



Revista Eletrônica
Paulista de Matemática

ISSN 2316-9664
Volume 13, dez. 2018

Vinicius Freitas de Oliveira

UNESP - Universidade Estadual
Paulista “Júlio de Mesquita
Filho”
vinicius.oliveira@ucb.org.br

Suzete Maria Silva Afonso

UNESP - Universidade Estadual
Paulista “Júlio de Mesquita
Filho”
smafonso@rc.unesp.br

Modelos matemáticos na genética de populações

Mathematical models in genetics of populations

Resumo

A genética de populações estuda o comportamento das frequências genotípicas e alélicas no decorrer do tempo em populações, bem como os eventos que podem modificá-las. Com a finalidade de realizar aplicações de Matemática em Biologia, este trabalho exhibe quatro modelos matemáticos que expressam o comportamento das frequências gênicas em populações que não possuem a presença de seleção natural. Assim, foram utilizadas noções de probabilidade, estatística e equações de diferenças como ferramentas principais na modelagem. O primeiro modelo a ser apresentado foi o modelo básico, que contou com cinco premissas iniciais. Já os demais modelos foram construídos alternando essas premissas. Por fim, para a análise dos modelos, foi avaliada a possibilidade de contemplação do equilíbrio de Hardy-Weinberg, o que reafirmou a ocorrência do equilíbrio apenas para populações que não possuem mutação, seleção natural ou migração.

Palavras-chave: Genética de Populações. Equações de Diferenças. Modelagem Matemática. Biomatemática.

Abstract

Population genetics studies the behavior of genotypic and allelic frequencies over time in populations, as well as the events that can modify them. In order to realize Mathematical applications in Biology, this work presents four mathematical models that express the behavior of the gene frequencies in populations that do not have the presence of natural selection. Thus, notions of probability, statistics and difference equations were used as main tools in modeling. The first model to be presented was the basic model, which had five initial assumptions. Already the other models were constructed alternating these premises. Finally, for the analysis of the models, the possibility of contemplating the Hardy-Weinberg equilibrium was evaluated, which reaffirmed the occurrence of equilibrium only for populations that do not have mutation, natural selection or migration.

Keywords: Population Genetics. Difference Equations. Mathematical Modeling. Biomathematics .



1 Introdução

Com a necessidade de estudar o comportamento das frequências genóticas e alélicas no tempo surge a ciência da genética de populações. Essa ciência é otimizada através de uma interdisciplinaridade entre as áreas da matemática e da biologia, e assim sendo, tem ascendido muito nos últimos anos.

Os primeiros estudos voltados para os comportamentos genéticos foram realizados por Francis Galton (1822-1911), que realizou aplicações da estatística na biologia. O avanço desses estudos foram muitos e atualmente eles contam com a possibilidade de estudar genética sem focar nas mutações visíveis e sem cruzamentos controlados, ampliando as opções de pesquisa (HARTL; CLARK, 2007).

As aplicações desse estudo são diversas, segundo Hartl e Clark (2007), além de revelar conhecimentos sobre os seres humanos, também é possível encontrar implicações na ética, políticas sociais, medicina, agricultura e outros. De forma mais direta, os autores ainda dizem que através da genética de populações é possível organizar programas para a conservação de espécies ameaçadas, desenvolver levantamentos populacionais de portadores de genes com doenças e identificar genes de suscetibilidade a doenças em humanos, incluindo o câncer.

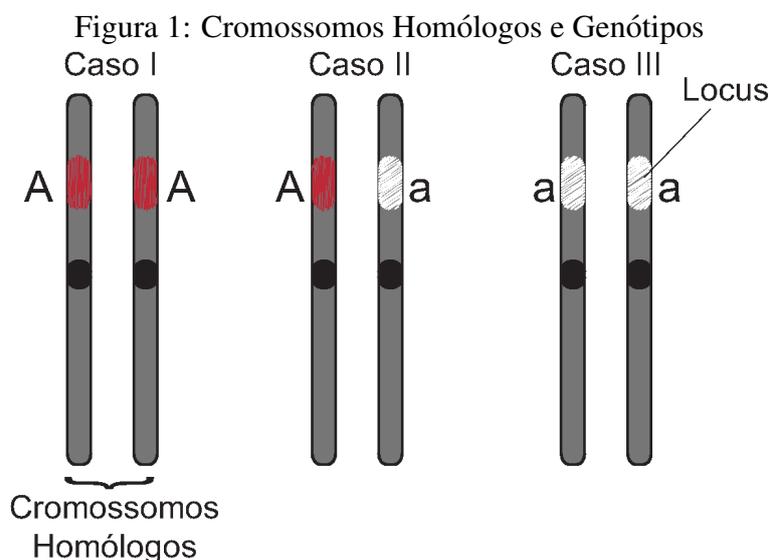
Os conceitos de genética e hereditariedade são fundamentais para o estudo da genética de populações e a sua compreensão tem como ponto de partida a célula, um sistema estrutural dos organismos vivos. As células são sistemas complexos, compostos por diversas organelas, cada uma com um objetivo de funcionamento. Todas as informações genéticas (gene) que são passadas de pai para filho representam um trecho da partícula de ácido desoxirribonucleico (ADN, em português: ácido desoxirribonucleico; ou DNA, em inglês: deoxyribonucleic acid) e são reservadas no núcleo da célula (AMABIS; MARTHO, 1994).

Os genes são localizados em um espaço chamado locus, ou loci no plural. Segundo Lewis (2004), nesses espaços os genes sofrem variações na forma — essas mudanças são denominadas alelo. Cada alelo tem direta influência na característica herdada. Representados por letras do nosso alfabeto, eles podem ser apontados como dominantes ou recessivos. Alelos dominantes, exibidos por letras maiúsculas, são sempre expressos, inibindo os demais alelos. Dessa forma, os alelos recessivos, representados por letras minúsculas, só possuem expressão quando estão sozinhos.

Outra estrutura fundamental são os cromossomos e eles podem ser classificados como sexuais ou autossômicos. Os cromossomos sexuais determinam o sexo dos indivíduos. No caso humano os cromossomos X pareados configuram o gênero feminino e os cromossomos X e Y juntos representam o gênero masculino. Já os cromossomos autossômicos não tem dependência sobre os gêneros, porém fazem parte do processo de todas as demais características de um indivíduo (LEWIS, 2004).

Os loci fazem parte da estrutura do cromossomo, conseqüentemente, cada cromossomo pode portar uma quantidade específica de alelos. Espécies denominadas diploides possuem todos os seus cromossomos dispostos em duplas homólogas, assim, o conjunto dos alelos arranjados nas duplas são chamados de genótipos (LEWIS, 2004). A Figura 1 ilustra esse conceito.

No primeiro caso da Figura 1 é possível apontar o genótipo **AA** — composto por dois alelos dominantes **A**. Para o segundo caso contempla-se o genótipo **Aa** e por fim, no terceiro caso, o genótipo **aa**. Classificam-se os genótipos **AA** e **aa** como homozigotos, por possuírem alelos iguais, e o genótipo **Aa** como heterozigoto, por possuir alelos diferentes.



