elemento de acaso envolvido na amostragem do

pool gênico. Algumas cópias podem, por acaso, ser coletadas mais de uma vez e outras podem absolutamente não ser coletadas. Posteriormente no capítulo, veremos como essas propriedades da amostragem do *pool* gênico podem levar a alterações no *pool* gênico ao longo do tempo.

Utilizamos a lei de Hardy-Weinberg para calcular as frequências genotípicas na próxima geração a partir das frequências alélicas na geração atual. Também podemos utilizar a lei de Hardy-Weinberg para calcular as frequências alélicas a partir das frequências genotípicas em uma única geração. Por exemplo, alguns tipos de albinismo em seres humanos ocorrem em virtude de alelos recessivos no *locus OCA2*. Na África, um tipo de albinismo, denominado albinismo oculocutâneo marrom, resulta de um alelo recessivo de *OCA2* (Figura 18.8). Os indivíduos com essa condição estão presentes a frequências tão altas quanto 1 em 1.100 em alguns grupos étnicos na África. Podemos utilizar a lei de Hardy-Weinberg para calcular as frequências alélicas:

$$f_{a/a} = q^2 = 1/1.100 = 0,0009$$

assim:

e:

$$q = \sqrt{0,0009} = 0,03$$

$$p = 1 - q = 0,97$$

Com a utilização das frequências alélicas, também podemos calcular a frequência de heterozigotos na população como:

$$2pq = 2 \times 0,97 \times 0,03 = 0,06$$

O último número prevê que aproximadamente 6% dessa população são heterozigotos ou portadores do alelo recessivo no *OCA2*.

Quando utilizamos a lei de Hardy-Weinberg para calcular as frequências alélicas ou genotípicas, realizamos algumas presunções críticas.

- *Primeiramente*, presumimos que o cruzamento é aleatório na população quanto ao gene em questão. O desvio do cruzamento aleatório viola essa presunção, tornando inadequada a aplicação da lei de Hardy-Weinberg. Por exemplo, uma tendência de indivíduos fenotipicamente semelhantes de cruzarem entre si viola a lei de Hardy-Weinberg. Se albinos cruzassem mais frequentemente com outros albinos do que com não albinos, então a lei de Hardy-Weinberg superestimaria a frequência do alelo recessivo



FIGURA 18.8 Indivíduo de ancestralidade africana com albinismo oculocutâneo marrom (BOCA), uma condição definida pela pele bronze-clara e pelos cabelos bege a marrom-claros. (Dra. Michele Ramsay, Department of Human Genetics, School of Pathology, the National Health Laboratory Service University of Witwatersrand.)

- *Em segundo lugar*, se um dos genótipos reduzir a viabilidade de tal modo que alguns indivíduos com aquele genótipo morrem antes que as frequências dos genótipos sejam contadas, então a estimativa das frequências gênicas será imprecisa
- *Em terceiro lugar*, para que a lei de Hardy-Weinberg seja aplicada, a

população não pode ser dividida em subpopulações que sejam parcial ou totalmente isoladas geneticamente. Se existirem subpopulações separadas, os alelos podem estar presentes em frequências distintas nas diferentes subpopulações. Nesse caso, a utilização das contagens genótípicas da população em geral pode não fornecer uma estimativa precisa das frequências alélicas globais

- *Finalmente*, a lei de Hardy-Weinberg se aplica estritamente apenas às populações infinitamente grandes. Em relação às populações finitas, haverá desvios das frequências previstas pela lei de Hardy-Weinberg em virtude do acaso quando a amostragem do *pool* gênico produzir a próxima geração.

Observamos como podemos utilizar a lei de Hardy-Weinberg e as frequências gênicas na geração atual (t_0) para calcular as frequências genótípicas na próxima geração (t_1) por meio da amostragem aleatória do *pool* gênico para a produção de ovócitos e espermatozoides. De modo semelhante, as frequências genótípicas previstas em relação à geração t_1 podem ser utilizadas, por sua vez, para calcular as frequências gênicas em relação à próxima geração (t_2). As frequências gênicas na geração t_2 permanecerão as mesmas da geração t_1 . Sob a lei de Hardy-Weinberg, nem as frequências gênicas nem as genótípicas mudam de uma geração para a próxima quando uma população infinitamente grande é amostrada aleatoriamente para a formação de ovócitos e espermatozoides. Portanto, uma lição importante da lei de Hardy-Weinberg é que, em populações grandes, a variação genética não é criada nem destruída pelo processo de transmissão de genes de uma geração para a próxima. Diz-se que as populações que aderem a esse princípio estão em **equilíbrio de Hardy-Weinberg**.

Geração	Frequências genótípicas			Frequências gênicas	
	<i>A/A</i>	<i>A/a</i>	<i>a/a</i>	<i>A</i>	<i>a</i>
t_0	0,64	0,32	0,04	0,8	0,2

t_1	0,64	0,32	0,04	0,8	0,2
.
.
.
t_n	0,64	0,32	0,04	0,8	0,2

Aqui estão mais alguns poucos pontos a respeito da lei de Hardy-Weinberg.

1. Em relação a qualquer alelo que exista em uma frequência muito baixa, os indivíduos homocigotos serão observados apenas muito raramente. Se o alelo a apresenta uma frequência de 1 em mil ($q = 0,001$), então apenas 1 em 1 milhão (q^2) de indivíduos será homocigoto em relação àquele alelo. Conseqüentemente, os alelos recessivos para distúrbios genéticos podem ocorrer no estado heterocigoto em muito mais indivíduos do que aqueles que de fato expressam o distúrbio genético em questão.
2. A lei de Hardy-Weinberg ainda é aplicável quando existem mais de dois alelos por *locus*. Se existem n alelos, A_1, A_2, \dots, A_n , com as frequências p_1, p_2, \dots, p_n , então a soma de todas as frequências individuais é igual a 1,0. As frequências de cada um dos genótipos homocigotos são simplesmente o quadrado das frequências dos alelos, e as frequências das diferentes classes de heterocigotos são duas vezes o produto das frequências do primeiro e do segundo alelo. A [Tabela 18.1](#) fornece um exemplo com $p_1 = 0,5$, $p_2 = 0,3$ e $p_3 = 0,2$.
3. A lógica de Hardy-Weinberg também é aplicável aos *loci* ligados ao X. Os homens são hemizigotos em relação aos genes ligados ao X, o que significa que um homem apresenta uma única cópia desses genes. Portanto, em relação

aos genes ligados ao X em homens, as frequências genotípicas são iguais às frequências alélicas. Para as mulheres, as frequências genotípicas em relação aos genes ligados ao X seguem as expectativas normais de Hardy-Weinberg.

A calvície masculina padrão é um traço ligado ao X (Figura 18.9). O gene *AR* (em referência a receptor de andrógeno) é um gene ligado ao X envolvido no desenvolvimento masculino. Existe um haplótipo *AR* denominado *Eur-H1* que está fortemente associado à calvície-padrão. A calvície masculina padrão é comum na Europa, onde o haplótipo *Eur-H1* ocorre a uma frequência de 0,71, o que significa que 71% dos homens europeus o carregam. Com a utilização da lei de Hardy-Weinberg, podemos calcular que 50% das mulheres europeias são homozigotas para o *Eur-H1* e que 41% são heterozigotas. A herança da calvície é complexa e é afetada por múltiplos genes. Assim, nem todos os homens que apresentam o *Eur-H1* se tornam calvos.

4. Pode-se testar se as frequências genotípicas observadas em um *locus* correspondem às previsões de Hardy-Weinberg com a utilização do teste de χ^2 (ver Capítulo 3). Um exemplo é fornecido pelo gene do antígeno leucocitário humano, *HLA-DQA1*, do complexo principal de histocompatibilidade (MHC). O MHC é um agrupamento de genes no cromossomo 6 que desempenham papéis no sistema imune. A Tabela 18.2 apresenta as frequências genotípicas em relação a um SNP (rs9272426) no *HLA-DQA1* para 84 habitantes da Toscana, na Itália. Esse SNP apresenta os alelos *A* e *G*. A partir das frequências genotípicas na Tabela 18.2, podemos calcular as frequências alélicas: $f(A) = p = 0,53$ e $f(G) = q = 0,47$. Em seguida, podemos calcular as frequências genotípicas esperadas sob a lei de Hardy-Weinberg: $p^2 = 0,281$, $2pq = 0,498$ e $q^2 = 0,221$. A multiplicação das frequências genotípicas esperadas pelo tamanho da amostra ($N = 84$) nos fornece o número esperado de indivíduos para cada genótipo. Agora, podemos calcular a estatística do χ^2 como 8,29. Com a utilização da Tabela 3.1, observamos que a probabilidade sob a hipótese nula de que os dados observados correspondam às previsões de Hardy-Weinberg é $P < 0,005$,

com $gl = 1$. (Temos apenas um grau de liberdade, tendo em vista que temos três categorias genotípicas e utilizamos dois números a partir dos dados [N e p] para calcular os valores esperados [$3 - 2$ resulta em 1 grau de liberdade]. Não precisamos utilizar q , tendo em vista que $q = p - 1$.) Essa análise nos conduz à forte suspeita de que os toscanos não estão em conformidade com as expectativas de Hardy-Weinberg em relação ao *HLA-DQA1*. Observaremos adicionalmente a genética de populações do MHC na Seção 18.3 sobre os sistemas de acasalamento e na Seção 18.5 sobre a seleção natural.

Tabela 18.1 Frequências genotípicas de Hardy-Weinberg em relação a um *locus* com três alelos A_1 , A_2 , e A_3 com frequências 0,5, 0,3 e 0,2, respectivamente.

Genótipo	Expectativa	Frequência
A_1A_1	p_1^2	0,25
A_2A_2	p_2^2	0,09
A_3A_3	p_3^2	0,04
A_1A_2	$2 p_1 p_2$	0,30
A_1A_3	$2 p_1 p_3$	0,20
A_2A_3	$2 p_2 p_3$	0,12
Soma		1,00

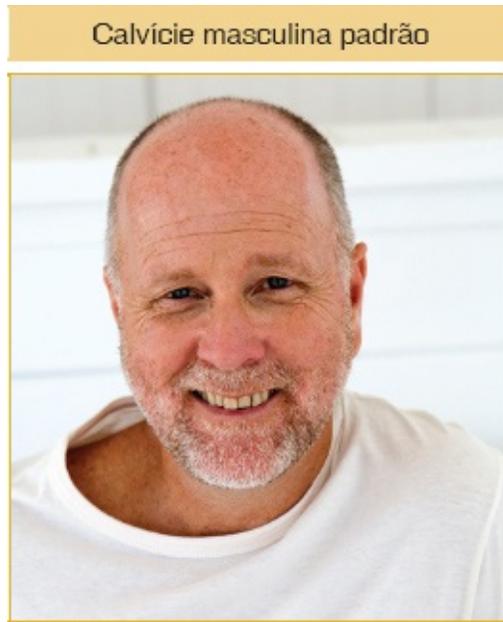


FIGURA 18.9 Indivíduo demonstrando calvície masculina padrão, uma condição ligada ao cromossomo X. (B2M Productions/Getty Images.)

Tabela 18.2 Frequências de genótipos do SNP rs9272426 no *locus HLA-DQA1* do MHC para pessoas da Toscana, na Itália.

	Genótipos			Soma
	A/A	A/G	G/G	
Número observado	17	55	12	84
Frequência observada	0,202	0,655	0,143	1
Frequência esperada	0,281	0,498	0,221	1
Número observado	23,574	41,851	18,574	84

(Observado —
esperado)²/esperado

1,833	4,131	2,327	8,29
-------	-------	-------	------

Fonte: International HapMap Project (www.hapmap.org).

A lei de Hardy-Weinberg é parte do fundamento da genética de populações. Ela se aplica a uma população idealizada de tamanho infinito e na qual o cruzamento é aleatório. Ela também presume que todos os genótipos são igualmente adaptados — ou seja, que todos eles são igualmente viáveis e que apresentam o mesmo sucesso na reprodução. As populações reais se desviam dessa idealizada. No restante do capítulo, examinaremos como fatores tais como o cruzamento não aleatório, o tamanho de população finito e a aptidão desigual de diferentes genótipos causam desvios das expectativas de Hardy-Weinberg. Também veremos como a lei de Hardy-Weinberg pode ser modificada para compensar esses fatores.

CONCEITO-CHAVE A lei de Hardy-Weinberg descreve a relação entre as frequências alélicas e genóticas. Essa lei nos informa que a variação genética não é criada nem destruída pelo processo de transmissão de genes de uma geração para a próxima. A lei de Hardy-Weinberg se aplica estritamente a populações infinitamente grandes e de cruzamento aleatório.

18.3 Sistemas de acasalamento

O cruzamento aleatório é uma suposição crítica da lei de Hardy-Weinberg. A suposição de cruzamento aleatório é atendida se todos os indivíduos na população apresentam igual probabilidade de representar uma escolha quando um parceiro é escolhido. Entretanto, se um parente, vizinho ou indivíduo fenotipicamente semelhante é um parceiro mais provável do que um indivíduo aleatório, então a suposição de cruzamento aleatório foi violada. As populações que não têm cruzamento aleatório não exibirão as proporções de Hardy-Weinberg

exatas em relação aos genótipos em alguns dos genes, ou todos. Três tipos de desvios na escolha do parceiro que violam a suposição de cruzamento aleatório são o cruzamento preferencial, o isolamento pela distância e o endocruzamento.

Cruzamento preferencial

Ocorre o cruzamento preferencial se os indivíduos escolhem parceiros com base na semelhança consigo mesmos. O **cruzamento preferencial positivo** ocorre quando tipos semelhantes acasalam; por exemplo, se indivíduos altos acasalam preferencialmente com outros indivíduos altos e indivíduos baixos acasalam com outros indivíduos baixos. Nesses casos, os genes que controlam a diferença na altura não seguirão a lei de Hardy-Weinberg. Em vez disso, esperaríamos observar um excesso de homozigotos em relação aos alelos “altos” entre a progênie de pares de cruzamento altos e um excesso de homozigotos em relação aos alelos “baixos” entre a progênie de pares de cruzamento baixos. Em seres humanos, ocorre um cruzamento preferencial positivo em relação à altura.

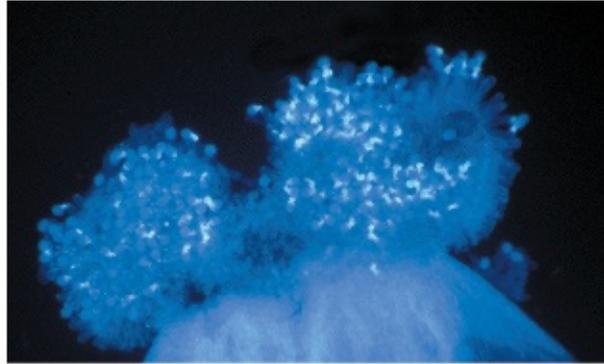
O **cruzamento preferencial negativo** ou **não preferencial** ocorre quando indivíduos diferentes acasalam — ou seja, quando os opostos se atraem. Um exemplo de cruzamento preferencial negativo é fornecido pelo *locus* da autoincompatibilidade, ou *S*, em plantas como *Brassica* (brócolis e seus parentes). Existem diversos alelos no *locus S*, S_1 , S_2 , S_3 e assim por diante. O estigma de uma planta não será receptivo ao pólen que carrega qualquer um de seus dois próprios alelos (Figura 18.10). Por exemplo, o estigma de um heterozigoto S_1/S_2 não possibilitará que os grãos de pólen que carregam um alelo S_1 ou S_2 germinem e fertilizem os seus óvulos, embora os grãos de pólen que carregam os alelos S_3 ou S_4 possam fazer isso. Esse mecanismo bloqueia a autofertilização, forçando, assim, a polinização cruzada. O *locus S* viola a lei de Hardy-Weinberg, uma vez que não são formados genótipos homozigotos para *S*.

Um segundo exemplo de cruzamento preferencial negativo é fornecido pelo complexo principal de histocompatibilidade (MHC), que sabidamente influencia a escolha do parceiro em vertebrados. O MHC afeta o odor corporal em camundongos e ratos, proporcionando uma base para a escolha do parceiro. Nos que são conhecidos como os “experimentos da camiseta suada”, pesquisadores

solicitaram a um grupo de homens que vestissem camisetas durante 2 dias. Em seguida, eles solicitaram a um grupo de mulheres que sentissem o odor das camisetas e as classificassem em relação à “preferência”. As mulheres preferiram o odor de homens cujos haplótipos do MHC eram diferentes dos seus próprios. Os dados do projeto HapMap humano confirmaram que os casais americanos são significativamente mais heterozigotos quanto ao MHC do que o esperado ao acaso. O MHC desempenha um papel central na nossa resposta imune aos patógenos, e os heterozigotos podem ser mais resistentes aos patógenos. Portanto, a nossa descendência se beneficia se acasalamos de modo não preferencial em relação ao nosso genótipo MHC. Esse mecanismo pode explicar o motivo de o SNP no gene *HLA-DQA1* de MHC a respeito do qual discutimos anteriormente não seguir a lei de Hardy-Weinberg entre os residentes da Toscana. Observe novamente a [Tabela 18.2](#) e você notará que existem mais heterozigotos do que o esperado, 55 *versus* 42. Os toscanos aparentam estar praticando o cruzamento não preferencial em relação a esse SNP.

A autoincompatibilidade leva ao cruzamento preferencial negativo em *Brassica*

A. Inibição do pólen



B. Crescimento do tubo de pólen

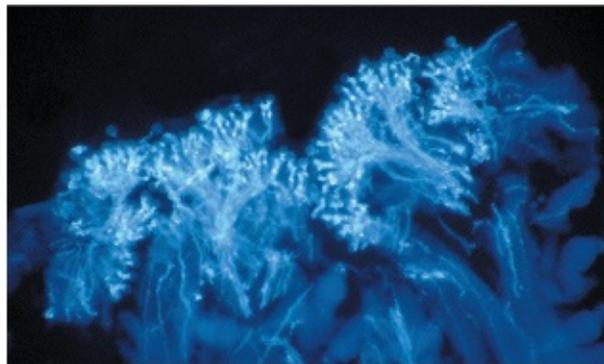


FIGURA 18.10 Cruzamento preferencial negativo causado pelo *locus* de autoincompatibilidade (*S*) do gênero de plantas florescentes *Brassica*. **A.** Um estigma autopolinizado S_1/S_2 demonstra ausência de crescimento do tubo de pólen. **B.** Ocorre o crescimento do tubo de pólen para um estigma S_1/S_2 submetido à polinização cruzada com o pólen de um heterozigoto S_3/S_4 . (June Bowman Nasrallah.)

Isolamento pela distância

Outro tipo de desvio na escolha do parceiro tem origem a partir da distância geográfica entre os indivíduos. Os indivíduos são mais aptos a acasalarem com um vizinho do que com outro membro da sua espécie do lado oposto do continente — ou seja, os indivíduos podem demonstrar **isolamento pela distância**. Conseqüentemente, as frequências alélicas e genóticas com frequência são diferentes entre peixes em lagos separados ou entre pinheiros em diferentes

regiões de um continente. Diz-se que as espécies ou as populações que exibem tal padrão de variação genética demonstram **estrutura populacional**. Uma espécie pode ser dividida em uma série de subpopulações, tais como as rãs em diferentes lagoas, ou as pessoas em diferentes cidades.

Se uma espécie apresenta estrutura populacional, a proporção de homozigotos será maior na espécie como um todo do que o esperado sob a lei de Hardy-Weinberg. Considere um exemplo hipotético de uma espécie de girassóis selvagens distribuída por todo o estado do Kansas com um gradiente na frequência do alelo *A* de 0,9 perto de Kansas City a 0,1 perto de Elkhart ([Figura 18.11 A](#)). Obtivemos amostras de 100 plantas de girassol de cada uma dessas duas cidades, mais 100 de Hutchinson, no centro do estado, e calculamos as frequências alélicas. Cada cidade representa uma subpopulação. Em relação a qualquer uma das três cidades, a lei de Hardy-Weinberg funciona bem. Por exemplo, em Elkhart, esperamos que $Nq^2 = 100 \times (0,9)^2 = 81$ homozigotos *a/a*, e isso é o que observamos. Entretanto, para todo o estado, preveríamos que $Nq^2 = 300 \times (0,5)^2 = 75$ homozigotos *a/a*, mas observamos 107. Em virtude da estrutura populacional, existem mais plantas de girassol homozigotas do que o esperado.

Número de indivíduos

	<i>N</i>	<i>A/A</i>	<i>A/a</i>	<i>a/a</i>	<i>p</i>	<i>q</i>
Kansas City	100	81	18	1	0,90	0,10
Hutchinson	100	25	50	25	0,50	0,50
Elkhart	100	1	18	81	0,10	0,90
Todo o						

estado (observado)	300	107	86	107	0,50	0,50
Todo o estado (esperado)	300	75	150	75	—	—

Eis um exemplo real de estrutura populacional de nossa própria espécie. Na África, o alelo FY^{nulo} do grupo sanguíneo Duffy demonstra um gradiente com uma baixa frequência no leste e no norte da África uma frequência moderada no sul da África, e uma alta frequência ao longo da África central ([Figura 18.11 B](#)). Esse alelo é raro fora da África. Em virtude desse gradiente, não podemos utilizar as frequências alélicas globais na África para calcular as frequências genotípicas com a utilização da lei de Hardy-Weinberg. Posteriormente neste capítulo e no [Capítulo 20](#), discutiremos a relação entre o FY^{nulo} e a malária.