FONTE: Os Autores (2018)

De acordo com Amabis e Martho (1994), os genótipos possuem a informação necessária para a produção de uma proteína que dá origem a uma característica — fenótipo. Hartl e Clark (2007) dizem que existe uma grande diferença entre o genótipo e o fenótipo, pois o fenótipo pode sofrer influência direta com o ambiente em que o indivíduo vive, ao contrário do genótipo, que é apenas hereditário. Por exemplo, a cor da pele é uma característica, portanto, um fenótipo. Assim, existe um genótipo responsável por carregar a informação de produção da proteína melanina, porém, a exposição ao sol, ambiente externo, pode influenciar na cor da pele do indivíduo.

Utilizando os conceitos biológicos aqui expostos, o presente artigo tem como objetivo realizar aplicações matemáticas na genética de populações através da construção de quatro modelos matemáticos. Cabe mencionar que essa construção não é inédita, podendo também ser verificada em Ocone (2014). Para a modelagem serão utilizadas noções de probabilidade, estatística e equações de diferenças como ferramentas.

2 Preliminares sobre equações de diferenças

Uma ferramenta indispensável na construção de modelos de genética populacional são as equações de diferenças. As equações de diferenças exprimem o comportamento de fenômenos no decorrer do tempo cuja medida é determinada por intervalos discretos, $t \in \mathbb{N}$. Essas equações, também chamadas de equações discretas, podem ser aplicadas em diversas áreas, como biologia, economia e física.

Todas as noções aqui expostas também podem ser vistas em Elaydi (2005).



2.1 Equações de diferenças lineares de primeira ordem

Sejam $a(t)$ e $g(t)$ funções reais definidas para $t \in \mathbb{N}$, $t \geq t_0 \geq 0$. Um problema de valor inicial para uma equação linear não-homogênea de primeira ordem é dado por

$$x(t+1) = a(t)x(t) + g(t), \quad x(t_0) = x_0, \quad t \geq t_0, \quad (1)$$

em que $x_0 \in \mathbb{R}$.

Quando $g(t) = 0$, a equação presente em (1) é dita homogênea. Portanto, podemos dizer que um típico problema de valor inicial para uma equação linear homogênea de primeira ordem tem a forma

$$y(t+1) = a(t)y(t), \quad y(t_0) = y_0, \quad t \geq t_0, \quad (2)$$

com $y_0 \in \mathbb{R}$.

Em (1) e (2), assumimos que $a(t) \neq 0$ para todo $t \in \mathbb{N}$, $t \geq t_0 \geq 0$.

A solução de (2), dada a condição inicial $y(t_0) = y_0$, pode ser obtida através de iterações:

$$y(t_0+1) = a(t_0)y(t_0) = a(t_0)y_0;$$

$$y(t_0+2) = a(t_0+1)y(t_0+1) = a(t_0+1)a(t_0)y_0;$$

$$y(t_0+3) = a(t_0+2)y(t_0+2) = a(t_0+2)a(t_0+1)a(t_0)y_0.$$

Pelo Princípio da Indução Finita, é possível constatar que

$$y(t) = a(t-1)a(t-2)\dots a(t_0)y_0,$$

ou seja,

$$y(t) = \left[\prod_{i=t_0}^{t-1} a(i) \right] y_0, \quad t \geq t_0. \quad (3)$$

A única solução para o problema (1), com a condição inicial $x(t_0) = x_0$, pode ser encontrada de forma similar, através de iterações:

$$x(t_0+1) = a(t_0)x_0 + g(t_0),$$

$$x(t_0+2) = a(t_0+1)x(t_0+1) + g(t_0+1) = a(t_0+1)a(t_0)x_0 + a(t_0+1)g(t_0) + g(t_0+1).$$

Novamente pelo Princípio da Indução Finita pode-se concluir que

$$x(t) = \left[\prod_{i=t_0}^{t-1} a(i) \right] x_0 + \sum_{r=t_0}^{t-1} \left[\prod_{i=r+1}^{t-1} a(i) \right] g(r), \quad t \geq t_0. \quad (4)$$

2.1.1 Casos Especiais

Existem dois casos de equações de diferenças lineares de suma importância para algumas aplicações. O primeiro deles supõe, para o problema (1), que $a(t)$ seja uma função constante real, $a(t) = a \in \mathbb{R}$, e $t_0 = 0$. Portanto, em lugar de (1), temos o seguinte problema:

$$x(t+1) = ax(t) + g(t), \quad x(0) = x_0, \quad t \geq 0. \quad (5)$$

Consequentemente, por (4), a solução $x(t)$ de (5) é dada por:

$$x(t) = a^t x_0 + \sum_{k=0}^{t-1} a^{t-k-1} g(k), \quad t \geq 0. \quad (6)$$

O segundo caso especial supõe, para o problema (1), que $a(t)$ e $g(t)$ sejam funções constantes e reais, $a(t) = \alpha \in \mathbb{R}$ e $g(t) = \beta \in \mathbb{R}$, e $t_0 = 0$. Sendo assim, em lugar de (1), temos o seguinte problema:

$$x(t+1) = \alpha x(t) + \beta, \quad x(0) = x_0, \quad t \geq 0. \quad (7)$$

Utilizando a fórmula (6), obtém-se a seguinte solução para o problema (7)

$$x(t) = \begin{cases} \alpha^t x_0 + \beta \left[\frac{\alpha^t - 1}{\alpha - 1} \right], & \text{se } \alpha \neq 1, \\ x_0 + \beta t, & \text{se } \alpha = 1. \end{cases} \quad (8)$$

2.2 Equações de diferenças lineares de segunda ordem

Uma equação de diferenças autônoma de segunda ordem tem a seguinte forma geral

$$x(t+2) = f(x(t+1), x(t)), \quad t \in \mathbb{N}, t \geq 0 \quad (9)$$

em que $f : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ é uma função dada.

Para determinar uma única solução para (9), é necessário estabelecer dois valores iniciais, $x(0) = x_0$ e $x(1) = x_1$. Essa única solução pode ser obtida através de recursos computacionais, fazendo $x(2) = f(x_1, x_0)$, $x(3) = f(x(2), x(1)) = f(f(x_1, x_0), x_1)$, e assim sucessivamente, em todo o domínio da f .

Um equação de diferenças de segunda ordem linear é um caso particular da equação (9) e possui a forma geral

$$x(t+2) = p_1 x(t+1) + p_2 x(t) + \beta, \quad \text{com } p_2 \neq 0. \quad (10)$$

Quando $\beta = 0$, a Equação (10) transforma-se na equação

$$x(t+2) = p_1 x(t+1) + p_2 x(t), \quad (11)$$

denominada equação de diferenças de segunda ordem linear homogênea.

A virtude da equação homogênea é que as soluções para ela obedecem princípio da superposição, conforme se pode observar na proposição abaixo.



Proposição 1. Se $z(t)$ e $y(t)$ são soluções de (11) e A e B são constantes, então a função $x(t) = Az(t) + By(t)$ também é solução de (11).

Demonstração. Se $z(t+2) = p_1z(t+1) + p_2z(t)$ e $y(t+2) = p_1y(t+1) + p_2y(t)$, então

$$\begin{aligned}x(t+2) &= Az(t+2) + By(t+2) = A[p_1z(t+1) + p_2z(t)] + B[p_1y(t+1) + p_2y(t)] \\ &= p_1[Az(t+1) + By(t+1)] + p_2[Az(t) + By(t)] \\ &= p_1x(t+1) + p_2x(t),\end{aligned}$$

de onde segue que $x(t)$ também é solução de (11). \square

Este princípio de superposição é a chave para o método de solução da equação (11). Duas soluções $(z(t))_{t \geq 0}$ e $(y(t))_{t \geq 0}$ de (11) são ditas independentes se $(y(t))_{t \geq 0}$ não é um múltiplo constante de $(z(t))_{t \geq 0}$, ou seja, não existe $\mu \in \mathbb{R}$, $\mu \neq 0$, tal que $y(t) = \mu z(t)$ para todo $t \geq 0$. Se pudermos encontrar duas soluções independentes para a equação (11), $(y(t))_{t \geq 0}$ e $(z(t))_{t \geq 0}$, então toda solução de (11) terá a forma $Az(t) + By(t)$, em que A e B são constantes. Tendo essa solução geral para a equação de diferenças (11), podemos resolvê-la sujeita a quaisquer condições iniciais $x(1) = x_1$ e $x(0) = x_0$. Portanto, a solução de (11) deve assumir a forma $x(t) = Az(t) + By(t)$ e cumprir as condições iniciais

$$Az(0) + By(0) = x_0 \quad \text{e} \quad Az(1) + By(1) = x_1. \quad (12)$$

Encontrando A e B , obtém-se a solução.

Como acabamos de ver, resolver uma equação de diferenças de segunda ordem, linear e homogênea, requer encontrar duas soluções independentes. Isso é feito procurando soluções da forma $z(t) = \lambda^t$, em que λ é um número complexo. Substituindo $x(t)$ por $z(t)$ em (11), obtemos

$$\lambda^{t+2} = p_1\lambda^{t+1} + p_2\lambda^t,$$

e dividindo toda a equação por λ^t , encontramos

$$\lambda^2 - p_1\lambda - p_2 = 0. \quad (13)$$

Chamamos (13) de equação característica da equação (11) e suas raízes λ são denominadas raízes características. Resolvendo essa equação, temos

$$\lambda_1 = \frac{p_1 + \sqrt{p_1^2 + 4p_2}}{2} \quad \text{e} \quad \lambda_2 = \frac{p_1 - \sqrt{p_1^2 + 4p_2}}{2}.$$

Se λ é uma raiz, então λ^t é de fato uma solução para (11). Dispomos de três casos para determinar valores de λ :

- **Caso 1:** Quando $p_1^2 + 4p_2 > 0$, existem duas raízes características reais e distintas para o polinômio (13).
- **Caso 2:** Quando $p_1^2 + 4p_2 = 0$, existem duas raízes características reais e iguais para o polinômio (13).
- **Caso 3:** Quando $p_1^2 + 4p_2 < 0$, as raízes características do polinômio (13) são complexas.

É de interesse deste trabalho apenas o Caso 1, quando existem duas raízes características reais e distintas, λ_1 e λ_2 . Neste caso, $x(t) = A\lambda_1^t + B\lambda_2^t$ é a solução geral de (11).



3 Princípios de genética de populações

O principal objetivo dos modelos de genética de populações é descobrir qual é a probabilidade de um indivíduo que possui o genótipo G ter um descendente com o mesmo genótipo. Para responder essa pergunta é necessária a compreensão de algumas premissas que serão apresentadas nesta seção.

3.1 Frequências genótípicas e alélicas

Quando discorremos a respeito das frequências genótípicas e das frequências alélicas, evidenciamos que ambas podem ser representadas a partir da definição de frequência relativa apresentada por Magalhães e Lima (2010). Portanto, considerando uma população qualquer com os possíveis genótipos $G_1 \dots G_k$, a frequência genotípica $f_{G_1 \dots G_k}$ é definida pela seguinte fórmula

$$f_{G_1 \dots G_k} = \frac{\text{número de indivíduos com genótipo } G_1 \dots G_k}{\text{tamanho da população}}. \quad (14)$$

Quando o tempo for considerado, $f_{G_1 \dots G_k}(t)$ denotará a frequência genotípica no tempo t .

As frequências alélicas são calculadas contando somente alelos e ignorando como eles são organizadas em genótipos. Seja l um locus. Dada uma população, a coleção alélica para o locus l é a coleção de todos os alelos de genótipos de população que ocorrem no locus l . Lembramos que uma população é tratada como uma coleção de sequências de caracteres genotípicos. Considere um alelo particular A que pode ocorrer em l . A frequência f_A de A é definida como sua frequência relativa à coleção alélica correspondente a l da seguinte forma

$$f_A = \frac{\text{número de } A\text{'s na coleção alélica de } l}{\text{tamanho da coleção alélica de } l}, \quad (15)$$

e, de forma análoga à frequência genotípica, denotamos a frequência alélica no tempo t por $f_A(t)$.

Exemplo 2. Considere o estudo de um locus com dois alelos, A e a , para uma população de espécies diploides. Uma coleta de 15 amostras dessa população obteve o seguinte resultado: AA, Aa, AA .

Desta forma, pelas fórmulas (14) e (15) obtemos as seguintes frequências genótípicas e alélicas: $f_{AA} = \frac{4}{15}$, $f_{Aa} = \frac{6}{15} = \frac{2}{5}$, $f_{aa} = \frac{5}{15} = \frac{1}{3}$, $f_A = \frac{14}{30} = \frac{7}{15}$ e $f_a = \frac{16}{30} = \frac{8}{15}$.

Note que para um locus com dois alelos, A e a , os possíveis genótipos são AA , Aa e aa , assim, é fácil constatar que: $f_A + f_a = 1$ e $f_{AA} + f_{Aa} + f_{aa} = 1$. Ademais, existe uma relação entre as frequências genótípicas e alélicas, conforme podemos observar na Proposição 3.

Proposição 3. Se \mathcal{P} é uma população de uma espécie diploide com tamanho N e l um locus nessa população que admite dois alelos, A e a , então

$$f_A = f_{AA} + \frac{f_{Aa}}{2} \quad e \quad f_a = f_{aa} + \frac{f_{Aa}}{2}. \quad (16)$$

Demonstração. Como N é o tamanho da população \mathcal{P} e ela é diploide, segue que o tamanho da coleção de alelos no locus l , onde A ocorre, é $2N$. No entanto, para contabilizarmos a quantidade de A 's no grupo de alelos do locus l , usaremos as frequências genótípicas. Pela definição de f_{AA}



(veja (14)), existem Nf_{AA} genótipos AA na população, e como cada um deles contribui com dois alelos A 's, no total existem $2Nf_{AA}$ desses alelos na população \mathcal{P} . Do mesmo modo, existem Nf_{Aa} genótipos Aa que contribuem com um total de Nf_{Aa} alelos A . Como os indivíduos com genótipos aa não contribuem para a frequência f_A , por (15) segue que

$$f_A = \frac{2Nf_{AA} + Nf_{Aa}}{2N} = f_{AA} + \frac{f_{Aa}}{2}.$$

A frequência do alelo a , f_a , pode ser deduzida de forma similar. □

3.2 Acasalamento aleatório

O acasalamento aleatório é o surgimento de um novo indivíduo pela união de dois gametas escolhidos por experimentação aleatória. Desta forma, cada informação genética é passada de maneira aleatória pelos progenitores.