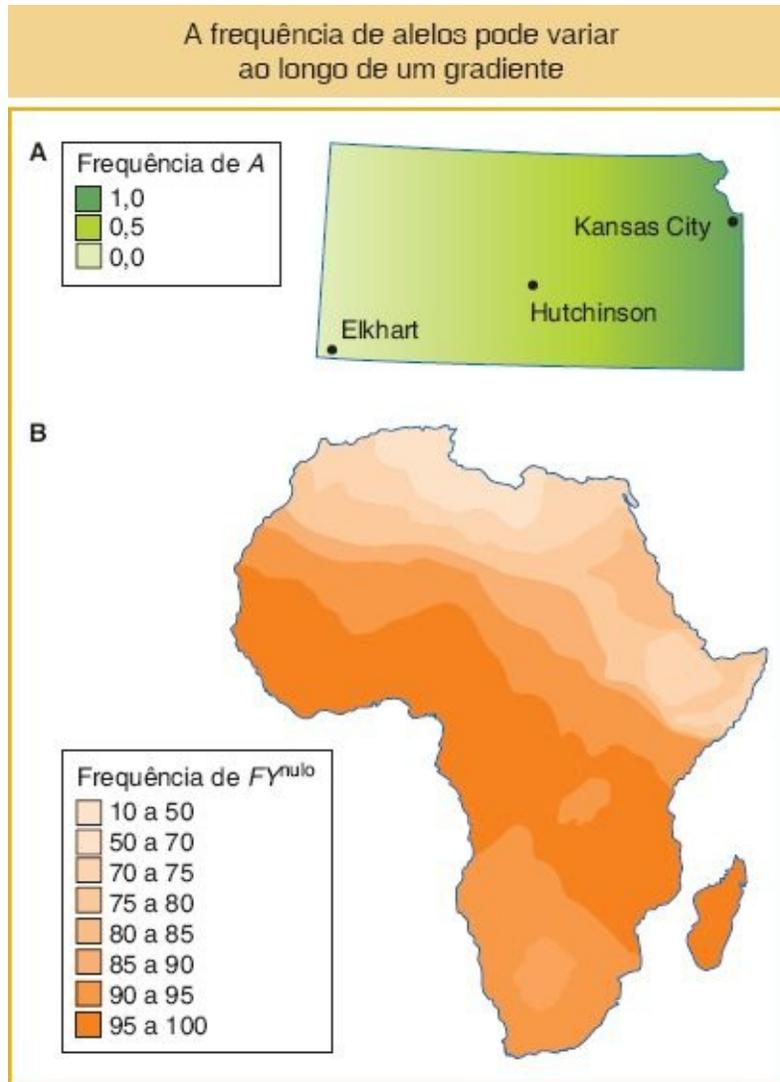


FIGURA 18.11 A. Variação da frequência alélica em todo o Kansas em relação a uma espécie hipotética de girassol selvagem. **B.** Variação da frequência em relação ao alelo FY^{nu0} do locus do grupo sanguíneo Duffy na África. (Dados de P. C. Sabeti *et al.*, *Science* 312, 2006, 1614-1620.)

CONCEITO-CHAVE O cruzamento preferencial e o isolamento pela distância violam a lei de Hardy-Weinberg e podem fazer com que as frequências genotípicas desviem-se das expectativas de Hardy-Weinberg.

Endocruzamento

O terceiro tipo de desvio no acasalamento é o **endocruzamento**, ou cruzamento

entre parentes. Muito antes que qualquer pessoa soubesse a respeito dos alelos recessivos deletérios, algumas sociedades reconheciam que distúrbios como mudez, surdez e cegueira eram mais comuns entre os filhos de casamentos consanguíneos. Por isso, os casamentos de irmãos e irmãs e primos em primeiro grau são ilegais ou desencorajados. Apesar disso, muitos indivíduos famosos se casaram com primos, incluindo Charles Darwin, Albert Einstein, J. S. Bach, Edgar Allan Poe, Jesse James e a Rainha Victoria. Conforme veremos, a descendência dos casamentos consanguíneos tem risco mais alto de apresentar um distúrbio hereditário.

A progênie de endocruzamento apresenta maior probabilidade de ser homozigota em qualquer *locus* do que a progênie de cruzamentos não consanguíneos. Portanto, ela apresenta maior probabilidade de ser homozigota em relação a alelos recessivos deletérios. Por esse motivo, o endocruzamento pode levar a uma redução no vigor e no sucesso reprodutivo, denominada **depressão por endocruzamento**. Entretanto, o endocruzamento também apresenta vantagens. Em muitas espécies de plantas observa-se alta incidência de autopolinização e endogamia. Essas incluem a planta-modelo *Arabidopsis*, uma erva daninha de sucesso, e os cultivos de cereais produtivos de arroz e de trigo. Tendo em vista que a maior parte das espécies de plantas contém órgãos masculinos e femininos no mesmo indivíduo, a autopolinização pode ser conquistada mais facilmente do que o cruzamento externo. Outra vantagem da autopolinização é que, quando uma única semente é dispersa em um novo local, a planta que cresce a partir da semente apresenta um parceiro pronto — ela própria, o que possibilita que uma nova população seja estabelecida a partir de uma única semente. Finalmente, se uma planta individual apresenta uma combinação benéfica de alelos em diferentes *loci*, o endocruzamento então preserva aquela combinação. Nas espécies de plantas endogâmicas, benefícios como esses oferecem vantagens que superam o custo associado à depressão por endocruzamento.

Coeficiente de endocruzamento

O endocruzamento aumenta o risco de que um indivíduo seja homozigoto em relação a um alelo recessivo deletério e exiba uma doença genética. O aumento

desse risco depende de dois fatores: (1) frequência do alelo deletério na população e (2) grau de endocruzamento. Para medir o grau de endocruzamento, os geneticistas utilizam o **coeficiente de endocruzamento** (F), que é a probabilidade de que dois alelos em um indivíduo remontem à mesma cópia em um ancestral comum. Consideraremos primeiramente como calcular F com a utilização de heredogramas e em seguida examinaremos como F pode ser utilizado para determinar o aumento no risco de herdar uma doença recessiva.

Considere um heredograma simples em relação a um cruzamento entre meios-irmãos, indivíduos que apresentam um genitor em comum ([Figura 18.12 A](#)). Na figura, B e C são meios-irmãos que apresentam a mesma mãe, A, mas pais diferentes; B e C têm uma filha, I. Observe que existe uma alça fechada de I para B e A e de volta para I por C. A presença de uma alça fechada no heredograma nos informa que I é resultante de uma união consanguínea. As duas cópias do gene em A estão coloridas em azul e cor-de-rosa — o azul do pai de A e o cor-de-rosa de sua mãe. Conforme desenhado, I herdou a cópia cor-de-rosa tanto através de seu pai (B) quanto de sua mãe (C). Tendo em vista que as duas cópias do gene de I remontam à mesma cópia em sua avó, as suas duas cópias são **idênticas por descendência (IBD)**. Em termos mais gerais, se duas cópias de um gene em um indivíduo remontam à mesma cópia em um ancestral, então as cópias são IBD. Gostaríamos de ter um modo para calcular a probabilidade de que os dois alelos de I sejam IBD. Essa probabilidade é o coeficiente de endocruzamento para I, que é, na forma de símbolo, F_I .

Primeiramente, tendo em vista que estamos interessados apenas em traçar o trajeto dos alelos IBD, podemos simplificar o heredograma para que ele contenha apenas os indivíduos na alça fechada e ainda seguir a transmissão de quaisquer alelos IBD ([Figura 18.12 B](#)). Além disso, tendo em vista que o sexo do indivíduo não importa, utilizamos círculos para ambos os sexos. Os alelos transmitidos a cada cruzamento são rotulados w , x , y e z . Utilizamos “ \sim ” para simbolizar IBD. Gostaríamos de calcular a probabilidade de que w e x sejam IBD, mas faremos esse cálculo passo a passo. Primeiramente, qual é a probabilidade de que x e y sejam IBD ou, simbolicamente, qual é $P(x \sim y)$? Essa é a probabilidade de C transmitir a cópia herdada de A para I, que é $1/2$, ou $P(x \sim y) = 1/2$. De modo

semelhante, a probabilidade de B transmitir a cópia herdada de A para I é $1/2$, ou $P(w \sim z) = 1/2$.

Agora precisamos calcular a probabilidade de que z e y sejam IBD. Existem dois modos por meio dos quais z e y podem ser IBD. O primeiro modo é quando z e y são ambos a mesma cópia (ambas cor-de-rosa ou ambas azuis). Isso ocorre em $1/2$ das ocasiões, tendo em vista que em $1/4$ das ocasiões elas são ambas azuis e em $1/4$, ambas cor-de-rosa. O segundo modo é quando z e y são cópias diferentes (uma cor-de-rosa e a outra azul), mas o indivíduo A foi endocruzado. Se o indivíduo A é endocruzado, existe então uma probabilidade de que as suas duas cópias do gene sejam IBD. A probabilidade de que as duas cópias de A sejam IBD é o coeficiente de endocruzamento de A, F_A . A probabilidade de que z e y sejam cópias diferentes (uma cor-de-rosa, a outra azul) é de $1/2$. Assim, a probabilidade de que z e y sejam cópias diferentes que são IBD é $1/2$, multiplicada pelo coeficiente de endocruzamento (F_A) para fornecer $\frac{1}{2} F_A$. Em conjunto, a probabilidade de que z e y sejam IBD é a probabilidade de que eles sejam a mesma cópia ($1/2$) mais a probabilidade de que eles sejam cópias diferentes que são IBD ($\frac{1}{2} F_A$). Simbolicamente, escrevemos:

Os heredogramas demonstram quando os genes são idênticos por descendência

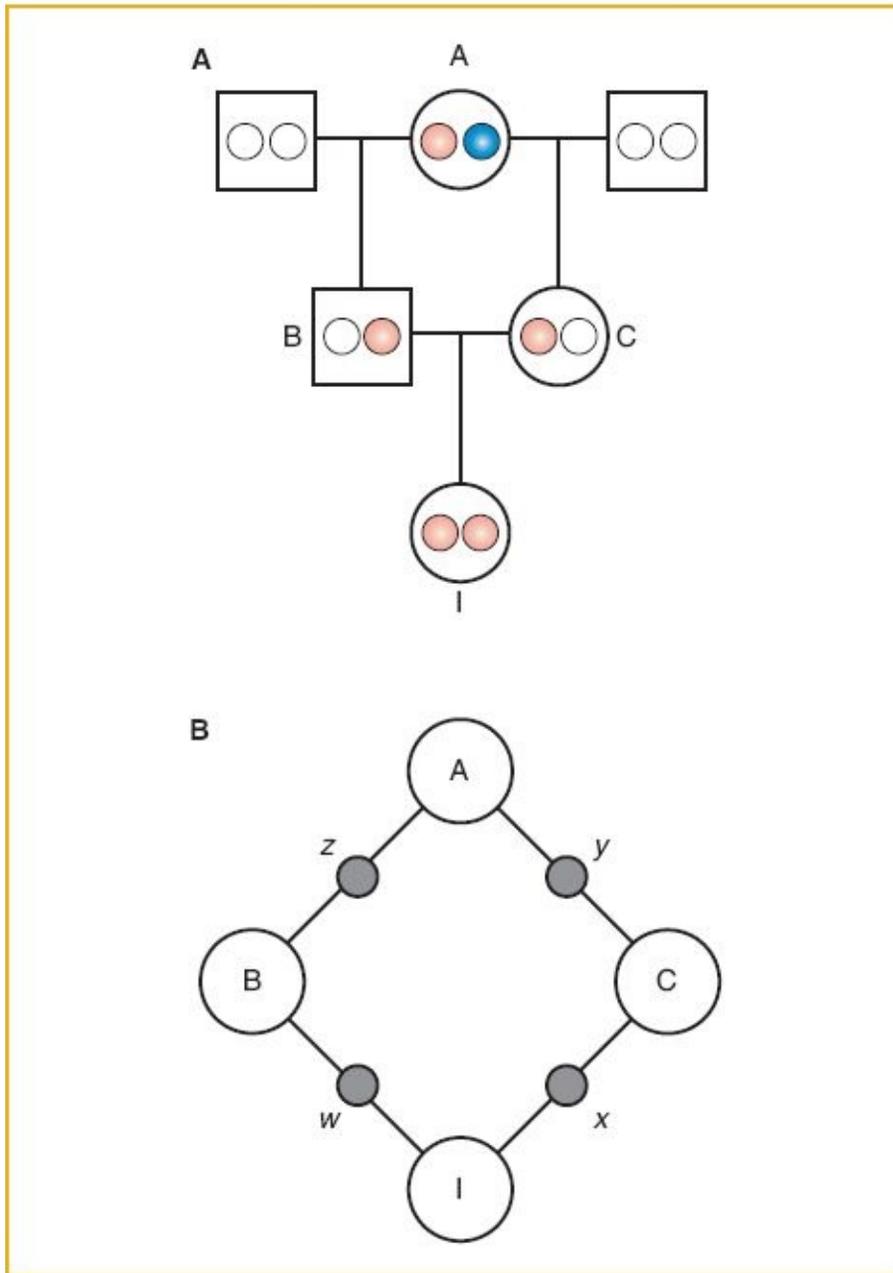


FIGURA 18.12 A. Heredograma em relação a um cruzamento entre meios-irmãos desenhado no formato-padrão. As *pequenas bolas coloridas* representam uma única cópia de um gene. No indivíduo A, as cópias *cor-de-rosa* e *azul* representam as cópias do gene que ela herdou de sua mãe e de seu pai, respectivamente. **B.** Heredograma em relação a um cruzamento entre meios-irmãos desenhado no formato simplificado utilizado para a análise de endocruzamento. Apenas as linhas que conectam o genitor à descendência estão desenhadas e apenas os indivíduos na “alça de endocruzamento fechada” estão incluídos. *w*, *x*, *y* e *z* são símbolos para o alelo transmitido do genitor para a descendência.

$$P(z \sim y) = + F_A$$

$P(x \sim y)$, $P(w \sim z)$ e $P(z \sim y)$ são probabilidades independentes; portanto, podemos utilizar a regra do produto e reunir tudo para obter:

$$\begin{aligned} F_I &= P(x \sim y) \times P(w \sim z) \times P(z \sim y) \\ &= \frac{1}{2} \times \frac{1}{2} \times \left(\frac{1}{2} + \frac{1}{2}F_A\right) \\ &= \left(\frac{1}{2}\right)^3 (1 + F_A) \end{aligned}$$

Na análise de heredogramas consanguíneos, podemos substituir o valor de F_A na equação anterior, se ele for conhecido. De outro modo, podemos presumir que F_A é zero, se não houver informação para sugerir que o indivíduo A seja endocruzado. No exemplo atual, se presumirmos que $F_A = 0$, então:

$$F_I = \left(\frac{1}{2}\right)^3 = \frac{1}{8}$$

Esse cálculo nos informa que a descendência de cruzamentos entre meios-irmãos será homozigota em relação aos alelos que são IBD em relação a no mínimo 1/8 de seus genes. Ela pode ser superior a 1/8 se F_A for superior a zero. Heredogramas de endocruzamentos adicionais e uma fórmula geral para o cálculo de F podem ser encontrados no [Quadro 18.2](#).

Quando existe um endocruzamento em uma população, a presunção de cruzamento aleatório de Hardy-Weinberg será violada. Entretanto, a lei de Hardy-Weinberg pode ser modificada para corrigir as proporções genotípicas previstas em relação a diferentes graus de endocruzamento por meio da utilização de F , o coeficiente de endocruzamento médio para a população. As frequências de Hardy-Weinberg modificadas são:

$$\begin{aligned} f_{AA} &= p^2 + pqF \\ f_{Aa} &= 2pq - 2pqF \\ f_{aa} &= q^2 + pqF \end{aligned}$$

Essas proporções de Hardy-Weinberg modificadas fazem sentido intuitivamente, demonstrando como o endocruzamento reduz a frequência de heterozigotos em

$2pqF$ e adiciona metade dessa quantidade a cada uma das classes homozigotas. Com essas equações de Hardy-Weinberg modificadas, você também observará que quando não há endocruzamento ($F = 0$), você obtém novamente as frequências genóticas de Hardy-Weinberg padrão e, quando existe um endocruzamento completo ($F = 1$), você obtém $f_{AA} = p$ e $f_{aa} = q$.

Em quanto o endocruzamento aumenta o risco de que aquela descendência exiba uma condição de doença recessiva? A [Tabela 18.3](#) demonstra os coeficientes de endocruzamento em relação à descendência e alguns diferentes cruzamentos consanguíneos e o número previsto de homozigotos recessivos em relação a diferentes frequências (q) do alelo recessivo. Quando $q = 0,01$, existe um aumento de 7 vezes (7,19/1,0) na descendência homozigota recessiva para cruzamentos entre primos de primeiro grau, em comparação aos cruzamentos entre indivíduos não relacionados. O aumento no risco aumenta para 13 vezes (3,36/0,25) quando $q = 0,005$ e para 63 vezes (0,63/0,01) quando $q = 0,001$. Em outras palavras, o risco aumenta dramaticamente em relação aos alelos raros. Cruzamentos entre irmão e irmã e entre genitores e filhos são os mais arriscados: quando $q = 0,001$, eles demonstram um risco 250 vezes (2,51/0,01) maior em comparação aos cruzamentos entre indivíduos não relacionados.

Quadro 18.2

Cálculo dos coeficientes de endocruzamento a partir de heredogramas.

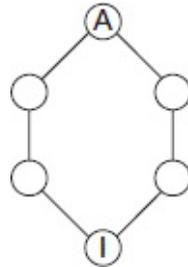
No texto, vimos que o coeficiente de endocruzamento (F_I) em relação à descendência de um cruzamento entre meios-irmãos é:

$$F_I = \left(\frac{1}{2}\right)^3 (1 + F_A)$$

em que F_A é o coeficiente de endocruzamento do ancestral. Essa expressão inclui o termo $1/2$ elevado à terceira potência, $(1/2)^3$. Na [Figura 18.12](#), você verá que existem três indivíduos na alça de endocruzamento, sem contar I. A fórmula geral para computar os coeficientes de endocruzamento a partir de heredogramas é:

$$F_I = \left(\frac{1}{2}\right)^n (1 + F_A)$$

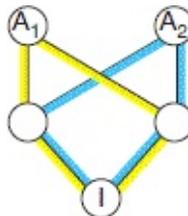
em que n é o número de indivíduos na alça de endocruzamento, sem contar I. Vejamos outro heredograma, no qual os avós de I são meios-irmãos:



Existem cinco indivíduos na alça de endocruzamento além de I; assim, se presumirmos que o ancestral não foi endocruzado ($F_A = 0$), então:

$$F_I = \left(\frac{1}{2}\right)^5 (1 + F_A) = 0,03125$$

Em alguns heredogramas, existe mais de uma alça de endocruzamento. Aqui está um heredograma no qual I é a descendência de um cruzamento entre irmãos legítimos:



Em relação aos heredogramas com diversas alças de endocruzamento, soma-se a contribuição em todas as alças em que F_A é o coeficiente de endocruzamento do ancestral (A) da alça em questão:

$$F_I = \sum_{\text{alças}} \left(\frac{1}{2}\right)^n (1 + F_A)$$

Portanto, em relação ao heredograma no qual I é a descendência de um cruzamento entre irmãos legítimos, obtemos:

$$F_I = \left(\frac{1}{2}\right)^3 (1 + F_{A1}) + \left(\frac{1}{2}\right)^3 (1 + F_{A2}) = \frac{1}{4}$$

presumindo que os coeficientes de endocruzamento em relação a ambos os ancestrais sejam 0.

Tabela 18.3 Número de homozigotos recessivos por 10.000 indivíduos em relação a diferentes frequências alélicas (q).

Cruzamento	F	$q = 0,01$	$q = 0,005$	$q = 0,001$
Genitores não relacionados	0,0	1,00	0,25	0,01
Genitor e filho ou irmão e irmã	1/4	25,75	12,69	2,51
Meios-irmãos	1/8	13,38	6,47	1,26
Primos de primeiro grau	1/16	7,19	3,36	0,63
Primos de segundo grau	1/64	2,55	1,03	0,17

O impacto do endocruzamento sobre a frequência de distúrbios genéticos em populações humanas pode ser observado na [Figura 18.13](#). Os filhos de casamentos de primos em primeiro grau demonstram uma frequência duas vezes mais alta de distúrbios em comparação aos filhos de genitores não relacionados.

Registros históricos sugerem que os riscos do endocruzamento eram compreendidos muito antes da existência da genética.

Tamanho da população e endocruzamento

O tamanho da população é um fator de contribuição importante para o nível de endocruzamento nas populações. Em populações pequenas, os indivíduos apresentam maior probabilidade de cruzar com um parente do que em populações maiores. O fenômeno é observado em pequenas populações humanas, tais como a das Ilhas Tristão da Cunha, no Atlântico Sul, que tem menos de 300 pessoas. Vejamos o efeito do tamanho da população sobre o nível geral de endocruzamento em uma população, conforme medido por F .