A forma como o acasalamento aleatório ocorre depende exclusivamente do tipo da espécie. Quando está em análise as espécies monoicas, indivíduos que possuem órgãos sexuais dos dois sexos, os dois gametas virão de um mesmo grupo de acasalamento. Em algumas dessas espécies, existe a possibilidade da autofecundação, e quando isso ocorre, ambos os gametas são provenientes de um mesmo progenitor. A probabilidade da autofecundação ocorrer para uma população de tamanho N é $1/N$. Consequentemente, para uma população muito grande, essa probabilidade é extremamente pequena. Em contrapartida, quando a análise é feita em espécies dioicas, indivíduos segregados por machos e fêmeas, os gametas serão provenientes de dois grupos de acasalamentos, um para cada gênero.

Lema 4. *Sejam S uma população com reprodução por acasalamento aleatório e A um possível alelo dos indivíduos. Se f_A^S é a frequência relativa do alelo A na população S e p_A^S é a probabilidade do gameta possuir esse alelo, então*

$$p_A^S = f_A^S. \quad (17)$$

A equação (17) retrata que, embora as probabilidades possuam definições próprias e diferenciadas, em alguns casos as frequências relativas podem ser utilizadas como probabilidades. Para o caso em análise, essa equação é um resultado da escolha aleatória do parceiro e mostra a naturalidade da reprodução sexual. A demonstração do Lema 4 pode ser contemplada em Ocone (2014). O próximo exemplo, também retirado de Ocone (2014), mostra uma aplicação útil do lema anterior.

Exemplo 5. Considere um acasalamento aleatório ocorrendo em uma população monoica. Consideraremos também, em análise, um locus para dois alelos: A e a . Para os genótipos AA , Aa e aa , sejam p_A^1 e $p_a^1 = 1 - p_A^1$ as probabilidades dos progenitores masculinos passarem os respectivos alelos para a prole no acasalamento aleatório. E sejam p_A^2 e $p_a^2 = 1 - p_A^2$ as probabilidades dos progenitores femininos passarem os respectivos alelos para a prole no acasalamento aleatório. Como a população é monoica, os progenitores são escolhidos de um mesmo grupo de acasalamento, portanto, segue que

$$p_A^1 = p_A^2 = f_A \quad \text{e} \quad p_a^1 = p_a^2 = f_a.$$

Por conseguinte, por (16) e (17), obtém-se

$$P(\text{prole ser } AA) = p_A^1 \cdot p_A^2 = (f_A)^2 = \left(f_{AA} + \frac{f_{Aa}}{2} \right)^2, \quad (18)$$

e

$$P(\text{prole ser } aa) = p_a^1 \cdot p_a^2 = (f_a)^2 = \left(f_{aa} + \frac{f_{Aa}}{2} \right)^2. \quad (19)$$

Para o genótipo Aa deve-se levar em conta a contribuição de ambos os alelos. Portanto, uma prole possuirá o genótipo Aa quando receber o alelo A do pai e o alelo a da mãe, ou quando receber o alelo a do pai e o alelo A da mãe. Assim, tem-se

$$P(\text{prole ser } Aa) = p_A^1 p_a^2 + p_a^1 p_A^2 = 2f_A f_a = 2 \left(f_{AA} + \frac{f_{Aa}}{2} \right) \left(f_{aa} + \frac{f_{Aa}}{2} \right). \quad (20)$$

3.3 População infinita

Nesta subseção apresentamos a premissa da população infinita, que é a suposição de que o limite do tamanho da população tende ao infinito. Para considerarmos este conceito é necessário admitirmos que a população acasala de forma aleatória. Denotamos por $p_{G_1 \dots G_k}$ a probabilidade de que um genótipo seja produzido em um único acasalamento aleatório e por $f_{G_1 \dots G_k}$ a frequência genotípica das proles. Assim, para cada genótipo, é válida a seguinte identidade

$$f_{G_1 \dots G_k} = p_{G_1 \dots G_k}. \quad (21)$$

A identidade (21) amplifica o Lema 4. A explicação por trás da identidade pode ser apreciada em Ocone (2014).

A utilização da premissa de população infinita é útil para uma população tão grande de tal forma que a identidade (21) represente o que ocorre quando a população tende ao infinito, tornando-se uma boa aproximação.

3.4 Gerações não-sobrepostas

Outra noção importante é obtida a partir das interações entre as gerações com o grupo de possíveis pais, isto é, se as proles ao nascerem vão ou não para o grupo de acasalamentos dos seus progenitores. O modelo mais simples para estas interações são as gerações não-sobrepostas.

Considere que os indivíduos de uma geração t acasalam entre si para obter a geração $t + 1$ e então não acasalam mais. Analogamente, os indivíduos da geração $t + 1$ se acasalam entre si, obtêm a geração $t + 2$ e não se acasalam mais. Definimos a premissa de gerações não-sobrepostas quando os acasalamentos de todas as gerações seguem este modelo.

3.5 Mutação, seleção e migração

Existem três premissas que anulam a validade do Lema 4, pois quando elas ocorrem, a definição de acasalamento aleatório é debilitada. As três premissas são: mutação, seleção e migração.



A mutação ocorre quando existe uma variação na sequência do DNA de um único gene de forma que um alelo pode se transformar em outro, anulando as leis da hereditariedade. Essas variações ocorrem por algum erro no processo de reprodução ou por efeitos da radiação (BEIGUELMAN, 2008).

Segundo Hartl e Clark (2007), a seleção natural indica que existem diferentes probabilidades de sobrevivência entre os indivíduos antes de se tornarem possíveis reprodutores. Essas diferenças podem ocorrer por diversos fatores, sejam eles naturais, ambientais, ou até mesmo por influência genotípica.

Por fim, a migração ocorre quando existe uma entrada e uma saída excessiva e desorganizada no grupo de acasalamentos aleatórios. Desta forma, indivíduos podem acasalar ou não antes de entrar ou sair da população, trazendo divergências nos conceitos de acasalamento aleatório (HARTL; CLARK, 2007). Um exemplo clássico de migrações são as cidades universitárias, onde os alunos vão para a cidade, fazem parte do grupo de reprodutores por um período curto, se formam e muitas vezes vão embora.

4 Modelagem

Todos os modelos contemplados neste artigo dispensam a presença de seleção natural e a construção será feita a partir do modelo mais simples até chegarmos aos mais complexos. O primeiro modelo a ser desenvolvido será o modelo básico, a partir das premissas contempladas anteriormente, assim, alterná-las-emos a fim de deduzir os próximos modelos.

4.1 Modelo básico

Para o modelo básico considere as seguintes premissas:

- (P1) Acasalamento aleatório;
- (P2) Gerações não-sobrepostas;
- (P3) População infinita;
- (P4) Espécies monoicas;
- (P5) Sem seleção, mutação ou migração.

No desenvolvimento do modelo básico deduziremos noções lógicas das premissas (P1)-(P5), para tanto, consideremos um locus l com dois alelos, A e a . É válido lembrar que a premissa (P5) permite a validade do acasalamento aleatório e, portanto, é justificável a utilização do Lema 4. Assim, sejam $f_{AA}(t)$, $f_{Aa}(t)$ e $f_{aa}(t)$ as frequências genotípicas da geração t , e $f_A(t)$ e $f_a(t)$ as frequências alélicas da mesma geração, com $t \geq 0$.

As gerações são bem definidas, pois (P2) atesta que elas não são sobrepostas. Dessa forma, indivíduos da geração t só se acasalam com indivíduos da geração t , e assim sucessivamente. Note também que, por (P5), as frequências em cada geração permanecem imutáveis desde o nascimento até o momento em que a prole entra para o grupo de acasalamento.

Desejamos sistematizar um conjunto de equações de diferenças de primeira ordem que representem as frequências genotípicas. Assim sendo, considere $f_{AA}(t+1)$ a frequência do genótipo



AA na geração $t + 1$, para qualquer $t \geq 0$. Como estamos lidando com gerações não-sobrepostas, (P2), a geração $t + 1$ é estabelecida por acasalamentos aleatórios com progenitores da geração t . Por conseguinte, pela identidade (21), proposta por (P3), temos

$$f_{AA}(t + 1) = P(\text{um acasalamento aleatório da geração } t, \text{ produzir AA}).$$

Mas, conforme (18), temos

$$P(\text{um acasalamento aleatório da geração } t, \text{ produzir AA}) = f_A^2(t),$$

e assim, pela Proposição 3, segue que

$$f_{AA}(t + 1) = f_A^2(t) = \left(f_{AA}(t) + \frac{f_{Aa}(t)}{2} \right)^2. \quad (22)$$

Utilizando os mesmos princípios, podemos encontrar as equações de diferenças para os genótipos Aa e aa , as quais são:

$$f_{Aa}(t + 1) = 2f_A(t)f_a(t) = 2 \left(f_{AA}(t) + \frac{f_{Aa}(t)}{2} \right) \left(f_{aa}(t) + \frac{f_{Aa}(t)}{2} \right), \quad (23)$$

$$f_{aa}(t + 1) = f_a^2(t) = \left(f_{aa}(t) + \frac{f_{Aa}(t)}{2} \right)^2. \quad (24)$$

Finalmente, as equações de diferenças (22), (23) e (24), constituem o modelo básico. O grande problema é que por se tratarem de equações de diferenças não-lineares, encontrar soluções para essas equações pode ser um desafio. Porém, essas equações trazem uma particularidade das frequências alélicas que simplificam a observação. De fato, pela Proposição 3, temos

$$f_A(t + 1) = f_{AA}(t + 1) + \frac{f_{Aa}(t + 1)}{2},$$

e aplicando (22) e (23), segue que

$$f_A(t + 1) = f_A^2(t) + f_A(t)f_a(t) = f_A(t)(f_A(t) + f_a(t)),$$

e como $f_A(t) + f_a(t) = 1$, finalmente

$$f_A(t + 1) = f_A(t). \quad (25)$$

Uma identidade análoga é obtida para o alelo a , ou seja, $f_a(t + 1) = f_a(t)$. Veja que o modelo foi reduzido para uma equação simples e trivial onde todas as soluções de (25) são constantes no tempo

$$f_A(t) = f_A(t - 1) = \dots = f_A(0), \quad \text{para todo } t \geq 0.$$

As soluções constantes implicam que as frequências alélicas não mudam de geração para geração, e como as frequências alélicas podem ser vistas em função das frequências genotípicas, conforme mostra a Proposição 3, é fácil encontrar soluções completas para as equações (22), (23) e (24), sabendo que $f_A(t) = f_A(0)$ e $f_a(t) = 1 - f_A(t) = 1 - f_A(0)$. De fato:

$$f_{AA}(t + 1) = f_A^2(t) = f_A^2(0), \quad (26)$$

$$f_{Aa}(t + 1) = 2f_A(t)f_a(t) = 2f_A(0)(1 - f_A(0)), \quad (27)$$

$$f_{aa}(t + 1) = f_a^2(t) = (1 - f_A(0))^2. \quad (28)$$



Godfrey H. Hardy e Wilhem Weinberg foram os primeiros pesquisadores a enxergarem essas soluções e, como forma de homenagem, elas passaram a possuir os seus nomes. Esse resultado permitiu um grande avanço para a biologia, alcançando diversas outras consequências importantes.

Definição 6. As frequências genóticas f_{AA} , f_{Aa} e f_{aa} , com $f_{AA} + f_{Aa} + f_{aa} = 1$, estão no equilíbrio de Hardy-Weinberg se existe $p \in \mathbb{R}$, com $0 \leq p \leq 1$, de tal modo que

$$f_{AA} = p^2, \quad f_{Aa} = 2p(1-p) \quad e \quad f_{aa} = (1-p)^2.$$

Note que as frequências genóticas obtidas em (26), (27) e (28) estão no equilíbrio de Hardy-Weinberg, conforme a Definição 6, para todo $t \geq 1$, com $p = f_A(0) = f_{AA}(0) + \frac{1}{2}f_{Aa}(0)$. O equilíbrio é válido com certeza apenas a partir da primeira geração uma vez que o acasalamento aleatório produz cada geração completamente aleatória e redistribui o grupo de alelos dos indivíduos da próxima geração.

Saber se uma população está no equilíbrio de Hardy-Weinberg é muito importante pois essa informação reafirma a premissa (P5), uma vez que populações que sofrem de mutação, seleção ou migração não estarão no equilíbrio.

Proposição 7. As frequências genóticas f_{AA} , f_{Aa} e f_{aa} estão no equilíbrio de Hardy-Weinberg se, e somente se,

$$f_{Aa}^2 = 4f_{AA}f_{aa}.$$

Demonstração. Inicialmente, suporemos que as frequências genóticas f_{AA} , f_{Aa} e f_{aa} estão no equilíbrio de Hardy-Weinberg. Então, pela Definição 6, existe $0 \leq p \leq 1$ tal que

$$f_{AA} = p^2, \quad f_{Aa} = 2p(1-p) \quad e \quad f_{aa} = (1-p)^2. \quad (29)$$

Portanto, de (29) segue que

$$f_{Aa}^2 = [2p(1-p)]^2 = 4p^2(1-p)^2 = 4f_{AA}f_{aa}.$$

Agora, devemos mostrar que as frequências genóticas f_{AA} , f_{Aa} e f_{aa} estão no equilíbrio de Hardy-Weinberg sob a hipótese de que $f_{Aa}^2 = 4f_{AA}f_{aa}$. Com efeito, pela Proposição 3, temos

$$f_A = f_{AA} + \frac{f_{Aa}}{2} \Rightarrow f_{AA} = f_A - \frac{f_{Aa}}{2} \quad e \quad f_a = f_{aa} + \frac{f_{Aa}}{2} \Rightarrow f_{aa} = f_a - \frac{f_{Aa}}{2}.$$

Conseqüentemente, como $f_{Aa}^2 = 4f_{AA}f_{aa}$, concluímos que

$$f_{Aa}^2 = 4f_{AA}f_{aa} = 4 \left(f_A - \frac{f_{Aa}}{2} \right) \left(f_a - \frac{f_{Aa}}{2} \right) = 4f_Af_a - 2f_Af_{Aa} - 2f_a f_{Aa} + f_{Aa}^2.$$

Dessa forma, $2f_{Aa}(f_A + f_a) = 4f_Af_a$ e, como $f_A + f_a = 1$, segue que $f_{Aa} = 2f_Af_a$. Portanto, para que $f_{Aa}^2 = 4f_{AA}f_{aa}$ ocorra, é necessário que

$$f_{AA}f_{aa} = f_A^2 f_a^2,$$

o que vale se $f_{AA} = f_A^2$ e $f_{aa} = f_a^2 = (1-f_A)^2$. Destarte, para as frequências genóticas f_{AA} , f_{Aa} e f_{aa} , com $f_{AA} + f_{Aa} + f_{aa} = 1$, existe um $p = f_A$, com $0 \leq p \leq 1$, tal que $f_{AA} = p^2$, $f_{Aa} = 2p(1-p)$ e $f_{aa} = (1-p)^2$, conforme queríamos demonstrar. \square

Para finalizarmos as considerações acerca modelo básico, o próximo exemplo mostrará a relevância do equilíbrio de Hardy-Weinberg e da Proposição 7.