Considere uma população com F_t sendo o nível de endocruzamento na geração t . Para formar um indivíduo na próxima geração $t + 1$, selecionamos o primeiro alelo do *pool* gênico. Suponha que o tamanho da população é N . Após a seleção do primeiro alelo, a probabilidade de que o segundo alelo que coletamos seja exatamente a mesma cópia é $1/2N$ e o coeficiente de endocruzamento em relação a esse indivíduo é 1,0. A probabilidade de que o segundo alelo que coletamos seja uma cópia diferente do primeiro alelo é $1 - 1/2N$ e o nível de endocruzamento em relação ao indivíduo resultante seria F_t , o coeficiente de endocruzamento médio em relação à população inicial na geração t . O nível de endocruzamento na próxima geração é a soma desses dois desfechos possíveis, ou:

$$F_{t+1} = \left(\frac{1}{2N}\right)1 + \left(1 - \frac{1}{2N}\right)F_t$$

O endocruzamento leva a um aumento nos distúrbios genéticos recessivos

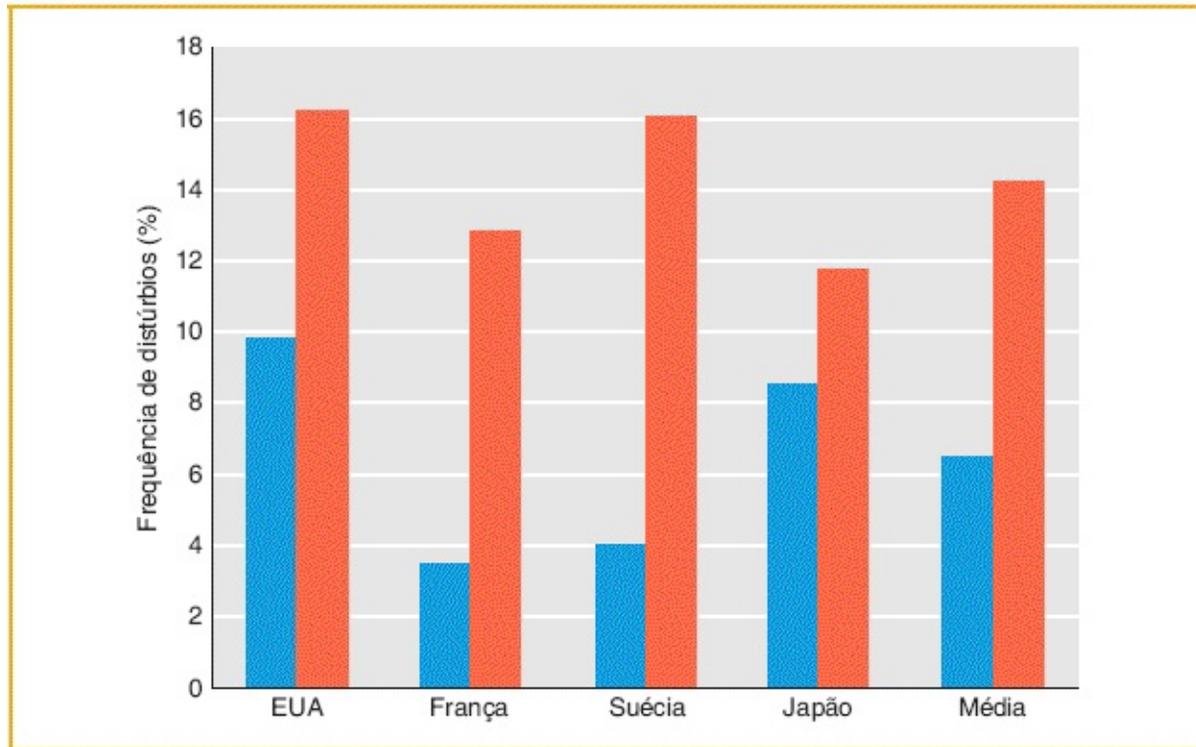


FIGURA 18.13 A frequência de distúrbios genéticos entre filhos de genitores não relacionados (colunas azuis) em comparação àquela de filhos de genitores que são primos de primeiro grau (colunas vermelhas). (Dados de C. Stern, *Principles of Human Genetics*, W. H. Freeman, 1973.)

Essa equação nos informa que a F aumentará ao longo do tempo como uma função do tamanho da população. Quando N é grande, F aumenta lentamente ao longo do tempo. Quando N é pequeno, F aumenta rapidamente ao longo do tempo. Por exemplo, suponha que F_t na população inicial seja 0,1 e $N = 10.000$. Em seguida, F_{t+1} seria 0,10005, um valor apenas ligeiramente mais alto. Entretanto, se $N = 10$, então F_{t+1} seria 0,145, um valor muito mais alto. Também podemos utilizar essa equação recursivamente para calcular F_{t+2} por meio da utilização de F_{t+1} em vez de F_t do lado direito. O resultado com $N = 10$ e $F_t = 0,1$ será $F_{t+2} = 0,188$. Os efeitos do tamanho da população sobre o endocruzamento em populações são adicionalmente explorados no [Quadro 18.3](#).

Uma consequência do aumento de endocruzamento é que os indivíduos em populações pequenas apresentam maior probabilidade de ser homozigotos em relação a alelos deletérios, assim como a descendência de casamentos entre

primos em primeiro grau apresenta maior probabilidade de ser homozigota em relação a tais alelos. Esse efeito é observado em grupos étnicos que vivem em pequenas comunidades isoladas em termos reprodutivos. Por exemplo, um tipo de nanismo no qual os indivíduos afetados apresentam seis dedos ocorre com uma frequência superior a 1 em 200 em uma população de aproximadamente 13.000 Amish no Condado de Lancaster, na Pensilvânia, embora sua frequência na população norte-americana em geral seja de apenas 1 em 60.000.

Quadro 18.3

Endocruzamento em populações finitas.

No texto principal, derivamos a fórmula em relação ao aumento no endocruzamento entre gerações em populações finitas como:

$$F_{t+1} = \left(\frac{1}{2N}\right)1 + \left(1 - \frac{1}{2N}\right)F_t$$

que pode ser reescrita como:

$$(1 - F_{t+1}) = \left(1 - \frac{1}{2N}\right)(1 - F_t)$$

Também apresentamos a fórmula em relação à frequência de heterozigotos (H) com endocruzamento como:

$$H = f_{A/a} = 2pq - 2pqF$$

que pode ser reescrita como:

$$(1 - F) = H/2pq$$

Combinando essas duas equações, obtemos:

$$H_{t+1}/2pq = \left(1 - \frac{1}{2N}\right)H_t/2pq$$

e em seguida:

$$H_{t+1} = \left(1 - \frac{1}{2N}\right)H_t$$

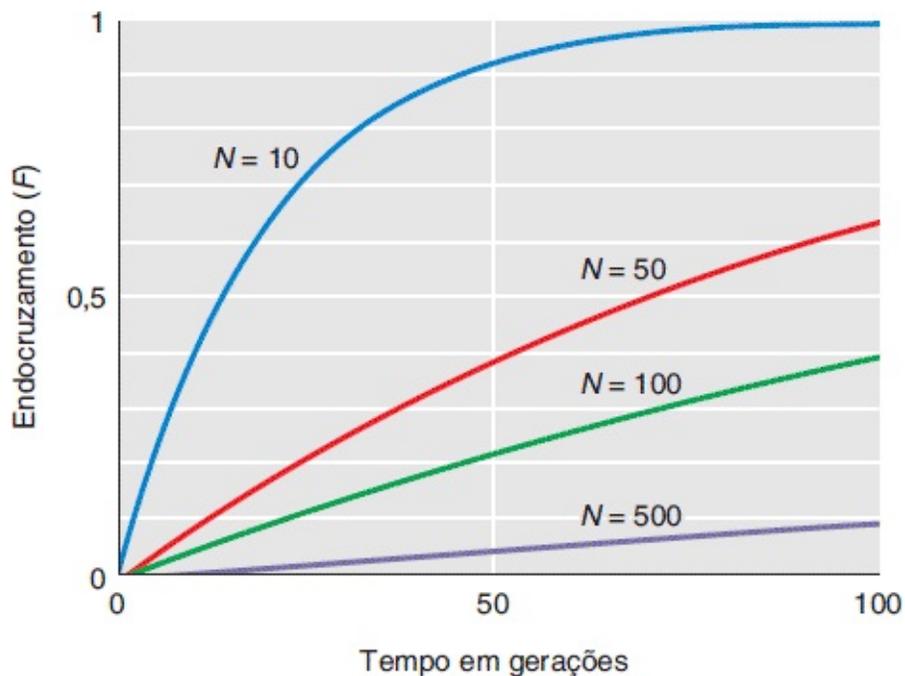
Portanto, em relação a cada geração, o nível de heterozigosidade é reduzido pela fração $(1 - 1/2N)$. A redução em H ao longo de t gerações é:

$$H_t = \left(1 - \frac{1}{2N}\right)^t H_0$$

e a alteração em F ao longo das gerações t é fornecida por:

$$F_t = 1 - \left(1 - \frac{1}{2N}\right)^t (1 - F_0)$$

Conforme demonstrado na figura a seguir, o endocruzamento aumentará com o tempo em uma população finita, até mesmo quando não houver endocruzamento na população inicial.



Aumento no endocruzamento (F) ao longo do tempo em relação a diversos tamanhos de populações diferentes.

CONCEITO-CHAVE O endocruzamento aumenta a frequência de homozigotos em uma população, e pode resultar em uma frequência mais alta de distúrbios genéticos recessivos. O coeficiente de endocruzamento (F) é a probabilidade de que dois alelos em um indivíduo remontem à mesma cópia em um ancestral comum.

18.4 Variação genética e sua medida

Para estudar a quantidade e a distribuição da variação genética em populações, precisamos de alguns modos para quantificar a variação. Para descrever como podemos quantificar a variação, utilizaremos dados em relação ao gene da glicose-6-fosfato desidrogenase (*G6PD*) de seres humanos. O *G6PD* é um gene ligado ao X que codifica uma enzima que catalisa uma etapa na glicólise. O alelo do tipo selvagem (B) de *G6PD* apresenta atividade enzimática total. Um segundo alelo, denominado A^- , leva à atividade enzimática fortemente reduzida, e os indivíduos que carregam esse alelo desenvolvem anemia hemolítica. Entretanto, esse alelo também confere uma redução de 50% no risco de malária grave nos portadores. Em regiões da África nas quais a malária é endêmica, o alelo A^- alcança frequências próximas de 20%, embora esse alelo esteja ausente ou seja raro em outros locais. Outro alelo (A^+) leva a uma atividade enzimática apenas modestamente reduzida. Contrariamente aos indivíduos que carregam o alelo A^- , os indivíduos que carregam apenas os alelos A^+ ou B não desenvolvem anemia hemolítica.

A [Figura 18.14](#) demonstra os SNP em 18 sítios polimórficos que foram identificados por meio do sequenciamento de um segmento de 5.102 pb do *G6PD* a partir de uma amostra mundial de 47 homens. Os 5.084 sítios remanescentes eram **fixado**, ou invariantes: existe apenas um único alelo (nucleotídeo) na amostra inteira em relação a cada um desses sítios. Ao fazer amostragem apenas com homens, observamos apenas um alelo e um haplótipo para cada indivíduo, tendo em vista que o gene é ligado ao X. O alelo A^+ difere de B pela substituição de um único aminoácido (ácido aspártico em lugar de asparagina) no *SNP3* na

Figura 18.14. O alelo A^- difere do alelo B em dois aminoácidos: ele contém ambas a substituição do “ácido aspártico em lugar da asparagina” observada no alelo A^+ e uma diferença em um segundo aminoácido (metionina em lugar de valina) no *SNP2*.

Como podemos quantificar a variação no *locus G6PD*? Uma medida simples é o número de **sítios segregantes (S)** ou polimórficos. Em relação aos dados de *G6PD*, S é 18 na amostra total, 14 na amostra africana e 7 na amostra não africana. Os africanos contêm o dobro do número de sítios segregantes, apesar do fato de nossa amostra apresentar menos africanos. Outra medida simples é o **número de haplótipos (NH)**. O valor da NH é 12 na amostra total, 9 na amostra africana e 6 na amostra não africana. Novamente, a amostra africana apresenta maior variação. Uma falha de medidas tais como S e NH é que os valores que observamos dependem fortemente do tamanho da amostra. Se temos amostras com mais indivíduos, então os valores de S e NH tendem a aumentar. Por exemplo, nossa amostra apresenta 16 africanos em comparação a 31 não africanos. Embora S tenha o dobro do tamanho em africanos do que em não africanos, a diferença provavelmente seria ainda maior se apresentássemos um número igual (31) de africanos e não africanos.

Em lugar de S e NH , podemos calcular as frequências alélicas, que não são influenciadas por diferenças no tamanho da amostra. Em relação aos dados de *G6PD*, B , A^- e A^+ apresentam frequências mundiais de 0,83, 0,13 e 0,04, respectivamente. Entretanto, você observará que A^- apresenta uma frequência de 0,0 fora da África e de 0,38 em nossa amostra africana, que é uma diferença substancial. Podemos utilizar os dados da frequência alélica para calcular uma estatística denominada **diversidade gênica (GD)**, que é a probabilidade de que dois alelos coletados aleatoriamente do *pool* gênico sejam diferentes. A probabilidade de coleta de dois alelos diferentes é igual a 1 menos a probabilidade de coleta de duas cópias do mesmo alelo somada a todos os alelos do *locus*. Portanto:

$$\begin{aligned}GD &= 1 - \sum p_i^2 \\ &= 1 - (p_1^2 + p_2^2 + p_3^2 + \dots + p_n^2)\end{aligned}$$

em que p_i é a frequência do $i^{\text{ésimo}}$ alelo e Σ é o sinal de somatório, indicando que adicionamos os quadrados de todos os n valores observados de p para $i = 1, 2, \dots, n$, até o $n^{\text{ésimo}}$ alelo. O valor de GD pode variar de 0 a 1. Ele se aproximará de 1 quando houver um grande número de alelos de frequências aproximadamente iguais. Será 0 quando houver um único alelo, e estará próximo de 0 sempre que houver um único alelo muito comum com uma frequência de 0,99 ou superior. A [Tabela 18.4](#) demonstra que a diversidade gênica é razoavelmente alta em africanos (0,47). Tendo em vista que os não africanos apresentam apenas o alelo B , a diversidade gênica é 0,0.

O valor de GD é igual à proporção esperada de heterozigotos sob o equilíbrio de Hardy-Weinberg, **heterozigosidade (H)**. Entretanto, H como um conceito é aplicável apenas aos diploides e não se aplicaria aos *loci* ligados ao X em indivíduos do sexo masculino. Portanto, conceitualmente a diversidade gênica (GD) é mais apropriada até mesmo se for matematicamente a mesma quantidade de H para populações de diploides sob o equilíbrio de Hardy-Weinberg.

A diversidade gênica pode ser calculada em relação a um sítio de nucleotídeo único. Ela pode ser calculada para todos os sítios de nucleotídeos em um gene, caso em que é denominada **diversidade nucleotídica**. Tendo em vista que a maioria dos nucleotídeos em quaisquer duas cópias de um gene de uma espécie é tipicamente a mesma, os valores para a diversidade nucleotídica em relação aos genes são tipicamente muito pequenos. Em relação à $G6PD$, existem apenas 18 sítios de nucleotídeos polimórficos, mas 5.084 sítios invariantes. A diversidade nucleotídica média para toda a sequência do gene $G6PD$ é de 0,0008 em africanos, 0,0002 em não africanos, e 0,0006 na amostra total. Esses valores nos informam que os africanos apresentam quatro vezes mais diversidade nucleotídica em $G6PD$ do que os não africanos.

Tabela 18.4 Dados de diversidade em relação à glicose-6-fosfato desidrogenase ($G6PD$) em seres humanos.

	Amostra total	Africanos	Não africanos
--	---------------	-----------	---------------

Tamanho da amostra	47	16	31
Número de sítios segregantes	18	14	7
Número de haplótipos	12	9	6
Diversidade gênica (<i>GD</i>) no <i>SNP2</i>	0,22	0,47	0,00
Diversidade de nucleotídios	0,0006	0,0008	0,0002

Varição de nucleotídios no gene *G6PD* em seres humanos

Indivíduo	Origem	Alelo	SNP																		Haplótipo
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
			A	G	A	C	C	G	C	C	C	C	C	G	G	C	T	C	A	C	
1	África Austral	A-	G	A	G	.	G	.	.	T	.	T	C	.	G	.	1
2	África Central	A-	G	A	G	.	G	.	.	T	.	T	C	.	G	.	1
3	África Central	A-	G	A	G	.	G	.	.	T	.	T	C	.	G	.	1
4	Afro-americano	A-	G	A	G	.	G	.	.	T	.	T	C	.	G	.	1
5	Afro-americano	A-	G	A	G	.	G	.	.	T	.	T	C	.	G	.	1
6	África Central	A-	G	A	G	.	G	.	.	T	.	T	C	.	G	.	1
7	África Central	A+	G	.	G	T	C	.	G	.	2
8	África Central	A+	G	.	G	T	C	.	G	.	2
9	África Central	B	C	.	G	.	3
10	África Austral	B	A	.	.	C	T	G	.	4
11	África Austral	B	A	.	.	C	T	G	.	4
12	África Austral	B	.	.	.	T	C	T	G	.	5
13	África Austral	B	C	T	G	.	6
14	África Austral	B	T	.	A	.	C	.	G	.	7
15	África Central	B	T	C	.	G	.	8
16	Europa	B	T	C	.	G	.	8
17	Europa	B	T	C	.	G	.	8
18	Europa	B	T	C	.	G	.	8
19	Sudoeste Asiático	B	T	C	.	G	.	8
20	Ásia Oriental	B	C	.	G	.	3
21	Nativo americano	B	A	T	C	.	G	.	9
22	África Austral	B	10
23	Nativo-americano	B	10
24	Nativo-americano	B	10
25	Nativo-americano	B	10
26	Nativo-americano	B	10
27	Nativo-americano	B	10
28	Nativo-americano	B	10
29	Nativo-americano	B	10
30	Nativo-americano	B	10
31	Nativo-americano	B	10
32	Europa	B	10
33	Europa	B	10
34	Europa	B	10
35	Europa	B	10
36	Europa	B	10
37	Europa	B	10
38	Sudoeste Asiático	B	10
39	Ásia Oriental	B	10
40	Ásia Oriental	B	10
41	Ásia Oriental	B	10
42	Ásia Oriental	B	10
43	Ásia Oriental	B	10
44	Ásia Oriental	B	10
45	Ásia Oriental	B	10
46	Ilhas do Pacífico	B	T	11
47	Ásia Oriental	B	T	12

FIGURA 18.14 Variação de nucleotídios de 5.102 pb do gene *G6PD* em relação a uma amostra mundial de 47 homens. Apenas os 18 sítios variáveis estão demonstrados. A classe alélica funcional (A^- , A^+ , ou B) está

demonstrada em relação a cada sequência. O *SNP2* é um SNP não sinônimo que causa uma substituição de valina para metionina, que é causa das diferenças na atividade enzimática associada ao alelo A^- . O *SNP3* é um SNP não sinônimo que causa substituição do aminoácido ácido aspártico por asparagina. (Dados de M. A. Saunders *et al.*, *Genetics* 162, 2002, 1849-1861.)

A [Figura 18.15](#) demonstra o nível de diversidade nucleotídica em diversos organismos. Os eucariotos unicelulares são os mais diversos, seguidos pelas plantas e em seguida pelos invertebrados. Os vertebrados são o grupo menos diverso; entretanto, a maior parte dos vertebrados ainda apresenta muita diversidade nucleotídica. Em relação aos seres humanos, a diversidade nucleotídica é de aproximadamente 0,001, o que significa que dois cromossomos humanos escolhidos aleatoriamente diferirão em aproximadamente 1 pb por mil. Com 3 bilhões de pb em nosso genoma, isso soma um total de aproximadamente 3 milhões de diferenças entre o conjunto de cromossomos herdado da mãe de uma pessoa e o conjunto herdado do pai de uma pessoa para os indivíduos não consanguíneos.

CONCEITO-CHAVE As populações biológicas com frequência são ricas em variação genética. Essa diversidade pode ser quantificada por meio de diferentes estatísticas para comparar os níveis de variação entre as populações e as espécies.

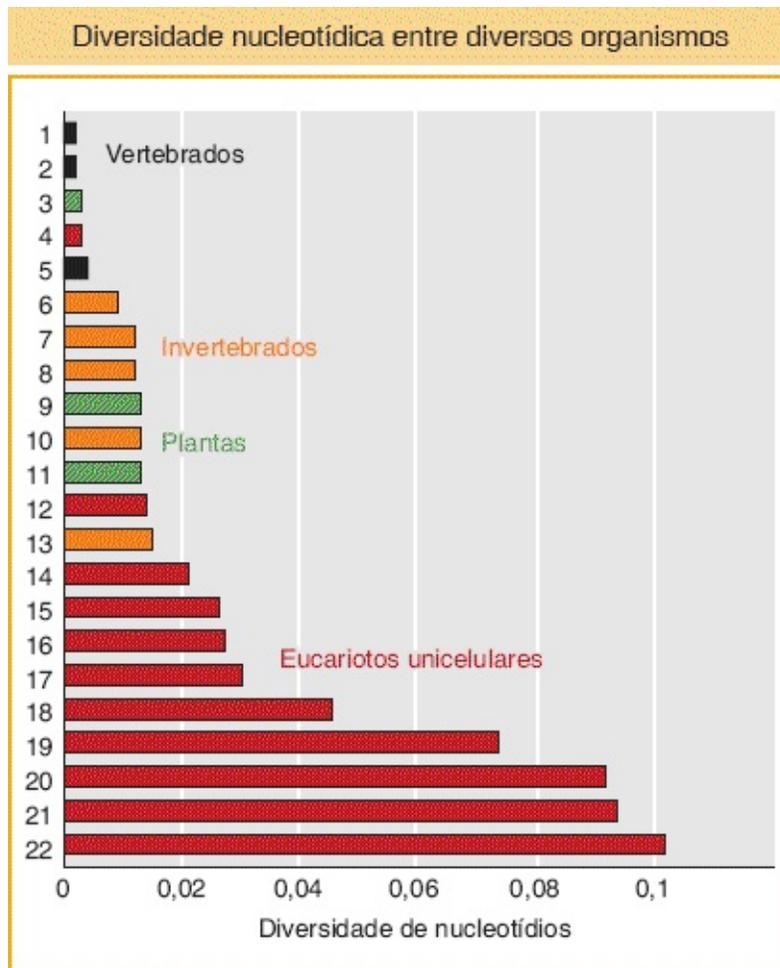


FIGURA 18.15 Níveis de diversidade nucleotídica em sítios sinônimos e silenciosos em alguns organismos diferentes. (1) *Mus musculus*, (2) *Homo sapiens*, (3) *Oryza sativa*, (4) *Plasmodium falciparum*, (5) *Fugu rubripes*, (6) *Strongylocentrotus purpuratus*, (7) *Anopheles gambiae*, (8) *Ciona intestinalis*, (9) *Arabidopsis thaliana*, (10) *Caenorhabditis elegans*, (11) *Zea mays*, (12) *Encephalitozoon cuniculi*, (13) *Drosophila melanogaster*, (14) *Leishmania major*, (15) espécies de *Trypanosoma*, (16) *Toxoplasma gondii*, (17) *Giardia lamblia*, (18) *Neurospora crassa*, (19) *Dictyostelium discoideum*, (20) *Saccharomyces cerevisiae*, (21) *Cryptosporidium parvum*, (22) *Cryptococcus neoformans*. (Dados de M. Lynch e J. S. Conery, *Science* 302, 2003, 1401-1404.)