Exemplo 8. Assuma as premissas (P1)-(P5) e tome $f_{AA}(0) = 0,2$, $f_{Aa}(0) = 0,4$, e $f_{aa}(0) = 0,4$. Note que na geração $t = 0$ a população não se encontra no equilíbrio de Hardy-Weinberg, pois $f_{Aa}^2(0) = (0,4)^2 = 0,16$ é diferente de $4f_{AA}f_{aa} = 4(0,2)(0,4) = 0,32$, mostrando mais uma vez a validade do equilíbrio apenas para $t \geq 1$. A frequência do alelo A na geração $t = 0$ é obtida da seguinte maneira:

$$p = f_A(0) = f_{AA}(0) + \frac{f_{Aa}(0)}{2} = 0,2 + \frac{0,4}{2} = 0,4.$$

Então, pela Definição 6, as frequências genótípicas alcançarão o equilíbrio de Hardy-Weinberg, com $p = 0,4$, na geração $t = 1$. Consequentemente, para $t \geq 1$, $f_{AA}(t) = p^2 = (0,4)^2 = 0,16$, $f_{Aa}(t) = 2p(1-p) = 2(0,4)(0,6) = 0,48$ e $f_{aa}(t) = (1-0,4)^2 = 0,36$.

4.2 O modelo básico para múltiplos alelos em um gene

Considere o modelo básico, ainda utilizando um locus para apenas um gene, porém desta vez com um número m de alelos, denotados por A_1, \dots, A_m , com $m \geq 3$. Assim sendo, os possíveis genótipos possuirão a forma de A_iA_j , para $i, j \in [1, m] \cap \mathbb{N}$. Para simplificarmos a notação, denotaremos por f_{ij} as frequências genótípicas de $f_{A_iA_j}$, e por f_i e f_j as frequências alélicas de f_{A_i} e f_{A_j} , respectivamente.

A construção deste modelo é muito similar a do modelo básico para dois alelos, podemos até considerá-la como sendo uma generalização, portanto será feita de forma resumida. O primeiro passo para essa construção é generalizar a Propriedade 3 para os múltiplos alelos. Para $i \neq j$, temos

$$f_i = \frac{2Nf_{ii} + (Nf_{i1} + Nf_{i2} + Nf_{i3} + \dots + Nf_{ij})}{2N} = f_{ii} + \sum_{j=1}^m \frac{f_{ij}}{2}. \quad (30)$$

Pela premissa dos acasalamentos aleatórios, (P1), e pela Equação (30), seguem abaixo as equações de diferenças que representam o modelo básico para múltiplos alelos, para quaisquer $i, j \in [1, m] \cap \mathbb{N}$, $i \neq j$:

$$f_{ii}(t+1) = f_i^2(t) = \left(f_{ii}(t) + \sum_{j=1}^m \frac{f_{ij}(t)}{2} \right)^2, \quad (31)$$

$$f_{ij}(t+1) = 2f_i(t)f_j(t) = 2 \left(f_{ii}(t) + \sum_{j=1}^m \frac{f_{ij}(t)}{2} \right) \left(f_{jj}(t) + \sum_{i=1}^m \frac{f_{ij}(t)}{2} \right), \quad (32)$$

$$f_{jj}(t+1) = f_j^2(t) = \left(f_{jj}(t) + \sum_{i=1}^m \frac{f_{ij}(t)}{2} \right)^2. \quad (33)$$

Para encontrar as soluções dessas equações de diferenças, da mesma forma que para (22), (23) e (24), é necessário analisar as frequências alélicas. Assim como foi obtido em (25), temos $f_i(t+1) = f_i(t)$, para quaisquer $i, j \in [1, m] \cap \mathbb{N}$, $i \neq j$, de fato:

$$\begin{aligned}
 f_i(t+1) &= f_{ii}(t+1) + \sum_{j=1}^m \frac{f_{ij}(t+1)}{2} \\
 &= f_i^2(t) + \sum_{j=1}^m f_i(t)f_j(t) \\
 &= f_i^2(t) + f_i(t) \sum_{j=1}^m f_j(t) \\
 &= f_i(t) \left[f_i(t) + \sum_{j=1}^m f_j(t) \right] \\
 &= f_i(t).
 \end{aligned}$$

Dessa forma, as frequências alélicas, quando considerados múltiplos alelos, também são constantes no decorrer do tempo, ou seja, $f_i(t) = f_i(0)$ e $f_j(t) = 1 - \sum_{i=1}^m f_i(0)$. Logo, fica fácil encontrar as soluções de (31), (32) e (33), para quaisquer $i, j \in [1, m] \cap \mathbb{N}$, $i \neq j$, isto é,

$$\begin{aligned}
 f_{ii}(t+1) &= f_i^2(t) = f_i^2(0), \\
 f_{ij}(t+1) &= 2f_i(t)f_j(t) = 2f_i(0) \left(1 - \sum_{i=1}^m f_i(0) \right), \\
 f_{jj}(t+1) &= f_j^2(t) = \left(1 - \sum_{i=1}^m f_i(0) \right)^2.
 \end{aligned}$$

Generalizando a Definição 6, dizemos que as frequências $f_{A_i A_j}$, com $i, j \in [1, m] \cap \mathbb{N}$, estão no equilíbrio de Hardy-Weinberg quando existe $p_i \in \mathbb{R}$, com $0 \leq p_i \leq 1$, de forma que

$$f_{ii} = p_i^2, \quad f_{ij} = 2p_i \left(1 - \sum_{i=1}^m p_i \right) \quad \text{e} \quad f_{jj} = \left(1 - \sum_{i=1}^m p_i \right)^2,$$

para $i \neq j$.

Consequentemente, as frequências $f_{A_i A_j}$, com $i, j \in [1, m] \cap \mathbb{N}$, estão no equilíbrio de Hardy-Weinberg quando $t \geq 1$, com $p_i = f_i(0)$. A seguinte proposição, que é uma generalização da Proposição 7, também é válida para os múltiplos alelos e sua demonstração é análoga.

Proposição 9. *As frequências f_{ij} , $i, j \in [1, m] \cap \mathbb{N}$, estão no equilíbrio de Hardy-Weinberg, se, e somente se, para quaisquer $i \neq j$, tem-se*

$$f_{ij}^2 = 4f_{ii}f_{jj}.$$

Podemos, então, constatar que quando ampliamos o número de alelos no gene de um único locus, encontramos uma generalização direta para todos os resultados do modelo básico.

4.3 O modelo básico para uma população dioica

O próximo modelo a ser estudado preserva as premissas (P1), (P2), (P3) e (P5), porém, ao invés de considerarmos espécies monoicas, (P4), apresentaremos uma nova premissa, (P6), as espécies dioicas.

(P6) Espécies dioicas.

Neste estudo, para contemplarmos as espécies dioicas, é necessário o desenvolvimento de dois modelos, um que considera cromossomos autossômicos e outro que considera cromossomos sexuais, pois em cada situação a forma como o sexo e os genótipos são herdados diferem. No primeiro caso, nos cromossomos autossômicos, a herança genética é passada independentemente do sexo da prole, e esse princípio é a base para a construção do primeiro modelo.

Aqui, $G_1 \dots G_k$ denotará qualquer genótipo referente aos loci nos cromossomos autossômicos e $p_{G_1 \dots G_k}^m(t+1)$ denotará a probabilidade que uma prole masculina de um acasalamento aleatório com pais da geração t tem de herdar os genótipos $G_1 \dots G_k$. Analogamente, $p_{G_1 \dots G_k}^f(t+1)$ denotará a probabilidade que uma prole feminina de um acasalamento aleatório com pais da geração t tem de herdar os genótipos $G_1 \dots G_k$. Então, por (P1), podemos dizer que

$$p_{G_1 \dots G_k}^m(t+1) = p_{G_1 \dots G_k}^f(t+1). \quad (34)$$

A Equação (34) é fundamental para a compreensão do caso dioico, pois diz que a probabilidade de uma prole masculina receber um genótipo do progenitor é exatamente a mesma probabilidade de uma prole feminina, ou seja, não existem preferências no gênero do indivíduo. Agora, considere um locus num cromossomo autossômico com o alelo A . A premissa da população infinita, aplicada separadamente para machos e fêmeas, para todo $t \geq 0$, sugere que

$$f_{AA}^m(t+1) = p_{AA}^m(t+1) \quad \text{e} \quad f_{AA}^f(t+1) = p_{AA}^f(t+1), \quad (35)$$

em que $p_{AA}^m(t+1)$ e $p_{AA}^f(t+1)$ foram definidas no parágrafo anterior.

Unindo as considerações feitas pelas equações (34) e (35) é fácil notar uma relação de igualdade entre as frequências do genótipo AA de cada gênero, a saber, $f_{AA}^m(t+1) = f_{AA}^f(t+1)$, para todo $t \geq 0$. As mesmas conclusões podem ser extraídas para quaisquer alelos de loci de cromossomos autossômicos. Então, as frequências genotípicas e alélicas masculinas são iguais às correspondentes frequências genotípicas e alélicas femininas em cada geração t , a partir da primeira geração.

Com essas informações, baseando-nos nas premissas assumidas, podemos encontrar equações de diferenças de primeira ordem que relacionam as frequências genotípicas entre as gerações, conforme seguem abaixo, para todo $t \geq 0$,

- $f_{AA}(t+1) = f_A^m(t)f_A^f(t) = \left(f_{AA}^m(t) + \frac{f_{Aa}^m(t)}{2} \right) \left(f_{AA}^f(t) + \frac{f_{Aa}^f(t)}{2} \right),$
- $f_{Aa}(t+1) = f_A^m(t)f_a^f(t) + f_a^m(t)f_A^f(t)$
 $= \left(f_{AA}^m(t) + \frac{f_{Aa}^m(t)}{2} \right) \left(f_{aa}^f(t) + \frac{f_{Aa}^f(t)}{2} \right) + \left(f_{aa}^m(t) + \frac{f_{Aa}^m(t)}{2} \right) \left(f_{AA}^f(t) + \frac{f_{Aa}^f(t)}{2} \right),$
- $f_{aa}(t+1) = f_a^m(t)f_a^f(t) = \left(f_{aa}^m(t) + \frac{f_{Aa}^m(t)}{2} \right) \left(f_{aa}^f(t) + \frac{f_{Aa}^f(t)}{2} \right).$

Novamente, para encontrarmos as soluções dessas equações de diferenças, é necessário obtermos relações entre as frequências alélicas de cada geração. Assim como no modelo básico, as frequências alélicas de ambos os gêneros serão constantes no tempo a partir da geração t , porém desta vez com $t \geq 1$, devido a (34) e (35). Logo,

$$f_A^m(t+1) = f_A^m(t) \quad \text{e} \quad f_A^f(t+1) = f_A^f(t), \quad \text{para todo } t \geq 1.$$

Portanto,

$$f_{AA}(t+1) = f_A^m(t)f_A^f(t) = f_A^m(1)f_A^f(1) = \left(f_{AA}^m(1) + \frac{f_{Aa}^m(1)}{2} \right) \left(f_{AA}^f(1) + \frac{f_{Aa}^f(1)}{2} \right),$$

e como as frequências genotípicas masculinas e femininas são iguais quando $t \geq 1$, segue que

$$f_{AA}(t+1) = \left(f_{AA}(1) + \frac{f_{Aa}(1)}{2} \right)^2 = \left(f_A^m(0)f_A^f(0) + \frac{f_A^m(0)f_a^f(0)}{2} + \frac{f_a^m(0)f_A^f(0)}{2} \right)^2.$$

Por fim, simplificando a equação acima, por meio da Proposição 3 e manipulações algébricas, obtemos:

$$f_{AA}(t+1) = \left[\frac{1}{2} \left(f_{AA}^m(0) + \frac{f_{Aa}^m(0)}{2} + f_{AA}^f(0) + \frac{f_{Aa}^f(0)}{2} \right) \right]^2.$$

Assim sendo, é fácil verificar que o modelo estará no equilíbrio de Hardy-Weinberg a partir da segunda geração, ou seja, para $t \geq 2$, com

$$p = \frac{1}{2} \left[f_{AA}^m(0) + \frac{f_{Aa}^m(0)}{2} + f_{AA}^f(0) + \frac{f_{Aa}^f(0)}{2} \right] \in [0, 1].$$

O modelo básico aplicado para os cromossomos sexuais é bem diferente do que acabamos de construir para os cromossomos autossômicos. A principal disparidade vem do fato de que os cromossomos sexuais diferem nos gêneros. No caso humano, esses cromossomos são distribuídos em XX para mulheres e XY para os homens, logo, o sexo e a herança genotípica são dependentes. Para entendermos melhor essa questão, considere um locus l carregando os alelos A e a no cromossomo X de uma população humana. A prole feminina possui dois cromossomos X , um do pai e um da mãe, e conseqüentemente recebe um alelo de cada progenitor. Já a prole masculina recebe um cromossomo Y do pai e um cromossomo X da mãe, assim sendo, recebe apenas um alelo do locus l proveniente da mãe.