18.5 Modulação da variação genética

Quais são as forças que modulam a quantidade de variação genética em uma população? Como os novos alelos entram no *pool* gênico? Quais forças removem alelos do *pool* gênico? Como as variantes genéticas podem ser recombinadas para criar novas combinações de alelos? As respostas a essas questões são o

cerne da compreensão do processo da evolução. Nesta seção, examinaremos os papéis da mutação, da migração, da recombinação, da deriva genética (acaso) e da seleção na formação da composição genética das populações.

Novos alelos na população | Mutação e migração

A mutação é a fonte definitiva de toda a variação genética. No [Capítulo 16](#), discutimos os mecanismos moleculares que são a base das mutações em pequena escala, tais como mutações de ponto, indels e alterações no número de unidades de repetição em microssatélites. Os geneticistas de populações estão particularmente interessados na **taxa de mutação**, que é a probabilidade de que uma cópia de um alelo seja alterada para alguma outra forma alélica em uma geração. A taxa de mutação é tipicamente simbolizada pela letra grega μ . Conforme veremos a seguir, se soubermos a taxa de mutação e o número de diferenças nucleotídicas entre duas sequências, então poderemos estimar há quanto tempo as duas sequências divergiram.

Como os geneticistas conseguem estimar a taxa de mutação? Os geneticistas podem estimar as taxas de mutação iniciando com um único indivíduo homocigoto e seguindo o heredograma de seus descendentes durante diversas gerações. Então eles podem comparar a sequência de DNA do indivíduo fundador com as sequências de DNA dos descendentes diversas gerações depois e registrar quaisquer novas mutações que tenham ocorrido. O número de mutações observadas por genoma por geração fornece uma estimativa da taxa. Tendo em vista que estamos procurando eventos um tanto quanto raros, é necessário sequenciar bilhões de nucleotídeos para encontrar apenas algumas mutações de SNP. Em 2009, a taxa de mutação de SNP em relação a uma parte do cromossomo Y humano foi estimada por meio dessa abordagem como sendo de $3,0 \times 10^{-8}$ mutações/nucleotídeo/geração, ou aproximadamente uma mutação a cada 30 milhões de pb. Se extrapolarmos para o genoma humano inteiro (3 bilhões de pb), cada um de nós herdou então 100 novas mutações de cada um de nossos genitores. Felizmente, a maioria das mutações não é prejudicial, tendo em vista que elas ocorrem em regiões do genoma que não são críticas.

A [Tabela 18.5](#) lista as taxas de mutação em relação a SNP e microssatélites em

diversos organismos-modelo. A taxa de mutação de SNP é diversas ordens de magnitude mais baixa do que a taxa de microssatélites. A sua mais alta taxa de mutação e a maior variação tornam os microssatélites particularmente úteis na genética de populações e na ciência forense do DNA. A taxa de mutação de SNP por geração aparenta ser mais baixa em relação a organismos unicelulares do que em relação a organismos multicelulares. Essa diferença pode ser explicada no mínimo parcialmente pelo número de divisões celulares por geração. Existem aproximadamente 200 divisões celulares desde o zigoto até o gameta em seres humanos, mas apenas uma em *E. coli*. Se a taxa humana for dividida por 200, então a taxa por divisão celular em seres humanos está notavelmente próxima da taxa em *E. coli*.

Tabela 18.5 Taxas de mutação aproximadas por geração por genoma haploide.

Organismo	Mutações de SNP (por pb)	Microssatélite
<i>Arabidopsis</i>	7×10^{-9}	9×10^{-4}
Milho	3×10^{-8}	8×10^{-4}
<i>E. coli</i>	5×10^{-10}	—
Levedura	5×10^{-10}	4×10^{-5}
<i>C. elegans</i>	3×10^{-9}	4×10^{-3}
<i>Drosophila</i>	4×10^{-9}	9×10^{-6}
Camundongo	4×10^{-9}	3×10^{-4}

Ser humano

3×10^{-8}

6×10^{-4}

Nota: a taxa de microssatélites é em relação a repetições de dinucleotídeos ou trinucleotídeos.

Fonte: dados de diversos estudos publicados.

Além da mutação, o único outro meio para a entrada de nova variação em uma população é por meio de **migração** ou **fluxo gênico**, a movimentação de indivíduos (ou gametas) entre as populações. A maior parte das espécies é dividida em um conjunto de pequenas populações ou subpopulações locais. As barreiras físicas, tais como oceanos, rios ou montanhas, podem reduzir o fluxo gênico entre as subpopulações, mas com frequência ocorre algum grau de fluxo gênico, apesar de tais barreiras. Dentro das subpopulações, um indivíduo pode ter uma chance de acasalar com qualquer outro membro do sexo oposto; entretanto, os indivíduos de diferentes subpopulações não podem acasalar, exceto se houver migração.

As subpopulações isoladas tendem a divergir na medida em que cada uma acumula as suas próprias mutações únicas. O fluxo gênico limita a divergência genética entre as subpopulações. Uma das consequências genéticas da migração é a **mistura genética**, a mistura de genes que resulta quando os indivíduos apresentam antepassados de mais de uma subpopulação. Esse fenômeno é comum em populações humanas. Ele é prontamente observado na África do Sul, para onde migrantes de todo o mundo foram levados. Conforme demonstrado na [Figura 18.16](#), os genomas dos sul-africanos de ancestralidade mista são complexos e incluem partes da população indígena da África Austral, mais contribuições de migrantes da África Ocidental, da Europa, da Índia, da Ásia Oriental e de outras regiões.

CONCEITO-CHAVE A mutação é a fonte definitiva de toda a variação genética. A migração pode adicionar variação genética a uma população por meio do fluxo gênico de outra população da mesma espécie.

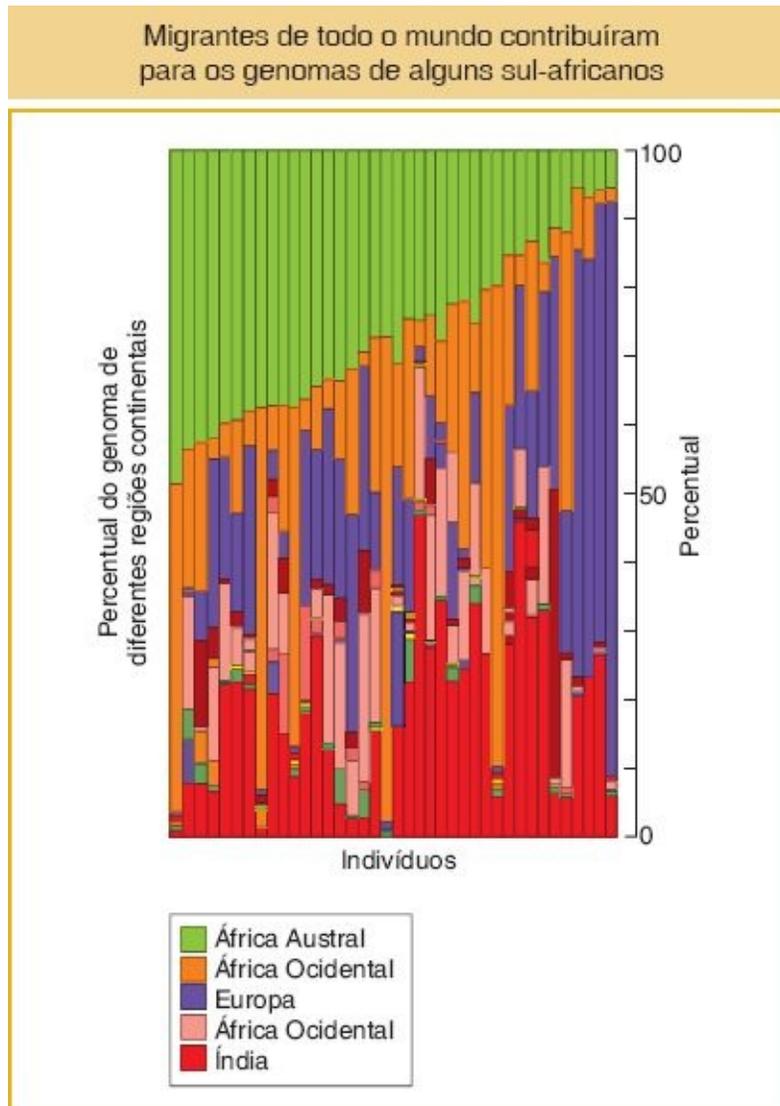


FIGURA 18.16 Representação gráfica da mistura genética em 39 pessoas de ancestralidade mista da África do Sul. Cada coluna representa o genoma de uma pessoa, e as cores representam as partes de seu genoma com contribuição por parte de seus ancestrais, que vieram de muitas regiões do mundo. A figura tem por base a análise genética de populações de mais de 800 microsatélites e 500 indels que foram classificados em quase 4.000 pessoas de todo o mundo, incluindo as 39 de ancestralidade mista da África do Sul. (Dados de S. A. Tishkoff *et al.*, *Science* 324, 2009, 1035-1044.)

Recombinação e desequilíbrio de ligação

A recombinação é uma força crítica que molda os padrões da variação genética nas populações. Nesse caso, não existe ganho ou perda de alelos; em vez disso, a

recombinação cria novos haplótipos. Vejamos como isso funciona. Considere os *loci* *A* e *B* ligados. Pode haver uma população na qual apenas dois haplótipos são observados na geração t_0 : *AB* e *ab*. Suponha que um indivíduo nessa população é heterozigoto em relação a esses dois haplótipos:

$$\frac{A \quad \quad \quad B}{a \quad \quad \quad b}$$

Se ocorrer um *crossing over* nesse indivíduo, então gametas com dois novos haplótipos, *Ab* e *aB*, podem ser formados e entrar na população na geração t_1 .

$$\frac{A \quad \quad \quad B}{a \quad \quad \quad b}$$

Portanto, a recombinação pode criar a variação que adota a forma de novos haplótipos. Os novos haplótipos podem apresentar propriedades únicas que alterem a função proteica. Por exemplo, suponha que um aminoácido variante em uma proteína em um haplótipo aumente a atividade enzimática da proteína em duas vezes e que um segundo aminoácido variante em outro haplótipo também aumente a atividade em duas vezes. Um evento de recombinação que combine essas duas variantes produziria uma proteína com uma atividade quatro vezes mais alta.

Agora consideraremos as frequências observadas e esperadas dos quatro possíveis haplótipos em relação a dois *loci*, cada um com dois alelos. Os *loci* ligados, *A* e *B*, apresentam os alelos *A* e *a* e *B* e *b*, com as frequências p_A , p_a , p_B , e p_b , respectivamente. Os quatro haplótipos possíveis são *AB*, *Ab*, *aB*, e *ab*, com as frequências observadas P_{AB} , P_{Ab} , P_{aB} e P_{ab} . A qual frequência esperamos observar cada um desses quatro haplótipos? Se houver uma relação aleatória entre os alelos nos dois *loci*, então a frequência de qualquer haplótipo será o produto das frequências dos dois alelos que compõem aquele haplótipo:

$$P_{AB} = p_A \times p_B$$

Por exemplo, suponha que a frequência de cada um dos alelos é de 0,5; ou seja, $p_A = p_a = p_B = p_b = 0,5$. Quando obtemos uma amostra do *pool* gênico, a probabilidade de coleta de um cromossomo com um alelo *A* é de 0,5. Se a relação entre os alelos no *locus A* e os alelos no *locus B* for aleatória, então a probabilidade de que o cromossomo selecionado apresente o alelo *B* também é de 0,5. Portanto, a probabilidade de coletarmos um cromossomo com o haplótipo *AB* é:

$$P_{AB} = p_A \times p_B = 0,5 \times 0,5 = 0,25$$

Se a associação entre os alelos em dois *loci* for aleatória, conforme descrito há pouco, diz-se então que os dois *loci* estão em **equilíbrio de ligação**. Nesse caso, as frequências observadas e esperadas serão as mesmas. A [Figura 18.17 A](#) diagrama um caso de dois *loci* em equilíbrio de ligação.

Se a associação entre os alelos em dois *loci* for não aleatória, diz-se então que os *loci* estão em **desequilíbrio de ligação (LD)**. Nesse caso, um alelo específico no primeiro *locus* está associado a um alelo específico no segundo *locus* mais frequentemente do que o esperado ao acaso. A [Figura 18.17 B](#) diagrama um caso de LD completo entre dois *loci*. O alelo *A* sempre está associado ao alelo *B*, enquanto o alelo *a* sempre está associado ao alelo *b*. Não existem cromossomos com haplótipos *Ab* ou *aB*. Nesse caso, as frequências observadas e esperadas não serão as mesmas.

Podemos quantificar o nível de LD entre dois *loci* como a diferença (*D*) entre a frequência observada de um haplótipo e a frequência esperada em relação a uma associação aleatória entre alelos nos dois *loci*. Se ambos os *loci* envolvidos apresentarem apenas dois alelos, então:

$$D = P_{AB} - p_A p_B$$

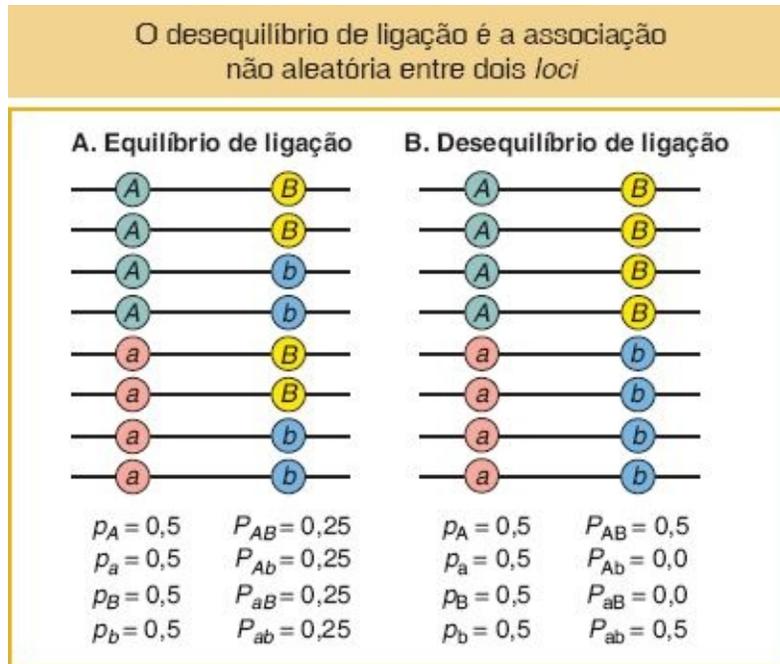


FIGURA 18.17 A. Equilíbrio de ligação. B. Desequilíbrio de ligação em relação a dois *loci* (A e B).

Na [Figura 18.17 A](#), $D = 0$, tendo em vista que não há LD e, na [Figura 18.17 B](#), $D = 0,25$, que é superior a 0, indicando a presença de LD.

Como o LD surge? Sempre que ocorre uma nova mutação em um *locus*, a mutação aparece em um único cromossomo específico e, assim, ela é instantaneamente ligada (ou associada) a alelos específicos em qualquer *loci* vizinho naquele cromossomo. Considere uma população na qual existem apenas dois haplótipos: *AB* e *Ab*. Se uma nova mutação (*a*) surgir no *locus A* em um cromossomo que já possui o alelo *b* no *locus B*, então um novo haplótipo *ab* será formado. Ao longo do tempo, esse novo haplótipo *ab* pode aumentar em frequência na população. Outros cromossomos na população possuiriam os haplótipos *AB* ou *Ab* nesses dois *loci*, mas nenhum cromossomo possuiria *aB*. Portanto, os *loci* estariam em LD. A migração também pode causar LD quando uma subpopulação apresenta apenas o haplótipo *AB* e outra apenas o haplótipo *ab*. Quaisquer migrantes entre as subpopulações dariam origem a LD na subpopulação que recebe os migrantes.

O LD entre dois *loci* declinará ao longo do tempo na medida em que os *crossovers* entre eles tornarem aleatória a relação entre os seus alelos. A taxa de

declínio no LD depende da taxa na qual o *crossing over* ocorre. A frequência de recombinantes (FR) entre os dois *loci* nos gametas que formam a próxima geração (ver [Capítulo 4](#)) fornece uma estimativa da taxa de recombinação, a qual, na genética de populações, é simbolizada pela letra minúscula r . Se D_0 é o valor do desequilíbrio de ligação entre dois *loci* na geração atual, então o valor na próxima geração (D_1) é fornecido por esta equação:

$$D_1 = D_0(1 - r)$$

Em outras palavras, o desequilíbrio de ligação, conforme medido por D , declina a uma taxa de $(1 - r)$ por geração. Quando r é pequena, D declina lentamente ao longo do tempo. Quando r está no seu máximo (0,5), então D declina em 1/2 a cada geração.

Tendo em vista que o LD decai como uma função do tempo e da fração de recombinação, geneticistas de populações podem utilizar o nível de LD entre uma mutação e os *loci* que a circundam para estimar o tempo em gerações desde que a mutação surgiu pela primeira vez na população. Mutações mais antigas apresentam pouco LD com os *loci* vizinhos, enquanto mutações recentes demonstram um alto nível de LD com os *loci* vizinhos. Se você observar novamente a [Figura 18.14](#), você verá que existe um LD considerável entre *SNP2* no *G6PD* e os *SNP* vizinhos. *SNP2* codifica a substituição do aminoácido valina por metionina no alelo A^- que confere resistência à malária. Geneticistas de populações utilizaram o LD na *G6PD* para estimar que o alelo A^- surgiu há aproximadamente 10.000 anos. Não se acreditava que a malária fosse prevalente na África até então. Portanto, o A^- surgiu por meio de uma mutação aleatória, mas foi mantido na população em virtude de proporcionar a proteção contra a malária.

CONCEITO-CHAVE O desequilíbrio de ligação decorre do fato de que novas mutações surgem em um único haplótipo. O desequilíbrio de ligação decairá ao longo do tempo em virtude da recombinação.

Deriva genética e tamanho da população

A lei de Hardy-Weinberg nos informa que as frequências alélicas permanecem as mesmas de uma geração até a próxima em uma *população infinitamente grande*. Entretanto, as reais populações de organismos na natureza são *finitas*, não infinitas. Em populações finitas, as frequências alélicas podem ser alteradas de uma geração para a próxima como resultado do acaso (erro de amostragem) quando os gametas são retirados do *pool* gênico para formar a próxima geração. A alteração nas frequências alélicas entre as gerações em virtude de erro de amostragem é denominada **deriva genética aleatória**, ou apenas deriva, abreviadamente.

Consideraremos um caso simples, porém extremo — uma população composta por um único indivíduo ($N = 1$) heterozigoto (A/a) na geração t_0 . Possibilitaremos a autofertilização. Nesse caso, o *pool* gênico pode ser descrito como apresentando dois alelos, A e a , cada um presente a uma frequência de $p = q = 0,5$. O tamanho da população permanece o mesmo, $N = 1$, na geração subsequente, t_1 . Qual é a probabilidade de que as frequências alélicas sejam alteradas (“derivem”) para $p = 1$ e $q = 0$ na geração t_1 ? Em outras palavras, qual é a probabilidade de que a população se torne fixada em relação ao alelo A , de modo que ela seja composta por um único indivíduo homozigoto A/A ? Tendo em vista que $N = 1$, precisamos coletar apenas dois gametas do *pool* gênico para formar um único indivíduo. A probabilidade de coleta de dois A é $p^2 = 0,5^2 = 0,25$. Portanto, em 25% das ocasiões essa população “derivará” para longe das frequências alélicas originais e se tornará fixada em relação ao alelo A após apenas uma geração.

O que ocorre se aumentarmos o tamanho da população para $N = 2$ e o *pool* gênico inicial ainda apresentar $p = q = 0,5$? As frequências alélicas serão alteradas para $p = 1$ e $q = 0$ na próxima geração apenas se a população for composta por dois indivíduos A/A . Para que isso ocorra, precisamos coletar quatro alelos A , cada um com uma probabilidade de $p = 0,5$; assim, a probabilidade de que próxima geração apresente $p = 1$ e $q = 0,0$ é $p^4 = (0,5)^4 = 0,0625$, ou apenas pouco mais de 6%. Portanto, uma população com $N = 2$

apresenta menos propensão à deriva para a fixação do alelo A do que uma população com $N = 1$. Em termos mais gerais, a probabilidade de que uma população sofra deriva para a fixação do alelo A em uma única geração é p^{2N} e, portanto, essa probabilidade se torna progressivamente menor na medida em que o tamanho da população (N) se torna maior. A deriva é uma força mais fraca em populações grandes.

Deriva significa qualquer alteração nas frequências alélicas em virtude de erro de amostragem, não apenas a perda ou a fixação de um alelo. Em uma população de $N = 500$ com dois alelos a uma frequência de $p = q = 0,5$, existem 500 cópias de A e 500 cópias de a . Se a próxima geração apresentar 501 cópias de A ($p = 0,501$) e 499 cópias de a ($q = 0,499$), terá ocorrido deriva genética, embora seja um nível de deriva muito modesto. Uma fórmula geral para o cálculo da probabilidade de observação de um número específico de cópias de um alelo na próxima geração, tendo em vista as frequências na geração atual, é apresentada no [Quadro 18.4](#).

Quando a deriva está operando em uma população finita, podemos calcular a probabilidade de diferentes desfechos, mas não podemos prever com precisão o desfecho específico que ocorrerá. O processo é como o de jogar dados. Em qualquer *locus*, a deriva pode continuar de uma geração até a próxima, até que um alelo tenha sido fixado. Além disso, em uma população em particular, a frequência do alelo A pode aumentar da geração t_0 para a t_1 , mas em seguida diminuir da geração t_1 até a t_2 . A deriva não prossegue em um sentido específico em direção à perda ou à fixação de um alelo.

As [Figuras 18.18 A](#) e [18.18 B](#) demonstram estudos aleatórios simulados por computador (jogadas de dados) em relação a seis populações de tamanho $N = 10$ e $N = 500$. Cada população inicia com a apresentação de dois alelos a uma frequência de $p = q = 0,5$, e em seguida os estudos aleatórios prosseguem durante 30 gerações. Primeiramente, observe a aleatoriedade do processo de uma geração até a próxima. Por exemplo, a frequência de A na população ilustrada pela linha amarela na [Figura 18.18 A](#) oscila para cima e para baixo de uma geração até a próxima, alcançando uma baixa de $p = 0,16$ na t_{16} , mas que em seguida retorna para $p = 0,75$ na t_{30} . Em segundo lugar, seja $N = 10$ ou $N = 500$, observe que duas

populações não apresentam exatamente a mesma trajetória. A deriva é um processo aleatório, e provavelmente não observaremos exatamente o mesmo desfecho com populações diferentes ao longo de muitas gerações, exceto quando N é muito pequeno. Em terceiro lugar, observe que quando $N = 10$, as populações se tornaram fixadas (seja $p = 1$ ou $p = 0$) antes da geração 20 em cinco dos seis estudos. Entretanto, quando $N = 500$, as populações retiveram ambos os alelos em todos os seis estudos, até mesmo depois de 30 gerações.

Quadro 18.4

A frequência alélica é alterada sob a deriva.

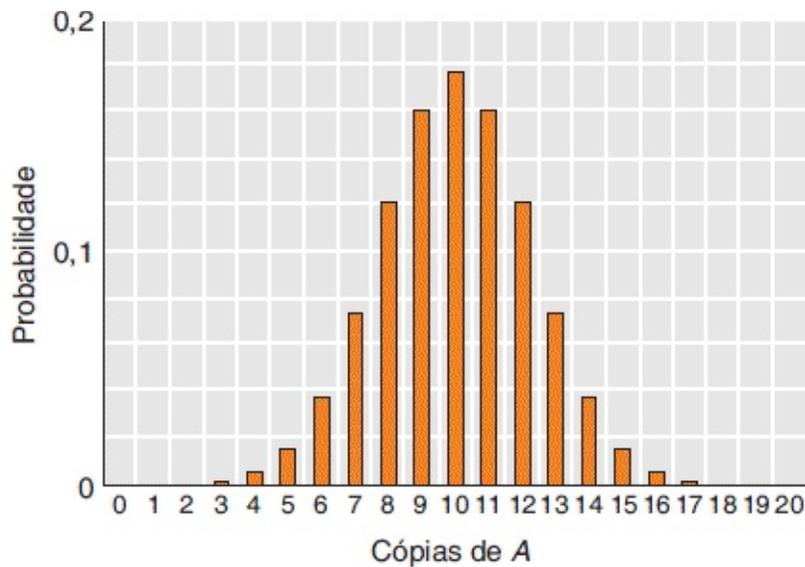
Considere uma população de N indivíduos diploides segregando para dois alelos A e a no *locus* A , com frequências p e q , respectivamente. A população é de cruzamento aleatório e o tamanho da população permanece o mesmo (N) em cada geração. Quando se obtém uma amostra do *pool* gênico para criar a próxima geração, o número exato de cópias do alelo A que são retiradas não pode ser previsto estritamente em virtude de erro de amostragem. Entretanto, a probabilidade de que um número específico de cópias de A seja retirado pode ser calculada com a utilização da fórmula binomial. Deixe que k seja o número específico de cópias do alelo A . A probabilidade de coleta de k cópias é:

$$\text{Prob}(k) = \left(\frac{2N!}{k! (2N - k)!} \right) p^k q^{(2N - k)}$$

Se estabelecermos $N = 10$ e $p = q = 0,5$, então a probabilidade de coleta de 10 cópias do alelo A é:

$$\text{Prob}(10) = \left(\frac{20!}{10! (20 - 10)!} \right) 0,5^{10} 0,5^{(20 - 10)} = 0,176$$

Portanto, em apenas 17,6% das ocasiões a próxima geração apresentará a mesma frequência de A e a da geração original. Podemos utilizar essa fórmula para calcular os desfechos em relação a todos os valores possíveis de k e obter uma distribuição de probabilidade, demonstrada na figura a seguir.



Distribuição de probabilidade demonstrando a probabilidade de que números diferentes de A estarão presentes após uma geração.

O desfecho único mais provável é a ausência de deriva, com $k = 10$ e uma probabilidade de 0,176. Entretanto, todos os outros desfechos envolvem alguma deriva e, assim, a probabilidade de que a população apresente alguma deriva é de 0,824.

Além do tamanho da população, o destino de um alelo é determinado por sua frequência na população. Especificamente, a probabilidade de que um alelo sofra deriva para fixação em uma geração futura é igual à sua frequência na geração atual. Um alelo que está a uma frequência de 0,5 apresenta uma chance de 50:50 de fixação ou perda da população em uma geração futura. Você pode observar o efeito da frequência alélica sobre o destino de um alelo na [Figura 18.18 C](#). Em relação a dez populações com uma frequência inicial de $p = 0,1$, oito populações apresentaram a perda do alelo A, uma a sua fixação, e uma população reteve ambos os alelos após 30 gerações. Isso está muito próximo da expectativa de que A chegue à fixação em 10% das ocasiões quando $p = 0,1$.

O fato de a frequência de um alelo ser igual à sua probabilidade de fixação significa que a maior parte das mutações surgidas recentemente será finalmente

perdida em uma população em virtude da deriva. A frequência inicial de uma nova mutação no *pool* gênico é:

$$\frac{1}{2N}$$

Se N for até mesmo modestamente grande, tal como 10.000, a probabilidade de que uma nova mutação finalmente alcance a fixação é, então, extremamente pequena, $1/2N = 1/20.000 = 5 \times 10^{-5}$. A probabilidade de que uma nova mutação finalmente seja perdida da população é:

$$\frac{2N-1}{2N} = 1 - \frac{1}{2N}$$

que está próxima de 1,0 em grandes populações. Ela é de 0,99995 em uma população de 10.000.

A [Figura 18.19 A](#) demonstra uma representação gráfica do destino de novas mutações em uma população. O eixo x representa o tempo e o eixo y o número de cópias de um alelo. As linhas pretas demonstram o destino da maior parte das novas mutações. Elas aparecem e logo em seguida são perdidas da população. As linhas coloridas demonstram as poucas novas mutações “de sorte” que se fixaram. A partir da teoria da genética de populações, pode-se demonstrar que o tempo médio necessário para que uma mutação tenha a sorte de ser fixada é de $4N$ gerações. A [Figura 18.19 B](#) demonstra uma população que tem $1/2$ do tamanho da população na [Figura 18.19 A](#). Portanto, $4N$ gerações correspondem a $1/2$ do tempo e as novas mutações de sorte são fixadas mais rapidamente.

Uma consequência importante da deriva é que alelos levemente deletérios podem ser fixados ou alelos vantajosos podem ser perdidos por meio desse processo aleatório. Considere um novo alelo que surge em uma população e que proporciona ao indivíduo que o carrega ter um sistema imune mais forte. Esse indivíduo pode transmitir o alelo vantajoso para a sua descendência, mas a descendência pode morrer antes da reprodução em virtude de um evento aleatório, tal como o fato de ser atingida por raios. Ou, se o indivíduo que carrega o alelo favorável for heterozigoto, ele pode transmitir apenas o alelo menos

favorável para a sua descendência ao acaso.

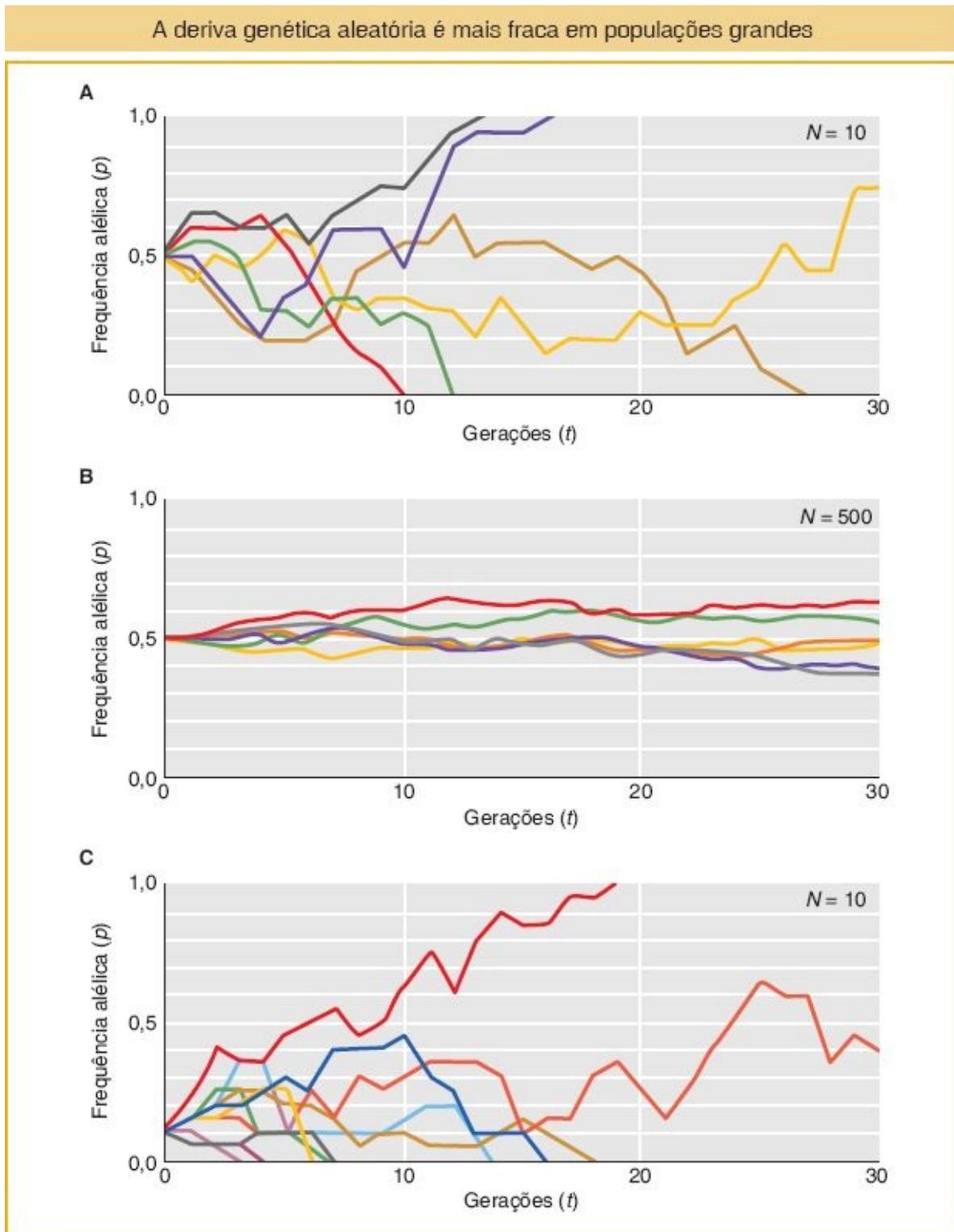


FIGURA 18.18 Simulações em computador de deriva genética aleatória. Cada *linha colorida* representa

uma população simulada ao longo de 30 gerações. **A.** $N = 10, p = q = 0,5$. **B.** $N = 500, p = q = 0,5$. **C.** $N = 10, p = q = 0,9$.

Ao calcular as probabilidades de diferentes desfechos sob deriva genética, estamos presumindo que os alelos A e a não conferem diferenças na viabilidade ou no sucesso reprodutivo aos indivíduos que os carregam. Presumimos que indivíduos AA , Aa e aa apresentam probabilidade igual de sobreviver e se reproduzir. Nesse caso, A e a seriam denominados **alelos neutros** (ou variantes) um em relação ao outro. A alteração nas frequências de alelos neutros ao longo do tempo em virtude da deriva é denominada **evolução neutra**. O processo de evolução neutra é o fundamento do **relógio molecular**, a taxa constante de substituição de variantes alélicas de surgimento recente em relação às preexistentes durante longos períodos ([Quadro 18.5](#)). A evolução neutra é distinta da evolução darwiniana, na qual os alelos favoráveis aumentam em frequência em virtude de os indivíduos que os carregam deixarem mais descendentes. Discutiremos sobre a evolução darwiniana na próxima seção deste capítulo e no [Capítulo 20](#).

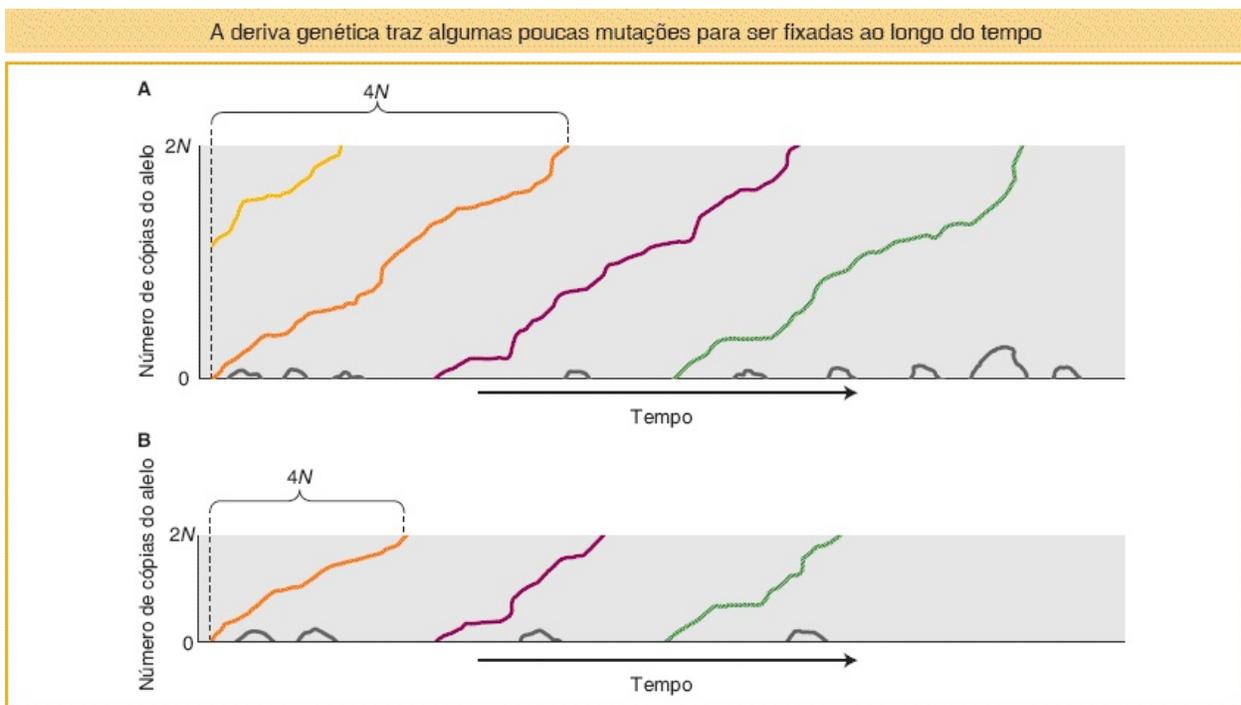


FIGURA 18.19 **A.** Representação gráfica do aparecimento, da perda e da incorporação final de novas

mutações em uma população ao longo do tempo sob a ação da deriva genética. As linhas de *cor cinza* demonstram o destino da maior parte das novas mutações, as quais aparecem e em seguida são perdidas da população em algumas poucas gerações. As *linhas coloridas* demonstram o destino das poucas mutações “de sorte” que continuam a aumentar em frequência, até alcançarem a fixação. **B.** Uma população que tem 1/2 do tamanho da população na parte **A.** Nessa população, $4N$ gerações correspondem a 1/2 do tempo e as novas mutações de sorte são fixadas mais rapidamente.

Até agora, temos considerado a deriva no contexto de populações que permanecem do mesmo tamanho de uma geração para a próxima. Na realidade, as populações com frequência contraem ou expandem em tamanho ao longo do tempo. Por exemplo, uma nova população de tamanho muito menor pode ser formada subitamente quando um número relativamente pequeno de membros de uma população migra para um novo local e estabelece uma nova população. Os migrantes, ou “fundadores” da nova população podem não carregar todos os alelos presentes na população original, ou podem carregar os mesmos alelos, porém em frequências diferentes. A deriva genética causada pela amostragem aleatória da população original para criar a nova população é conhecida como **efeito fundador**. Um dos muitos eventos fundadores na história humana ocorreu quando pessoas cruzaram a ponte terrestre de Bering da Ásia até as Américas durante a era do gelo, há aproximadamente 15.000 a 30.000 anos. Como resultado, a diversidade genética entre nativos americanos é mais baixa do que entre pessoas em outras regiões do mundo ([Figura 18.20](#)).

O tamanho da população também pode ser alterado em um único local. Um período de uma ou diversas gerações consecutivas de contração no tamanho da população é conhecido como um **gargalo** populacional. Os gargalos ocorrem em populações naturais em virtude de flutuações ambientais, tais como uma redução no suprimento alimentar ou um aumento na predação. O lobo-cinzento, o bisão-americano, a águia-de-cabeça-branca, o condor-da-califórnia, o grou-americano e muitas espécies de baleias são alguns exemplos familiares de espécies que sofreram gargalos recentes em virtude da caça pelos seres humanos, ou da invasão do seu *habitat* por parte dos seres humanos. A redução no tamanho da população durante um gargalo aumenta o nível de deriva em uma população. Conforme explicado anteriormente no capítulo, o nível de endocruzamento em

populações também depende do tamanho da população. Portanto, os gargalos também causam um aumento no nível de endocruzamento.

O condor-da-califórnia representa um exemplo extraordinário de um gargalo. Essa espécie chegou a habitar uma ampla área, mas, na década de 1980, ela declinou até uma população reprodutiva de apenas 14 aves em cativeiro. A população atualmente é superior a 400 indivíduos, mas a heterozigosidade média no genoma diminuiu em 8% durante o gargalo inicial. Além disso, um alelo recessivo deletério em relação a um tipo letal de nanismo ocorre a uma frequência de aproximadamente 9% entre os animais sobreviventes, presumivelmente como um resultado da deriva a partir de uma frequência mais baixa na população pré-gargalo. Para tratar desses problemas, biólogos de conservação estabelecem cruzamentos de animais em cativeiro para minimizar o endocruzamento adicional e remover os alelos deletérios da população.

Quadro 18.5

Relógio molecular

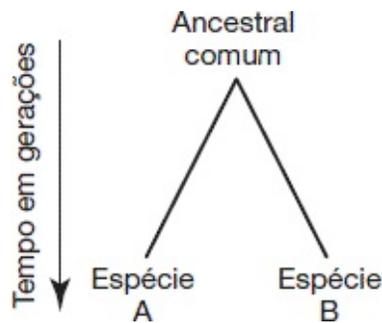
Visto que as espécies divergem ao longo do tempo, suas sequências de DNA se tornam cada vez mais diferentes na medida em que mutações surgem e são fixadas na população. A qual taxa as sequências divergem? Para responder a essa questão, considere uma população na geração t_0 . O número de mutações que aparecerá na geração t_1 é o produto do número de cópias da sequência no *pool* gênico ($2N$) vezes a taxa na qual elas sofrem mutação (μ); ou seja, $2N\mu$. Se uma mutação for neutra, então a probabilidade de que ela derive até a fixação é de $1/2N$. Assim, a cada geração, $2N\mu$ novas mutações entram no *pool* gênico, e $1/2N$ delas será fixada. O produto desses dois números é a taxa (k) na qual as sequências evoluem:

$$k = 2N\mu \times \frac{1}{2N} = \mu$$

O valor k é denominado taxa de substituição, e é igual à taxa de mutação em relação às mutações neutras. Se a taxa de mutação permanecer constante ao longo do tempo, a taxa de substituição então “baterá” regularmente como um

relógio, o relógio molecular.

Considere duas espécies, A e B, e seu ancestral comum. Definiremos d (divergência) como o número de substituições neutras em sítios de nucleotídios na sequência de DNA de um gene que ocorreram desde a divergência de A e B de seu ancestral.



O valor esperado em relação a d será o produto da taxa (k) na qual as substituições ocorrem e duas vezes o tempo em gerações ($2t$) durante as quais a substituição foi acumulada. O 2 é necessário em virtude de existirem duas linhagens que causam o distanciamento do ancestral comum. Portanto, temos:

$$d = 2tk$$

Essa equação pode se reescrita como:

$$t = \frac{d}{2k}$$

demonstrando como podemos calcular o tempo em gerações desde a divergência de duas espécies quando conhecemos d e k . A taxa de mutação de SNP por geração (μ) é conhecida em relação a muitos grupos de organismos (ver [Tabela 18.5](#)), e é a mesma que a taxa de substituição (k) em relação às mutações neutras. Pode-se sequenciar um ou mais genes de duas espécies e determinar a proporção de sítios de nucleotídios silenciosos (neutros) nos quais eles diferem e utilizar essa proporção como uma estimativa em relação a d . Portanto, pode-se calcular o tempo desde que duas sequências (duas espécies) divergiram com a utilização do relógio molecular. Entre seres

humanos e chimpanzés, existe aproximadamente 0,018 diferença de bases em sítios sinônimos nas sequências codificadoras. A taxa de mutação de SNP em relação aos seres humanos é de 3×10^{-8} , e o tempo de geração é de aproximadamente 20 anos. Com a utilização desses valores e da equação anterior, o tempo de divergência estimado em relação aos seres humanos e aos chimpanzés é de aproximadamente 6,0 milhões de anos. Esses cálculos presumem que as substituições são neutras e que a taxa de substituição tem sido constante ao longo do tempo.

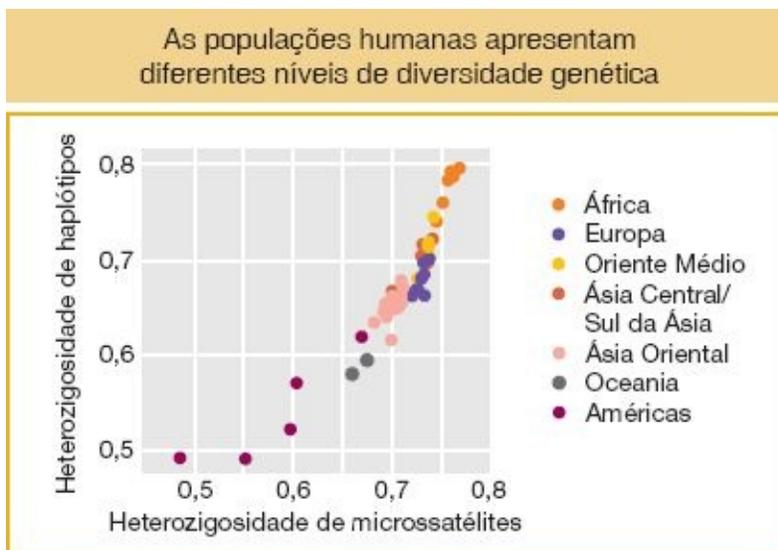
O [Quadro 18.6](#) discute sobre o gargalo bem-caracterizado que ocorreu durante a domesticação de espécies de cultivo. Esse gargalo explica o motivo de nossas plantas de cultivo apresentarem muito menos diversidade genética do que seus ancestrais selvagens.

CONCEITO-CHAVE O tamanho da população é um fator-chave que afeta a variação genética nas populações. A deriva genética exerce maior força sobre populações pequenas do que grandes. A probabilidade de que um alelo se torne fixado (ou seja perdido) em uma população por meio de deriva é uma função de sua frequência na população e do tamanho da população. A maior parte das novas mutações neutras é perdida das populações em decorrência de deriva.

Seleção

Até agora, consideramos como os novos alelos entram em uma população por meio de mutação e migração e como esses alelos podem se fixar (ou se perder) em uma população por meio da deriva aleatória. Mas a mutação, a migração e a deriva não conseguem explicar o motivo de os organismos aparentarem estar tão bem-adaptados aos seus ambientes. Elas não conseguem explicar as **adaptações**, as características da forma ou da fisiologia de um organismo que possibilitam que ele lide melhor com as condições ambientais sob as quais ele vive. Para explicar

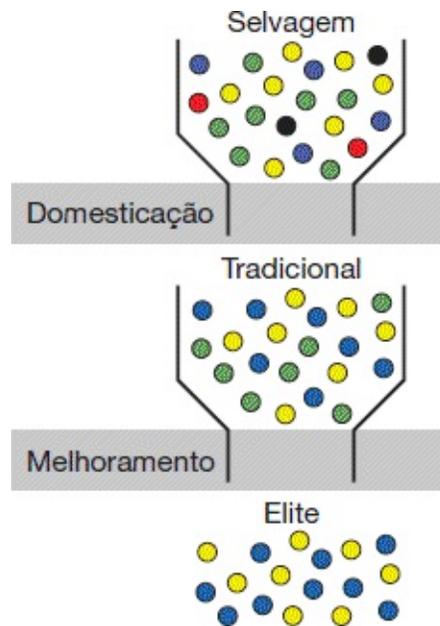
a origem das adaptações, em 1859, Charles Darwin, em seu histórico livro *A Origem das Espécies*, propôs que as adaptações surgem por meio da ação de outro processo, que ele denominou “seleção natural”. Nesta seção, exploraremos o papel da seleção natural na modulação da variação genética nas populações. Posteriormente, no [Capítulo 20](#), consideraremos os efeitos da seleção natural sobre a evolução de genes e de traços ao longo de períodos prolongados.



a domesticação, eles coletaram uma amostra do *pool* gênico selvagem. Essa amostra apresentava apenas um subconjunto da variação genética observada na vida selvagem. As populações domesticadas passaram por um gargalo. Conseqüentemente, as plantas de cultivo e os animais domesticados tipicamente apresentam menos variação genética do que os seus genitores selvagens.

O cultivo científico moderno de plantas que objetivava a melhora das plantações criou um segundo gargalo. Ao amostrar o *pool* gênico das variedades de cultivo tradicionais, agricultores modernos criaram variedades de elite, com traços de valor comercial, tais como alta produção e adequabilidade à coleta e ao processamento mecânicos. Conseqüentemente, as variedades de elite ou modernas apresentam ainda menos variação genética do que as variedades tradicionais.

A perda da variação genética resultante dos gargalos da domesticação e do melhoramento pode impor uma ameaça. Tendo em vista que existem menos alelos por *locus*, os cultivos apresentam um repertório menor de alelos nos genes de resistência a doenças e possivelmente maior suscetibilidade a patógenos emergentes. Para reduzir essa vulnerabilidade, cultivadores realizam cruzamentos entre as variedades modernas e os parentes selvagens (ou as variedades tradicionais) para reintroduzir alelos criticamente importantes nos cultivos modernos.



Gargalos da domesticação e do melhoramento de cultivos. Os *pontos coloridos* representam alelos diferentes. (M. Yamasaki *et al.*, *Plant Cell* 17, 2005, 2859-2872.)

Definiremos a **seleção natural** como o processo por meio do qual os indivíduos com determinadas características hereditárias apresentam maior probabilidade de sobreviver e se reproduzir do que outros indivíduos com ausência dessas características. Conforme resumido por Darwin, o processo funciona assim. A cada geração, são produzidos mais descendentes do que os que conseguem sobreviver e se reproduzir no ambiente. A natureza apresenta um mecanismo (mutação) para produzir novas formas ou variantes hereditárias. Indivíduos com variantes particulares de algumas características apresentam maior probabilidade de sobreviver e se reproduzir. Indivíduos com características que intensificam a sua capacidade de sobreviver e se reproduzir transmitirão essas características à sua descendência. Ao longo do tempo, essas características aumentarão em frequência na população. Portanto, as populações serão alteradas ao longo do tempo (*evoluirão*) na medida em que o ambiente (*a natureza*) favorecer (*selecionar*) as características que intensificam a capacidade

de sobreviver e se reproduzir. Essa é a teoria da evolução de Darwin por meio da seleção natural.

A evolução darwiniana com frequência é descrita com a utilização da frase “*sobrevivência do mais apto*”. Essa frase pode ser enganosa. Um indivíduo que é fisicamente forte, resistente a doenças e vive uma vida longa, mas que não apresenta descendência, não é adaptado no sentido darwiniano. A **adaptabilidade darwiniana** refere-se à capacidade de sobreviver e de *se reproduzir*. Ela considera tanto a viabilidade quanto a fecundidade. Uma medida da adaptabilidade darwiniana é simplesmente o tamanho da descendência que um indivíduo apresenta. Essa medida é denominada **adaptabilidade absoluta**, e a simbolizaremos com um W maiúsculo. Em relação a um indivíduo sem descendência, W é igual a 0, em relação a um indivíduo com um descendente, W é igual a 1, em relação a um indivíduo com dois descendentes, W é igual a 2, e assim por diante. W também é o número de alelos em um *locus* com que um indivíduo contribui para o *pool* gênico.

A adaptabilidade absoluta mistura o tamanho da população e diferenças no sucesso reprodutivo entre os indivíduos. Geneticistas de populações estão primariamente interessados no último e, assim, utilizam uma medida denominada **adaptabilidade relativa** (simbolizada por um w minúsculo), que é a adaptabilidade de um indivíduo em relação àquela de algum outro indivíduo, normalmente o indivíduo mais adaptado na população. Se o indivíduo X apresenta dois descendentes e o indivíduo mais adaptado, Y, apresenta 10 descendentes, a adaptabilidade relativa de X é $w = 2/10 = 0,2$. A adaptabilidade relativa de Y é $w = 10/10 = 1$. Para cada 10 alelos com os quais Y contribui para a próxima geração, X contribuirá com 2.

O conceito de adaptabilidade se aplica aos genótipos, bem como aos indivíduos. A adaptabilidade absoluta em relação ao genótipo A/A (W_{AA}) é o número médio de descendentes deixados pelos indivíduos com aquele genótipo. Se conhecermos a adaptabilidade absoluta em relação a todos os genótipos em um *locus*, podemos então calcular a adaptabilidade relativa de cada um dos genótipos.

Agora veremos como as frequências alélicas podem ser alteradas ao longo do

tempo quando genótipos diferentes apresentam adaptabilidade diferente; ou seja, quando a seleção natural está atuando. A seguir estão a adaptabilidade e as frequências genotípicas em relação a três genótipos no *locus A* em uma população. Nesse caso, *A* é um alelo dominante favorecido, tendo em vista que a adaptabilidade dos indivíduos *A/A* e *A/a* é a mesma e superior à dos indivíduos *a/a*. Estamos presumindo que essa população segue a lei de Hardy-Weinberg, com $p = 0,1$ e $q = 0,9$.

	<i>A/A</i>	<i>A/a</i>	<i>a/a</i>
Número médio de descendentes (<i>W</i>)	10	10	5
Adaptabilidade relativa (<i>w</i>)	1,0	1,0	0,5
Frequência genotípica	0,01	0,18	0,81

A contribuição relativa de cada genótipo para o *pool* gênico é determinada pelo produto de sua adaptabilidade com sua frequência. Quanto mais adaptado e mais alta a frequência de um genótipo, mais ele contribui.

Genótipo	<i>A/A</i>	<i>A/a</i>	<i>a/a</i>	Soma
Contribuição	$1 \times 0,01$	$1 \times 0,18$	$0,5 \times 0,81$	0,595
relativa	= 0,01	= 0,18	= 0,405	

As contribuições relativas não somam 1, então precisamos redimensioná-las ao

dividir cada uma pela soma das três (0,595) para obter as frequências genotípicas esperadas que contribuem para o *pool* gênico.

Genótipo	A/A	A/a	a/a	Soma
Frequências genotípicas	0,02	0,30	0,68	1,0

Com a utilização dessas frequências genotípicas esperadas e da lei de Hardy-Weinberg, podemos calcular as frequências alélicas na próxima geração:

$$p' = 0,02 + \left(\frac{1}{2} \times 0,3\right) = 0,17$$

e:

$$q' = 0,68 + \left(\frac{1}{2} \times 0,3\right) = 0,83$$

A diferença entre p' e q' ($\Delta p = p' - p$) é $0,17 - 0,1 = 0,07$, então concluímos que o alelo *A* aumentou 7% em uma geração em virtude da seleção natural. O [Quadro 18.7](#) apresenta as equações-padrão para o cálculo das alterações nas frequências alélicas ao longo do tempo em virtude da seleção natural.

Passaremos por esse processo recursivamente, utilizando as frequências alélicas da primeira geração para calcular aquelas na segunda geração, utilizando em seguida aquelas da segunda para calcular as da terceira e assim por diante. Se em seguida inserirmos em um gráfico p pelo tempo medido em número de gerações (t), teremos um quadro do tempo no qual as frequências alélicas são alteradas sob a força da seleção natural. A [Figura 18.21](#) demonstra um referido gráfico em relação a ambos um alelo dominante favorecido e um recessivo favorecido. O alelo dominante no início aumenta rapidamente, mas em seguida alcança um platô e se aproxima apenas lentamente da fixação. Após o alelo dominante favorecido alcançar uma alta frequência, o alelo recessivo não favorecido ocorre principalmente em heterozigotos e raramente em homozigotos

com adaptabilidade reduzida e, assim, a seleção é ineficaz para removê-lo da população. O recessivo favorecido se comporta da maneira oposta — primeiramente ele aumenta lentamente em frequência, tendo em vista que os homozigotos a/a com adaptabilidade intensificada são raros, mas posteriormente prossegue mais rapidamente até a fixação. Tendo em vista que a classe heterozigota apresenta adaptabilidade reduzida, o alelo dominante não favorecido finalmente pode ser removido da população.

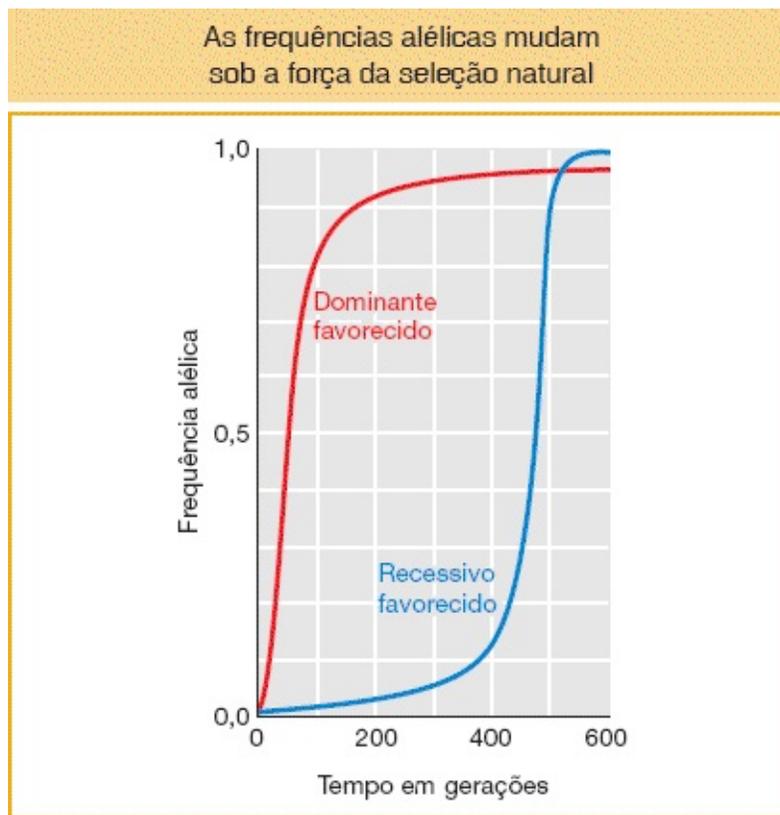


FIGURA 18.21 Alteração na frequência alélica de um alelo dominante favorecido (*vermelho*) e um alelo recessivo favorecido (*azul*), direcionada pela seleção natural ao longo de 600 gerações.

Quadro 18.7

Efeito da seleção sobre as frequências alélicas.

A seleção causa alterações nas frequências alélicas entre gerações, tendo em vista que alguns genótipos contribuem com mais alelos para o *pool* gênico do que outros. Descreveremos um conjunto de equações para prever as frequências gênicas na próxima geração quando a seleção está operando. As

frequências genótípicas e a adaptabilidade absoluta são simbolizadas como segue:

genótipo	A/A	A/a	a/a
frequência	p^2	$2pq$	q^2
adaptabilidade absoluta	$W_{A/A}$	$W_{A/a}$	$W_{a/a}$

O número médio de alelos com que indivíduos de um determinado genótipo contribuem é a frequência do genótipo vezes a adaptabilidade absoluta. Se N é o tamanho da população, o número total de alelos com que todos os indivíduos de determinado genótipo contribuem é N multiplicado pelo número médio de alelos com que indivíduos de determinado genótipo contribuem:

número médio	$p^2 W_{A/A}$	$2pq W_{A/a}$	$q^2 W_{a/a}$
número total	$N(p^2) W_{A/A}$	$N(2pq) W_{A/a}$	$N(q^2) W_{a/a}$

Portanto, o *pool* gênico apresentará:

$$\text{número de alelos } A = N(p^2)W_{A/A} + \frac{1}{2}[N(2pq)W_{A/a}]$$

$$\text{número de alelos } a = N(q^2)W_{a/a} + \frac{1}{2}[N(2pq)W_{A/a}]$$

A adaptabilidade média da população é:

$$\bar{W} = p^2 W_{A/A} + 2pq W_{A/a} + q^2 W_{a/a}$$

que é o número médio de alelos com que um indivíduo contribui para o *pool* gênico. $N\bar{W}$ é o número total de alelos no *pool* gênico.

Agora podemos calcular a proporção de alelos *A* no *pool* gênico na próxima geração como:

$$p' = \frac{Np^2 W_{A/A} + Npq W_{A/a}}{N\bar{W}}$$

Essa equação é reduzida para:

$$p' = p \frac{p W_{A/A} + q W_{A/a}}{\bar{W}}$$

Observe a expressão $p W_{A/A} + q W_{A/a}$. Essa é denominada adaptabilidade alélica ou adaptabilidade média de alelos *A* (W_A):

$$W_A = p W_{A/A} + q W_{A/a}$$

A partir da lei de Hardy-Weinberg, sabemos que uma proporção *p* de todos os alelos *A* está presente em homozigotos com outro *A*, caso em que eles apresentam uma adaptabilidade de $W_{A/A}$, enquanto uma proporção *q* de todos os alelos *A* está presente em heterozigotos com *a* e eles apresentam uma adaptabilidade de $W_{A/a}$. Substituindo W_A na equação anterior, obtemos:

$$p' = p \frac{W_A}{\bar{W}}$$

Essa equação pode ser utilizada para calcular a frequência de *A* na próxima geração e pode ser utilizada recursivamente para seguir a alteração em *p* ao longo do tempo.

Embora tenhamos derivado essas fórmulas com a utilização da adaptabilidade absoluta, em geral não estamos interessados no tamanho da população; assim, utilizamos formas dessas equações com a adaptabilidade relativa:

$$\bar{w} = p^2 w_{A/A} + 2pq w_{A/a} + q^2 w_{a/a}$$

$$w_A = p w_{A/A} + q w_{A/a}$$

$$p' = p \frac{w_A}{\bar{w}}$$

Finalmente, podemos expressar a alteração na frequência alélica entre as gerações como:

$$\begin{aligned} \Delta p &= p' - p = p \frac{w_A}{\bar{w}} - p \\ &= \frac{p(w_A - \bar{w})}{\bar{w}} \end{aligned}$$

Mas \bar{w} , a adaptabilidade relativa média da população, é a média de w_A e w_a , que são as adaptabilidades alélicas de A e a , respectivamente:

$$\bar{w} = p w_A + q w_a$$

Substituindo essa expressão por w na fórmula em relação a Δp e lembrando que $q = 1 - p$, obtemos:

$$p = \frac{pq(w_A - w_a)}{\bar{w}}$$

Tipos de seleção

A seleção natural pode operar de diversos modos. A **seleção direcional**, que temos discutido, movimentada a frequência de um alelo em uma direção até que ele alcance a fixação ou seja perdido. A seleção direcional pode ser *positiva* ou *purificadora*. A **seleção positiva** atua para trazer uma nova mutação ou um novo alelo favorável até uma frequência mais alta. Esse tipo de seleção está em atuação quando novas adaptações evoluem. Uma *varredura seletiva* ocorre quando um alelo favorável alcança a fixação. A seleção direcional também pode atuar para remover mutações deletérias da população. Esse tipo de seleção é denominado **seleção purificadora**, e evita que as características adaptativas

existentes sejam degradadas ou perdidas. A seleção nem sempre prossegue de modo direcional até a perda ou a fixação de um alelo. Se a classe heterozigota apresenta adaptabilidade mais alta do que qualquer uma das classes homozigotas, a seleção natural então favorecerá a manutenção de ambos os alelos na população. Nesse caso, o *locus* está sob **seleção balanceadora**, e a seleção natural irá movimentar a população até um ponto de equilíbrio, no qual ambos os alelos são mantidos na população (ver [Capítulo 20](#)).

Cada um dos diferentes tipos de seleção deixa uma assinatura distinta na sequência de DNA próxima do *locus*-alvo em uma população. Por exemplo, a seleção positiva pode ser detectada em sequências de DNA por meio de seus efeitos sobre a diversidade genética e o equilíbrio de ligação. A [Figura 18.22](#) demonstra haplótipos esquemáticos antes e depois de um episódio de seleção positiva. No painel que demonstra os haplótipos antes da seleção, a região entre chaves apresenta muitos polimorfismos e múltiplos haplótipos. Entretanto, após a seleção, existe apenas um único haplótipo nessa região e, portanto, nenhum polimorfismo. Quando a seleção é aplicada ao sítio-alvo (demonstrado em *vermelho*), todos os sítios-alvo e vizinhos podem ser varridos até a fixação antes que a recombinação fragmente o haplótipo no qual a mutação favorável ocorreu pela primeira vez. O resultado é uma diversidade mais baixa e um LD mais alto próximo do alvo. Na medida em que a distância do alvo aumenta, existem mais oportunidades para a recombinação e, assim, a diversidade retrocede gradualmente.

A [Figura 18.23](#) demonstra o padrão de diversidade na região que circunda o gene *SLC24A5* em seres humanos. Esse gene influencia a deposição da melanina na pele. Quando pessoas migraram da África para a Europa, uma varredura seletiva no *SLC24A5* causou a perda de toda a diversidade nesse *locus*. Consequentemente, existe um alelo único e um haplótipo único nesse *locus* na Europa. O alelo único que foi selecionado na Europa produz a cor da pele mais clara. Com o distanciamento do gene em qualquer direção, o número de haplótipos aumenta nas populações europeias, tendo em vista que a recombinação perturbou o desequilíbrio de ligação entre o *SLC24A5* e os sítios mais distantes. A pele clara pode ser adaptativa em latitudes ao norte. As pessoas são capazes de

sintetizar vitamina D, mas para tanto, elas precisam absorver a radiação UV pela pele. Nas latitudes equatoriais, as pessoas são expostas a altos níveis de luz UV e conseguem sintetizar a vitamina D até mesmo com a pele fortemente pigmentada. Mais distante do equador, as pessoas são expostas a menos luz UV, e a cor da pele mais clara pode facilitar a síntese de vitamina D nessas latitudes.

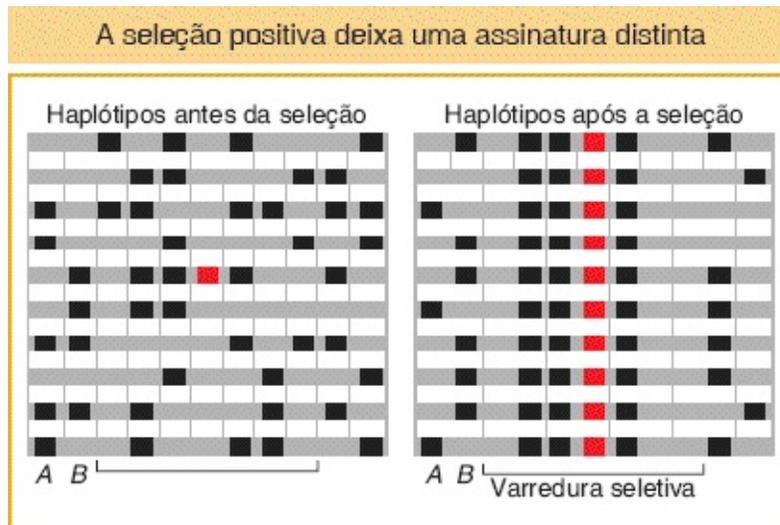


FIGURA 18.22 Esquema de haplótipos observados em uma população antes e depois de um alelo favorecido (*vermelho*) ser varrido até a fixação. Existem 11 *loci* no total. Existem dois alelos (*vermelho* e *cinza*) no *locus* que era o alvo da seleção. Existem dois alelos (*preto* e *cinza*) em cada *locus* ligado ao *locus*-alvo. Após a seleção, o sítio-alvo e alguns vizinhos foram varridos até a fixação.

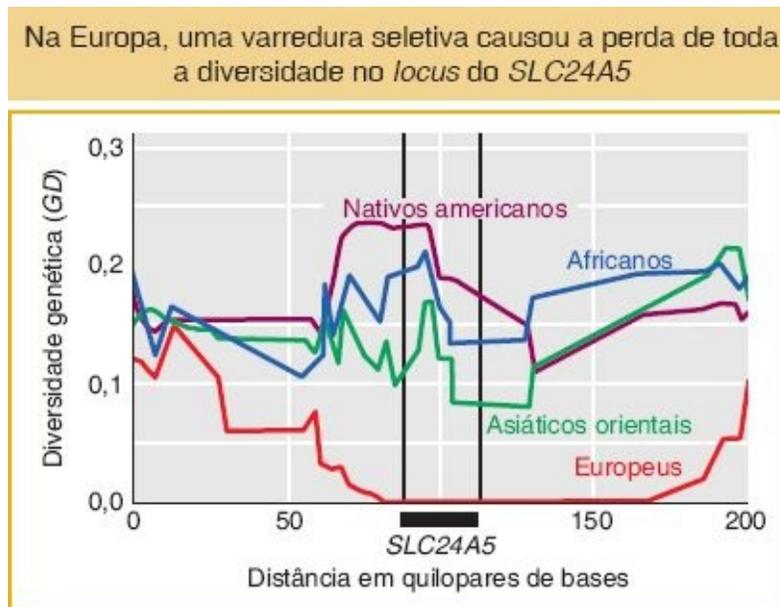


FIGURA 18.23 Diversidade genética em grupos continentais humanos ao longo de um segmento de 2 milhões de pb do cromossomo 15 humano que circunda o gene *SLC24A5*. (Dados de Human Diversity Genome Project, www.hgdp.uchicago.edu.)

A **Tabela 18.6** lista alguns dos genes que demonstram evidência de seleção natural em seres humanos modernos. Esses genes estão situados em algumas poucas categorias básicas. Um grupo fortalece a resistência a patógenos. Os genes *G6PD*, *FY^{nu}* e *Hb* (*hemoglobina B*, o gene da anemia falciforme), todos, ajudam a adaptar os seres humanos à ameaça da malária. A **Figura 18.11 B** demonstra que a frequência de *FY^{nu}* é mais alta na África central. A África central também apresenta a mais alta prevalência de malária, sugerindo que a seleção levou o *FY^{nu}* até a sua mais alta frequência na região na qual a pressão da seleção é maior. Recentemente, geneticistas descobriram o gene *CCR5* (*receptor de quimiocina 5*), que apresenta um alelo (*CCR5-Δ32*) que proporciona resistência à AIDS. Esse alelo atualmente é um alvo da seleção natural. Enquanto existirem patógenos, a seleção natural continuará a operar nas populações humanas.

Outro grupo de genes selecionados na **Tabela 18.6** adapta as pessoas às dietas regionais. Antes de 10.000 atrás, todos os seres humanos eram caçadores-coletores. Mais recentemente, a maior parte dos seres humanos realizou a transferência para os alimentos agrícolas, mas existem diferenças regionais nas dietas. No norte da Europa e em partes da África, os laticínios são uma parte substancial da dieta. Na maior parte das populações, a enzima lactase para a digestão do açúcar do leite (lactose) é expressa durante a infância, mas deixa de ser sintetizada em adultos. Entretanto, em partes da Europa e da África nas quais os adultos bebem leite, alelos especiais do gene da *lactase*, que continuam a expressar a enzima lactase durante a fase adulta, aumentaram em frequência em virtude da seleção natural. Finalmente, a **Tabela 18.6** inclui alguns genes para adaptação fisiológica ao clima. Entre esses estão os genes da pigmentação da pele, tais como o *SLC24A5*, discutido anteriormente.

Tabela 18.6 Alguns genes que demonstram evidência de seleção natural em populações humanas específicas.

Gene	Traço presumido	População
<i>EDA2R</i> (receptor A2 da ectodisplasina)	Calvície-padrão masculina	Europeus
<i>EDAR</i> (receptor A da ectodisplasina)	Morfologia dos cabelos	Asiáticos orientais
<i>FY^{nu}</i> (antígeno <i>Duffy</i>)	Resistência à malária	Africanos
<i>G6PD</i> (glicose-6-fosfato desidrogenase)	Resistência à malária	Africanos
<i>Hb</i> (hemoglobina B)	Resistência à malária	Africanos
<i>KITLG</i> (ligante KIT)	Pigmentação da pele	Asiáticos orientais e europeus
<i>LARGE</i> (glicosiltransferase)	Resistência à febre de Lassa	Africanos
<i>LCT</i> (lactase)	Persistência da lactase; capacidade de digerir o açúcar do leite quando adulto	Africanos, europeus
<i>LPR</i> (receptor de leptina)	Processamento de gorduras alimentares	Asiáticos orientais

<i>MC1R</i> (receptor de melanocortina)	Pigmentação dos cabelos/pelos e da pele	Asiáticos orientais
MHC (complexo principal de histocompatibilidade)	Resistência a doenças infecciosas	Múltiplas populações
<i>OCA2</i> (albinismo oculocutâneo)	Pigmentação da pele e cor dos olhos	Europeus
<i>PPARD</i> (receptor delta ativado por proliferador de peroxissomo)	Processamento de gorduras alimentares	Europeus
<i>SI</i> (sacarase-isomaltase)	Metabolismo da sacarose	Asiáticos orientais
<i>SLC24A5</i> (família transportadora de soluto 24)	Pigmentação da pele	Europeus e asiáticos ocidentais
<i>TYRP1</i> (proteína 1 relacionada à tirosinase)	Pigmentação da pele	Europeus

Fonte: P. C. Sabeti *et al.*, *Science* 312, 2006, 1614-1620; P. C. Saberi *et al.* *Nature* 449, 2007, 913-919; B. F. Voight *et al.* *PLoS Biology* 4, 2006, 446-458; J. K. Pickrell *et al.*, *Genome Research* 19, 2009, 826-837.

Enquanto a seleção direcional causa uma perda de variação genética na região que circunda o *locus*-alvo, a seleção balanceadora consegue prevenir a perda de

diversidade por meio de deriva genética aleatória, levando a regiões de diversidade genética incomumente alta no genoma. Uma região de alta diversidade genética circunda o complexo principal de histocompatibilidade (MHC) no cromossomo 6. A [Figura 18.24](#) demonstra um pico distinto no número de SNP no MHC. Esse complexo inclui os genes do antígeno leucocitário humano (HLA), que estão envolvidos no reconhecimento de (e na resposta a) patógenos por parte do sistema imune. A seleção balanceadora é uma hipótese proposta para explicar a alta diversidade observada no MHC. Tendo em vista que heterozigotos apresentam dois alelos, eles podem ser resistentes a um repertório maior de tipos de patógenos, proporcionando aos heterozigotos uma vantagem adaptativa.

Finalmente, a seleção pode ser imposta por outro agente além da natureza. Os seres humanos impuseram a seleção no processo de domesticação e melhoria de cultivos de plantas e criação de animais. Esse tipo de seleção é denominado **seleção artificial**. Nesse caso, os indivíduos com traços que os seres humanos preferem contribuem com mais alelos para o *pool* gênico do que os indivíduos com os traços não favorecidos. Ao longo do tempo, os alelos que conferem os traços favorecidos aumentam em frequência na população. As muitas raças de cães e vacas leiteiras e de variedades de vegetais de jardinagem e cultivos de cereais são, todas, produtos de seleção artificial.

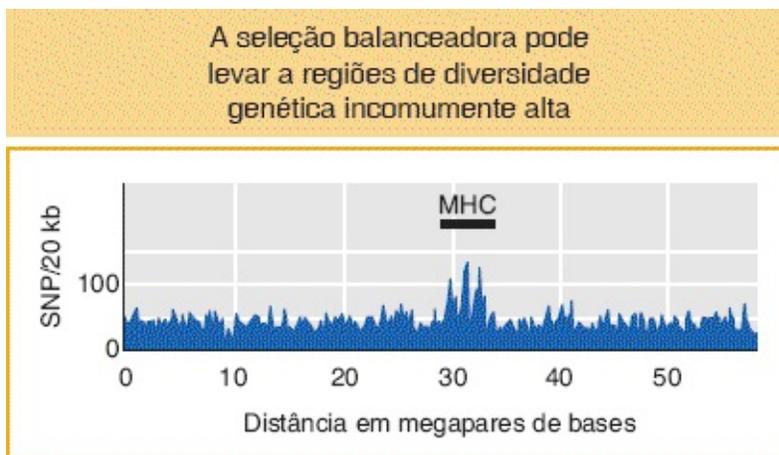


FIGURA 18.24 Número de sítios segregantes (S) ou SNP em janelas de 20 kilopares de bases ao longo do braço curto do cromossomo 6 humano. Existe um pico de alta diversidade no *locus* do MHC. (Dados de International HapMap Project, www.hapmap.org.)

CONCEITO-CHAVE A seleção natural é uma força que pode direcionar alelos favoráveis em um *locus* até a fixação, ou manter múltiplos alelos em um *locus* em uma população. A seleção deixa uma assinatura no genoma, no tipo de padrão de diversidade genética que circunda o alvo da seleção. Geneticistas de populações identificaram um número de genes que têm sido alvo de seleção em seres humanos.

Equilíbrio entre mutação e deriva

Consideramos as forças que regulam a variação nas populações individualmente. Agora consideraremos os efeitos opostos da mutação e da deriva, a primeira adicionando variação e a última removendo-a das populações. Quando essas duas forças estão em equilíbrio, uma população pode alcançar um equilíbrio no qual a perda e o ganho de variação são iguais. Utilizaremos a heterozigosidade (H) como uma medida de variação. Relembre que H estará próximo de 0 quando uma população estiver próxima da fixação em relação a um único alelo (variação baixa), e que H se aproximará de 1 quando houver muitos alelos de frequência igual (variação alta).

Utilizaremos H com acento circunflexo, \hat{H} , como símbolo do valor de equilíbrio de H . Para encontrar \hat{H} , iniciamos com duas equações matemáticas: uma equação que relaciona a alteração em H ao tamanho da população (deriva) e outra equação que relaciona a alteração em H à taxa de mutação. Em seguida, podemos estabelecer essas equações como iguais entre si e calcular \hat{H} .

Primeiramente, precisamos de uma equação para o declínio na variação (H) entre as gerações como uma função do tamanho da população (deriva). Desenvolvemos uma referida equação no [Quadro 18.3](#) quando discutimos sobre o endocruzamento:

$$H' = \left(1 - \frac{1}{2N}\right)H$$

Essa equação se aplica aos efeitos da deriva, bem como àqueles do

endocruzamento. A partir dessa equação, segue-se que a alteração em H entre as gerações em virtude da deriva é:

$$H = H - H' = \frac{1}{2N}H$$

Em segundo lugar, precisamos de uma equação para o aumento na variação, conforme medido por H , entre gerações em virtude da mutação. Qualquer mutação nova aumentará a heterozigosidade a uma taxa proporcional à frequência de homozigotos na população ($1 - H$) vezes a taxa na qual a mutação os converte em heterozigotos (2μ). (O 2 é necessário em virtude da existência de dois alelos que podem sofrer mutação em um diploide.) Portanto, a alteração em H entre as gerações em virtude de mutação é:

$$\Delta H = 2\mu(1 - \hat{H})$$

Quando a população alcançar um equilíbrio, a perda da heterozigosidade por meio de deriva será igual ao ganho por mutação. Portanto, temos:

$$\frac{1}{2N}\hat{H} = 2\mu(1 - \hat{H})$$

que pode ser reescrita como:

$$\hat{H} = \frac{4N\mu}{4N\mu + 1}$$

Essa equação fornece o valor de equilíbrio de \hat{H} quando a perda pela deriva e o ganho por mutação estão equilibrados. Essa equação é aplicável apenas à variação neutra; ou seja, estamos presumindo que a seleção não está atuando. Também estamos presumindo que cada nova mutação produz um alelo único.

Expressões tais como essa são úteis quando temos estimativas em relação a duas das variáveis e gostaríamos de conhecer a terceira. Por exemplo, a diversidade de nucleotídeos (H no nível de nucleotídeo) em relação às sequências não codificadoras, que são amplamente neutras, é de aproximadamente 0,0013 em

seres humanos, e μ em seres humanos é 3×10^{-8} (ver [Tabela 18.5](#)). A utilização desses valores e a solução da equação anterior em relação a N produzem uma estimativa do tamanho da população humana de 10.498 indivíduos. Essa estimativa está muito abaixo dos 7,2 bilhões que somos atualmente. O que acontece? Essa é uma estimativa do valor de equilíbrio. Os seres humanos modernos são um grupo jovem, de apenas aproximadamente 150.000 anos de idade. Ao longo dos últimos 150.000 anos, nossa população cresceu dramaticamente na medida em que preenchemos o globo, mas a mutação é um processo lento e, assim, a diversidade genética não foi mantida e a população humana não está em equilíbrio. O tamanho da população de 10.498 representa uma estimativa de nosso tamanho histórico, ou quantos membros em idade reprodutiva havia há aproximadamente 150.000 anos.

Equilíbrio entre mutação e seleção

As frequências alélicas também podem alcançar um equilíbrio estável quando a introdução de novos alelos por mutação repetida é balanceada por sua remoção por seleção natural. Esse equilíbrio provavelmente explica a persistência de doenças genéticas como polimorfismos de nível baixo em populações humanas. Novas mutações deletérias estão constantemente surgindo de modo espontâneo. Essas mutações podem ser completamente recessivas, ou parcialmente dominantes. A seleção remove-as da população, mas existe um equilíbrio entre o seu aparecimento e a sua remoção.

Iniciaremos com o caso mais simples — a frequência de um recessivo deletério quando é alcançado um equilíbrio entre mutação e seleção. Para essa finalidade, é conveniente expressar o valor adaptativo relativo em termos do **coeficiente de seleção (s)**, que é a desvantagem seletiva de (ou a perda de adaptabilidade em) um genótipo:

W_{AA}	W_{Aa}	W_{aa}
1	1	$1 - s$

Em seguida, conforme demonstrado no [Quadro 18.8](#), a equação em relação à

frequência de equilíbrio de um alelo recessivo deletério é:

$$\hat{q} = \sqrt{\frac{\mu}{s}}$$

Quadro 18.8

Equilíbrio entre seleção e mutação.

Se deixamos q ser a frequência do alelo deletério a e $p = 1 - q$ ser a frequência do alelo normal A , a alteração na frequência alélica em virtude da taxa de mutação μ então é:

$$\Delta q_{mut} = \mu p$$

Um modo simples de expressar os valores adaptativos dos genótipos no caso de um alelo recessivo deletério a é $w_{AA} = w_{Aa} = 1,0$ e $w_{aa} = 1 - s$, em que s , o coeficiente de seleção, é a perda de adaptabilidade nos homozigotos recessivos. Agora podemos substituir esses valores adaptativos em nossa expressão geral pela alteração da frequência alélica (ver [Quadro 18.7](#)) e obter:

$$\Delta q_{sel} = \frac{-pq(sq)}{1 - sq^2} = \frac{-spq^2}{1 - sq^2}$$

O equilíbrio significa que o aumento na frequência alélica em virtude de mutação equilibra exatamente a diminuição na frequência alélica em virtude da seleção; assim:

$$\mu \hat{p} = \frac{-s \hat{p} \hat{q}^2}{1 - s \hat{q}^2}$$

A frequência de um alelo recessivo deletério (q) no equilíbrio será consideravelmente pequena, então $1 - sq^2 \approx 1$, e temos:

$$\begin{aligned} \mu \hat{p} &= -s \hat{p} \hat{q}^2 \\ \hat{q} &= \sqrt{\frac{\mu}{s}} \end{aligned}$$

no equilíbrio.

Essa equação demonstra que a frequência no equilíbrio depende da razão μ/s . Quando a taxa de mutação de $A \rightarrow a$ se tornar maior e a desvantagem seletiva se tornar menor, a frequência no equilíbrio (\hat{q}) de um alelo recessivo deletério aumentará. Como um exemplo, um alelo letal recessivo ($s = 1$) que surge por meio de mutação do alelo selvagem na taxa de $\mu = 10^{-6}$ apresentará uma frequência de equilíbrio de 10^{-3} .

Consideraremos o equilíbrio entre a seleção e a mutação em relação ao caso discretamente mais complicado de um alelo deletério parcialmente dominante — ou seja, um alelo com algum efeito deletério em heterozigotos, bem como seu efeito em homozigotos. Definiremos h como o grau de dominância do alelo deletério. Quando h é 1, o alelo deletério é totalmente dominante, e quando h é 0, o alelo deletério é totalmente recessivo. Em seguida, os valores adaptativos são

W_{AA}	W_{Aa}	W_{aa}
1	$1 - hs$	$1 - s$

em que a é um alelo deletério parcialmente dominante. Uma derivação semelhante àquela no [Quadro 18.8](#) nos fornece:

$$\hat{q} = \frac{\mu}{hs}$$

Aqui está um exemplo. Se $\mu = 10^{-6}$ e o alelo letal não for totalmente recessivo, mas causar uma redução de 5% na adaptabilidade dos heterozigotos ($s = 1,0$, $h = 0,05$), então:

$$\hat{q} = \frac{\mu}{hs} = 2 \times 10^{-5}$$

Esse resultado é menor em duas ordens de magnitude do que a frequência de equilíbrio em relação ao caso puramente recessivo descrito anteriormente. Em

geral, então, podemos esperar que alelos deletérios completamente recessivos apresentem frequências muito mais altas do que aquelas de alelos parcialmente dominantes, tendo em vista que os alelos recessivos estão protegidos nos heterozigotos.

CONCEITOS-CHAVE A quantidade de variação genética nas populações representa um equilíbrio entre forças opostas: mutação e migração, que adicionam nova variação, *versus* deriva e seleção, que removem a variação. A seleção balanceadora também atua para manter a variação nas populações. Como resultado desses processos, as frequências alélicas podem alcançar valores de equilíbrio, explicando o motivo pelo qual as populações com frequência mantêm altos níveis de variação genética.

18.6 Aplicações biológicas e sociais

Assim como os princípios da física orientam os engenheiros que projetam pontes e jatos, também assim os princípios da genética de populações tocam todas as nossas vidas de muitas maneiras, despercebidas. No [Capítulo 19](#), você verá como a genética de populações figura de modo proeminente na busca por genes que contribuem para o risco de doenças em pessoas, com a utilização de conceitos tais como o desequilíbrio de ligação descrito neste capítulo. Nesta seção final do capítulo, examinaremos quatro outras áreas nas quais os princípios da genética de populações estão sendo aplicados em questões que afetam as sociedades modernas.

Genética da conservação

Biólogos conservacionistas que estão tentando salvar espécies selvagens ameaçadas, e profissionais de zoológicos que estão tentando manter pequenas populações de animais em cativeiro, com frequência realizam análises de genética de populações. Anteriormente, discutimos como um gargalo genético causou perda de variação genética no condor-da-califórnia e um aumento na

frequência de uma forma letal de nanismo. Os gargalos também podem aumentar o nível de endocruzamento em uma população, levando, talvez, a um declínio na adaptabilidade por meio da depressão por endocruzamento. Entretanto, a questão é complexa, tendo em vista que o endocruzamento nem sempre está associado a um declínio na adaptabilidade. O endocruzamento por vezes pode ajudar a remover alelos recessivos deletérios de uma população. A seleção purificadora é mais eficaz na eliminação de alelos recessivos deletérios, tendo em vista que a classe homozigota recessiva se torna mais frequente nas populações endocruzadas. Portanto, biólogos conservacionistas têm debatido se devem tentar maximizar a diversidade genética e minimizar o endocruzamento, ou submeter deliberadamente as populações de zoológicos ao endocruzamento com o objetivo de remover os alelos deletérios.

Para ajudar a abordar essa questão, pesquisadores procuraram por evidências de sucesso na remoção entre populações de zoológicos. Definiremos a depressão por endocruzamento como delta (δ):

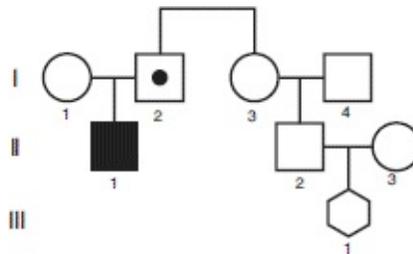
$$\delta = 1 - \frac{w_f}{w_0}$$

em que w_f é a adaptabilidade de indivíduos endocruzados e w_0 é o valor adaptativo de indivíduos não endocruzados. O valor de δ será positivo quando houver um declínio na adaptabilidade com o endocruzamento, mas negativo quando a adaptabilidade melhorar com o endocruzamento. Pesquisadores calcularam δ para 119 populações de zoológico, incluindo 88 espécies, e observaram evidências de que a remoção havia melhorado a adaptabilidade (valores negativos para δ) em 14 populações. Ainda assim, não está claro se o endocruzamento deliberado de animais de zoológico é recomendável. Por um lado, embora 14 das 119 populações tenham melhorado, a maioria das populações declinou na adaptabilidade quando foi endocruzada. Portanto, se iniciarmos com uma pequena população de zoológico e realizarmos propositalmente o endocruzamento dos animais, um declínio na adaptabilidade será o desfecho mais provável.

Cálculo dos riscos de doenças

No [Capítulo 2](#), vimos como os alelos de distúrbios genéticos podiam ser traçados em heredogramas e discutimos como calcular o risco de que um casal tenha um filho que herde tal distúrbio. Os princípios da genética de populações nos possibilitam estender esse tipo de análise. Consideraremos dois exemplos.

O alelo em relação à fibrose cística (FC) ocorre a uma frequência de aproximadamente 0,025 em caucasianos. No heredograma de uma família caucasiana a seguir, o indivíduo II-2 apresenta um primo em primeiro grau (II-1) com fibrose cística. II-2 é casado com uma caucasiana não relacionada (II-3), e eles estão planejando ter um filho. Qual é a chance de que o filho (III-1) apresente fibrose cística?



Um dos avós maternos de II-2 deve ter sido um portador. Começamos calculando a probabilidade de que III-1 herde esse alelo da fibrose cística desse avô ou dessa avó através de seu pai, II-2, utilizando métodos já familiares do [Capítulo 2](#). A probabilidade de que esse avô ou essa avó tenha transmitido o alelo da doença para I-3 é $1/2$. A probabilidade de que I-3 o tenha transmitido para II-2 e de que II-2 o transmita para III-1 também são, ambas, $1/2$. Assim, a probabilidade de III-1 herdar o alelo da FC de II-2 é $(\frac{1}{2})^3$, ou $1/8$. Agora estendemos o cálculo para determinar a probabilidade de que III-1 possa herdar o alelo da fibrose cística de sua mãe, II-3. O indivíduo II-3 não apresenta FC, mas não temos certeza se ela é ou não uma portadora. Se a frequência (q) do alelo da doença na população é de 0,025, então a probabilidade de que um indivíduo não afetado, tal como II-3, seja um portador é de $2pq/(1 - q^2) = 0,049$. Se II-3 for portadora, existe então uma chance de $\frac{1}{2}$ de que ela transmita o alelo da doença para III-1. Essas são todas probabilidades independentes; assim, podemos utilizar a regra do produto. A

probabilidade de que III-1 apresente fibrose cística é:

$$\frac{1}{8} \times \frac{1}{2} \times 0,049 = 0,003$$

A frequência de fibrose cística entre caucasianos é $p^2 = (0,025)^2 = 0,000625$. Esses cálculos nos informam que os indivíduos que têm um primo em primeiro grau com fibrose cística apresentam um risco $0,003 \div 0,000625 = 4,9$ vezes mais alto de ter um filho com a doença do que membros da população em geral.

Aqui está outra aplicação da genética de populações para avaliar o risco de doenças. A anemia falciforme, uma doença recessiva, apresenta uma frequência de aproximadamente 0,25%, ou 1 em 400, entre afro-americanos (ver [Capítulo 6](#)). Aplicando a lei de Hardy-Weinberg, estimamos a frequência do alelo da doença (Hb^S) como 0,05. Qual seria a frequência esperada dessa doença entre a descendência de afro-americanos que são primos de primeiro grau? Utilizando o método descrito no [Quadro 18.2](#), calculamos que o coeficiente de endocruzamento (F) em relação à descendência de casamentos entre primos de primeiro grau é de 1/16. Na seção anterior sobre o endocruzamento, vimos que a frequência de homozigotos aumenta quando existe endocruzamento, conforme demonstrado por essa equação:

$$f_{aa} = q^2 + pqF$$

Com a utilização dessa equação, obtemos:

$$f(Hb^S/Hb^S) = (0,05)^2 + (0,05 \times 0,95) \frac{1}{16} = 0,0055$$

Isso representa um aumento de 2,2 vezes no risco de ter um filho com a doença para casamentos de primos de primeiro grau, em comparação àquele em um casamento entre indivíduos não relacionados.

DNA forense

Criminosos podem deixar evidências de DNA na cena de um crime na forma de

sangue, sêmen, cabelos/pelos ou até mesmo células bucais da saliva em uma ponta de cigarro. A reação da cadeia de polimerase (PCR) possibilita que cientistas forenses amplifiquem quantidades muito pequenas de DNA e determinem o genótipo do indivíduo que deixou a amostra. Se o DNA encontrado em uma cena de crime corresponder àquele do suspeito, então eles “podem ser” o mesmo indivíduo. A frase-chave aqui é “podem ser”, e é aqui que a genética de populações realiza o seu papel. Vejamos como isso funciona.

Considere dois *loci* de microssatélites, cada um com múltiplos alelos: $A_1, A_2 \dots A_n$ e $B_1, B_2 \dots B_n$. Cientistas forenses determinam que uma amostra de DNA de uma cena de crime e uma do suspeito são ambas $A_3/A_8 B_1/B_7$. Eles determinaram que existe uma “correspondência” entre a evidência e o suspeito. A correspondência comprova que a evidência do DNA é originária do suspeito? Ela prova que o suspeito estava na cena do crime?

O que os geneticistas de populações fazem com esse tipo de evidência é testar uma hipótese específica: *a evidência é originária de outra pessoa além do suspeito*. Isso é o que os estatísticos denominam “hipótese nula”, ou a hipótese que é considerada verdadeira, exceto se evidências demonstrarem que ela é muito improvável (ver [Capítulo 4](#)). Para realizar o teste, calculamos a probabilidade de observar uma correspondência entre a evidência e o suspeito, dado que o suspeito e a pessoa que deixou a evidência são indivíduos diferentes. Simbolicamente, escrevemos:

$$\text{Prob}(\text{correspondência} \mid \text{indivíduos diferentes})$$

em que “|” significa “dado que”. Se essa probabilidade for muito pequena, podemos então rejeitar a hipótese nula e argumentar em favor de uma hipótese alternativa: *a evidência foi deixada pelo suspeito*. Nunca comprovamos formalmente que o suspeito deixou a evidência, tendo em vista que pode haver hipóteses alternativas, tais como: *a evidência foi deixada pelo gêmeo idêntico do suspeito*.

Para calcular a probabilidade de observar uma correspondência entre a evidência e o suspeito se a evidência for originária de um indivíduo diferente,

precisamos conhecer as frequências dos alelos de microssatélites na população.

A_4	0,03
A_6	0,05
B_1	0,01
B_7	0,12

A prob (correspondência | indivíduos diferentes) é a mesma probabilidade de que a evidência seja originária de um indivíduo escolhido aleatoriamente. Podemos calcular essa probabilidade utilizando as frequências alélicas anteriores. Primeiramente, presumiremos que a lei de Hardy-Weinberg seja aplicável e calcularemos a probabilidade de ser A_4/A_6 no primeiro *locus* e B_1/B_7 no segundo:

$$\text{Prob}(A_4/A_6) = 2pq = 2 \times 0,03 \times 0,05 = 0,003$$

$$\text{Prob}(B_1/B_7) = 2 \times 0,01 \times 0,12 = 0,0024$$

Para combinar essas duas probabilidades, precisamos realizar mais uma presunção. Precisamos presumir que os dois *loci* são *independentes*; ou seja, que os *loci* estão em *equilíbrio de ligação*. Ao realizar essa presunção, podemos aplicar a *regra do produto* a eventos independentes (ver [Capítulo 2](#)) e determinar que:

$$\begin{aligned} \text{Prob}(\text{correspondência} | \text{indivíduos diferentes}) = \\ \text{Prob}(A_4/A_6) \times \text{Prob}(B_1/B_7) = 7,2 \times 10^{-6} \end{aligned}$$

Portanto, a probabilidade sob a hipótese nula de que a evidência seja originária de alguma outra pessoa além do suspeito é de $7,2 \times 10^{-6}$, ou aproximadamente 7 em um milhão. Essa é uma probabilidade pequena e, assim, a hipótese nula aparenta ser improvável nesse caso. Entretanto, se a Prob (correspondência | indivíduos diferentes) fosse 0,1, 10% da população seria correspondente e poderia ter deixado a evidência. Naquele caso, não rejeitaríamos a hipótese nula.

Dois microssatélites não proporcionam muito poder de discriminação, e então o FBI, nos EUA, utiliza um conjunto de 13 microssatélites. Os *loci* de microssatélites apresentam tipicamente grandes números de alelos (10 a 20, ou mais); portanto, o número de possíveis genótipos com base em 13 microssatélites é astronomicamente grande. Com 10 alelos por *locus*, existem 55 possíveis genótipos em cada *locus* e 55^{13} , ou $4,2 \times 10^{22}$, possíveis genótipos *multilocus* para 13 *loci*. O FBI também reuniu uma base de dados denominada CODIS (Combined DNA Index System) que contém as frequências de diferentes alelos nesses *loci* na população, incluindo dados específicos de diferentes grupos étnicos e regiões do país.

Procura no Google por seus parceiros de DNA

Neste capítulo, revisamos os princípios básicos da genética de populações e discutimos muitas aplicações para a genética humana. A teoria básica da genética de populações tem estado presente há quase 100 anos, mas apenas na última década o desenvolvimento de tecnologias com base no alto processamento do DNA para a genotipagem de indivíduos trouxe à tona os complexos padrões de variação entre e dentro das populações humanas. Não apenas tem sido possível desembaraçar muitos dos detalhes a respeito de como e quando os seres humanos povoaram o globo a partir da sua origem na África, como também os geneticistas obtiveram uma profunda compreensão sobre como forças tais como a seleção natural e a deriva genética modelaram quem somos.

O que o futuro reserva para nós? Logo o sequenciamento de um genoma humano poderá custar um pouco mais do que uma nova bicicleta. Uma estudante universitária poderá esfregar o interior de sua bochecha com um *Q-tip* e depositar a amostra em um quiosque enquanto ouve música. Semanas depois, ela poderá visualizar a sequência de seu genoma em um *site*, compará-la àquela de seus amigos e parentes, e aprender a respeito de seus ancestrais. Conforme veremos no próximo capítulo, a nossa capacidade de prever os riscos de doenças, os talentos e outros traços de uma pessoa a partir do seu genótipo está melhorando. Até a medida na qual o gosto de uma pessoa pela música ou a admiração por esportes radicais apresenta fundamentos genéticos, teoricamente uma pessoa poderia

“buscar no Google” parceiros de DNA que provavelmente compartilham os mesmos interesses. A tecnologia do DNA e a teoria da genética de populações já estão funcionando; entretanto, existem questões sociais e éticas a serem abordadas. As informações poderão ser mantidas em sigilo, e como? Existem muitos limites sobre o que uma pessoa deve saber a respeito de sua própria sequência? O governo deve sequenciar o genoma de qualquer pessoa quando ela nasce? Planos de saúde podem exigir que seus clientes submetam as sequências do seu genoma? Uma compreensão sobre a ciência pode auxiliar na determinação de como essas questões são respondidas.

RESUMO

A genética de populações procura compreender as leis que regulam e as forças que influenciam a quantidade de variação genética nas populações e as alterações na variação genética ao longo do tempo. O conceito do *pool* gênico fornece um modelo para pensarmos a respeito da transmissão da variação genética de uma geração para a próxima em relação a uma população inteira. A teoria básica da genética de populações tem início com uma população idealizada de tamanho infinito e na qual o cruzamento é aleatório. Em uma referida população, a lei de Hardy-Weinberg define a relação entre as frequências alélicas no *pool* gênico e as frequências genóticas na população.

As populações reais normalmente se desviam, em pequeno ou grande grau, do modelo de Hardy-Weinberg. Uma fonte de desvio advém do tipo de cruzamento não aleatório ou preferencial. Se os indivíduos cruzam preferencialmente com outros que compartilham um fenótipo semelhante, então haverá um excesso de homozigotos para genes que controlam aquele fenótipo em comparação às expectativas de Hardy-Weinberg. Quando os indivíduos cruzam mais frequentemente com parentes do que o esperado ao acaso, haverá então um excesso de genótipos homozigotos em todo o genoma e a população se torna endocruzada. Até mesmo quando as populações locais de uma espécie estão de acordo com as expectativas de Hardy-Weinberg, aquelas estão aptas a ser isoladas de outras populações em locais distantes. Portanto, uma espécie com

frequência é composta por uma série de subpopulações geneticamente distintas; ou seja, as espécies demonstram estrutura genética de população.

Diversas forças podem adicionar novas variações a uma população ou remover dela a variação existente. A mutação é a fonte definitiva de toda a variação genética. Geneticistas de populações determinaram estimativas razoavelmente precisas da taxa na qual surgem novas mutações nas populações. A migração também pode trazer nova variação para uma população. A migração resulta em alguns indivíduos que são geneticamente misturados, que apresentam ancestrais de múltiplas populações. A recombinação genética também pode adicionar variação às populações por meio da recombinação de alelos em novos haplótipos.

Duas forças controlam o destino da variação genética nas populações. Primeiramente, a deriva genética é uma força aleatória que pode levar à perda ou à fixação de um alelo como resultado do erro de amostragem em populações finitas. A deriva age com mais força sobre populações pequenas e menor sobre populações grandes. Em segundo lugar, a seleção natural direciona alterações nas frequências alélicas nas populações ao longo do tempo. Os alelos que intensificam a adaptabilidade dos indivíduos que os carregam aumentarão em frequência e poderão se tornar fixos, enquanto alelos deletérios que reduzem a adaptabilidade serão removidos da população.

O objetivo fundamental da genética de populações é compreender as contribuições relativas dos sistemas de cruzamento, mutação, migração, recombinação, deriva e seleção natural para a quantidade e a distribuição da variação genética nas populações. Neste capítulo, vimos como pesquisas em genética de populações desenvolveram a teoria básica e coletaram uma vasta quantidade de dados para a conquista desse objetivo. Nossa compreensão sobre a genética de populações de nossa própria espécie é notavelmente detalhada.

Finalmente, os métodos e os resultados da genética de populações nos informam a respeito do processo evolutivo e apresentam aplicações práticas para questões que as sociedades modernas enfrentam. A teoria e as análises da genética de populações desempenham papéis importantes no manejo de espécies ameaçadas, na identificação de criminosos, no cultivo de plantas e na criação

animal, e na avaliação dos riscos de um casal ter um filho com uma doença.

TERMOS-CHAVE

adaptabilidade absoluta
adaptabilidade darwiniana
adaptabilidade relativa
adaptação
alelo neutro
coeficiente de endocruzamento
coeficiente de seleção (s)
cruzamento preferencial negativo
cruzamento preferencial positivo
depressão por endocruzamento
deriva genética aleatória
desequilíbrio de ligação (LD)
diversidade gênica (GD)
diversidade nucleotídica
efeito fundador
endocruzamento
equilíbrio de Hardy-Weinberg
equilíbrio de ligação
estrutura populacional
evolução neutra
fixado
fluxo gênico
frequência alélica
frequência genotípica
gargalo
genética de populações
haplótipo
HapMap

heterozigosidade (H)
idêntico por descendência (IBD)
isolamento pela distância
lei de Hardy-Weinberg
locus
microsatélite
migração
mistura genética
número de haplótipos (NH)
painel de descoberta
polimorfismo de nucleotídeo único (SNP)
pool gênico
população
rede de haplótipos
relógio molecular
seleção artificial
seleção balanceadora
seleção direcional
seleção natural
seleção positiva
seleção purificadora
sítios segregantes (S)
SNP comum
SNP raro
taxa de mutação (μ)

PROBLEMAS RESOLVIDOS

Problema resolvido 1. Aproximadamente 70% de todos os caucasianos conseguem degustar a substância química feniltiocarbamida, e o restante não consegue. A capacidade de degustar essa substância química é determinada pelo alelo dominante T , e a incapacidade de degustar é determinada pelo alelo

recessivo t . Se presumirmos que a população está em equilíbrio de Hardy-Weinberg, quais são as frequências genotípicas e alélicas nessa população?

Solução

Tendo em vista que 70% são degustadores (T/T e T/t), 30% devem ser não degustadores (t/t). Essa frequência de homozigotos recessivos é igual a q^2 ; assim, para obter q , simplesmente calculamos a raiz quadrada de 0,30:

$$q = \sqrt{0,30} = 0,55$$

Tendo em vista que $p + q = 1$, podemos escrever $p = 1 - q = 1 - 0,55 = 0,45$.

Agora podemos calcular:

$$p^2 = (0,45)^2 = 0,20, \text{ a frequência de } T/T$$

$$2pq = 2 \times 0,45 \times 0,55 = 0,50, \text{ a frequência de } T/t$$

$$q^2 = 0,3, \text{ a frequência de } t/t$$

Problema resolvido 2. Em uma grande população experimental de *Drosophila*, o valor adaptativo relativo de um fenótipo recessivo é calculado como 0,90, e a taxa de mutação em relação ao alelo recessivo é 5×10^{-5} . Se a população estiver em equilíbrio, quais frequências alélicas podem ser previstas?

Solução

Aqui, a mutação e a seleção estão atuando em direções opostas e, assim, é previsto um equilíbrio. Tal equilíbrio é descrito pela fórmula:

$$\hat{q} = \sqrt{\frac{\mu}{s}}$$

Na presente questão:

$$\mu = 5 \times 10^{-5} \text{ e } s = 1 - w = 1 - 0,9 = 0,1$$

Portanto:

$$\hat{q} = \sqrt{\frac{5 \times 10^{-5}}{0,1}} = 0,022$$

$$\hat{p} = 1 - 0,022 = 0,978$$

Problema resolvido 3. Uma colônia de 50 papagaios-do-mar (*Fratercula corniculata*) é estabelecida em um zoológico e ali é mantida durante 30 gerações.

a. Se o coeficiente de endocruzamento dos membros fundadores for zero ($F = 0,0$), qual é o coeficiente de endocruzamento esperado para essa população atualmente?

b. Em relação a um alelo deletério para doença, com uma frequência de 0,001 na natureza, qual é a frequência prevista de aves homozigotas afetadas na natureza e na população do zoológico atualmente?

Solução

a. No [Quadro 18.3](#), vimos que o endocruzamento aumentará como uma função do tamanho da população (N) ao longo do tempo (t), conforme medido nas gerações, de acordo com a equação a seguir:

$$F_t = 1 - \left(1 - \frac{1}{2N}\right)^t (1 - F_0)$$

Substituindo por $N = 50$, $t = 30$ e $F_0 = 0$, obtemos:

$$F_{30} = 1 - \left(1 - \frac{1}{2 \times 50}\right)^{30} (1 - 0) = 0,26$$

b. Se a frequência de um alelo de doença recessiva (q) na natureza é 0,001, então por meio da aplicação da lei de Hardy-Weinberg prevemos que a frequência de

indivíduos homozigotos afetados na natureza será $q^2 = 10^{-6}$. Em relação à população do zoológico, a frequência de homozigotos será mais alta em virtude do endocruzamento, de acordo com a equação a seguir:

$$f_{a/a} = q^2 + pqF$$

Substituindo por $q = 0,001$, $p = 0,999$ e $F = 0,26$, obtemos:

$$f_{a/a} = 10^{-6} + (0,001 \times 0,999 \times 0,26) = 2,61 \times 10^{-4}$$

A razão de $2,61 \times 10^{-4}$ a 10^{-6} nos demonstra que existe um aumento de 261 vezes na frequência esperada de indivíduos afetados na população do zoológico atual em comparação à população selvagem ancestral.

Problema resolvido 4. Em um julgamento penal, o promotor apresenta os genótipos em relação a três *loci* de microssatélites do conjunto CODIS do FBI. Ele relata que uma amostra de DNA da cena do crime e uma do suspeito apresentam, ambas, o genótipo FGA_1/FGA_4 , $TPOX_1/TPOX_3$, VWA_2/VWA_7 nesses três microssatélites. Ele também apresenta as frequências alélicas em relação à população em geral à qual o suspeito pertence (ver tabela a seguir). Qual é a probabilidade de que o genótipo da evidência de DNA corresponda àquele do suspeito, dado que a pessoa que cometeu o crime e o suspeito sejam indivíduos diferentes? Quais presunções você realiza quando calcula essa probabilidade?

Alelo	Frequência
FGA_1	0,30
FGA_4	0,26
$TPOX_1$	0,32
$TPOX_3$	0,65
VWA_2	0,23
VWA_7	0,59

Solução

A probabilidade de que o genótipo da evidência de DNA corresponda àquele do suspeito, dado que a pessoa que cometeu o crime e o suspeito sejam indivíduos diferentes, é a mesma probabilidade de que um membro da população escolhido aleatoriamente apresente o mesmo genótipo da evidência de DNA. A probabilidade de uma pessoa escolhida aleatoriamente ser $FGA_1/FGA_4 = 2pq = 2(0,30)(0,26) = 0,156$ e, de modo semelhante, a probabilidade de uma pessoa aleatória ser $TPOX_1/TPOX_3 = 0,416$ e $VWA_2/VWA_7 = 0,2714$. Aplicando a regra da multiplicação, a probabilidade de um membro aleatório da população ser $FGA_1/FGA_4, TPOX_1/TPOX_3, VWA_2/VWA_7 = 0,156 \times 0,416 \times 0,2714 = 0,0176$. Ao calcular essa probabilidade, presumimos que a população está em equilíbrio de Hardy-Weinberg e que os três *loci* em questão estão em equilíbrio de ligação entre si.

PROBLEMAS

QUESTÕES SOBRE AS FIGURAS

1. Qual indivíduo na [Figura 18.3](#) apresenta o maior número de *loci* heterozigotos, e qual indivíduo apresenta o menor número?
2. Suponha que os setes cromossomos na [Figura 18.4 A](#) representem uma amostra aleatória de cromossomos de uma população.
 - a. Calcule a diversidade genética (GD) em separado em relação ao indel, o *locus* de microssatélite e o SNP na posição 3.
 - b. Se a sequência foi encurtada de modo que você tivesse os dados apenas em relação às posições 1 a 24, quantos haplótipos haveria ali?
 - c. Calcule o parâmetro do desequilíbrio de ligação (D) entre os SNP nas posições 29 e 33.
3. Observando a [Figura 18.6](#), você consegue contar quantos haplótipos mitocondriais foram transportados da Ásia para as Américas?
4. Na [Figura 18.13](#), a coluna “não relacionados” (*azul*) em relação ao Japão é mais alta do que a coluna “não relacionados” em relação à França. O que isso lhe informa?