O fato de que o genoma masculino carrega apenas um cromossomo X implica que ele possui apenas um alelo no locus l e o genótipo masculino possui apenas uma letra: A ou a . Dessa forma, as frequências f_{AA}^m , f_{Aa}^m e f_{aa}^m não existem, abrindo espaço apenas para as frequências f_A^m e f_a^m . De forma dissemelhante, as mulheres possuem as frequências genotípicas usuais, bem como suas frequências alélicas. Assim, para as frequências genotípicas masculinas, valem as relações

$$f_A^m(t+1) = f_A^f(t) \quad \text{e} \quad f_a^m(t+1) = f_a^f(t), \quad t \geq 0. \quad (36)$$

E, para as frequências genóticas femininas, valem

$$f_{AA}^f(t+1) = f_A^m(t)f_A^f(t), \quad t \geq 0, \quad (37)$$

$$f_{Aa}^f(t+1) = f_A^m(t)f_A^f(t) + f_a^m(t)f_a^f(t), \quad t \geq 0, \quad (38)$$

$$f_{aa}^f(t+1) = f_a^m(t)f_a^f(t), \quad t \geq 0. \quad (39)$$

As equações de diferenças (36), (37), (38) e (39) compõem o modelo matemático para os cromossomos sexuais de populações dioicas. Vamos, então, analisar comportamento das frequências alélicas entre as gerações. Pela Proposição 3, segue que

$$\begin{aligned} f_A^f(t+1) &= f_{AA}^f(t+1) + \frac{f_{Aa}^f(t+1)}{2} \\ &= f_A^m(t)f_A^f(t) + \frac{f_A^m(t)f_a^f(t) + f_a^m(t)f_A^f(t)}{2} \\ &= \frac{f_A^m(t)(f_A^f(t) + f_a^f(t)) + f_A^f(t)(f_A^m(t) + f_a^m(t))}{2} \\ &= \frac{f_A^m(t) + f_A^f(t)}{2}. \end{aligned}$$

Por (36), podemos concluir que

$$f_A^f(t+1) = \frac{1}{2}f_A^f(t-1) + \frac{1}{2}f_A^f(t), \quad \text{para todo } t \geq 1. \quad (40)$$

A equação (40) é uma equação de diferenças de segunda ordem linear e, para resolvê-la, usaremos as considerações feitas na Seção 2.2. Veja que a equação característica de (40) é $\lambda^2 - (1/2)\lambda - (1/2) = 0$, cujas raízes são 1 e $-1/2$. Portanto, a solução geral de (40) é dada por $f_A^f(t) = A.(1)^t + B.(-1/2)^t$. Sabendo que a solução deve cumprir as condições iniciais $f_A^f(0)$ e $f_A^f(1) = (1/2)(f_A^m(0) + f_A^f(0))$, obtemos

$$A = (1/3) \left(2f_A^f(0) + f_A^m(0) \right) \quad \text{e} \quad B = (1/3) \left(f_A^f(0) - f_A^m(0) \right).$$

Por conseguinte, a solução de (40) em termos de $f_A^f(0)$ e $f_A^m(0)$ é

$$f_A^f(t) = \left(\frac{2f_A^f(0) + f_A^m(0)}{3} \right) (1)^t + \left(\frac{f_A^f(0) - f_A^m(0)}{3} \right) (-1/2)^t. \quad (41)$$

Através da Proposição 7, não é difícil comprovarmos que o modelo não estará no equilíbrio de Hardy-Weinberg em um número finito de gerações, porém, veremos que quando o tempo tende ao infinito, as frequências tenderão ao equilíbrio de Hardy-Weinberg. Com efeito, denotemos por $f_A^f(\infty)$ e $f_a^f(\infty) = 1 - f_A^f(\infty)$ e $f_{AA}^f(\infty)$, $f_{Aa}^f(\infty)$ e $f_{aa}^f(\infty)$ as frequências alélicas e as frequências

genotípicas quando $t \rightarrow \infty$, respectivamente. Então,

$$\begin{aligned} f_A^f(\infty) &= \lim_{t \rightarrow \infty} f_A^f(t) \\ &= \lim_{t \rightarrow \infty} \left(\frac{2f_A^f(0) + f_A^m(0)}{3} \right) (1)^t + \lim_{t \rightarrow \infty} \left(\frac{f_A^f(0) - f_A^m(0)}{3} \right) (-1/2)^t \\ &= \left(\frac{2f_A^f(0) + f_A^m(0)}{3} \right) \lim_{t \rightarrow \infty} (1)^t + \left(\frac{f_A^f(0) - f_A^m(0)}{3} \right) \lim_{t \rightarrow \infty} (-1/2)^t. \end{aligned}$$

Como $\lim_{t \rightarrow \infty} (1)^t = 1$ e $\lim_{t \rightarrow \infty} (-1/2)^t = 0$, obtemos

$$f_A^f(\infty) = \frac{2f_A^f(0) + f_A^m(0)}{3}. \quad (42)$$

Além disso,

$$\begin{aligned} f_{AA}^f(\infty) &= \lim_{t \rightarrow \infty} f_{AA}^f(t) \\ &= \lim_{t \rightarrow \infty} \left(f_A^m(t-1) f_A^f(t-1) \right) \\ &= \lim_{t \rightarrow \infty} \left(f_A^f(t-2) f_A^f(t-1) \right) \\ &= f_A^f(\infty) f_A^f(\infty), \end{aligned}$$

e por (42),

$$f_{AA}^f(\infty) = \left(\frac{2f_A^f(0) + f_A^m(0)}{3} \right)^2. \quad (43)$$

De forma análoga podemos obter

$$f_{Aa}^f(\infty) = 2 \left(\frac{2f_A^f(0) + f_A^m(0)}{3} \right) \left(1 - \frac{2f_A^f(0) + f_A^m(0)}{3} \right) \quad (44)$$

e

$$f_{aa}^f(\infty) = \left(1 - \frac{2f_A^f(0) + f_A^m(0)}{3} \right)^2. \quad (45)$$

Portanto, as frequências $f_{AA}^f(\infty)$, $f_{Aa}^f(\infty)$ e $f_{aa}^f(\infty)$ estão no equilíbrio de Hardy-Weinberg quando t tende ao infinito, com

$$p = \frac{2f_A^f(0) + f_A^m(0)}{3} \in [0, 1].$$

4.4 Populações com mutação

Nesta seção, estudamos novamente um único locus admitindo dois alelos A e a . Vamos impor novamente as premissas (P1)-(P4) e também excluir migração e seleção. No entanto, queremos permitir a mutação, que ocorre quando um alelo em um gameta parental se modifica no curso do acasalamento. As mutações, que podem ser induzidas pela cópia de erros ou exposição a toxinas químicas ou radiação, não podem ser previstas e por isso são consideradas aleatórias. Desta forma, um modelo simples, para um locus com dois alelos, supõe que o alelo A pode se transformara no alelo a , e vice-versa, de acordo com a premissa abaixo.

(P7) Em cada acasalamento aleatório e para cada progenitor independente, um alelo A sendo transmitido para uma prole se transforma no alelo a com probabilidade igual a u . De forma similar, um alelo a sendo transmitido para uma prole se transforma no alelo A com probabilidade igual a v . Além disso, $u + v > 0$.

A motivação para este modelo não é realmente científica. Porém, queremos explorar como o modelo básico pode mudar a partir da presença de mutação, e a premissa (P7) é o mecanismo de mutação mais simples que se pode imaginar. A suposição técnica $0 < u + v$ exclui o caso em que $u = v = 0$ e nenhuma mutação ocorre.

Conforme observamos nos modelos anteriores, a análise do modelo depende das frequências alélicas, e para obtermos essas frequências, precisamos recordar alguns conceitos. Pela premissa da população infinita, a frequência alélica na geração $t + 1$ é igual à probabilidade de uma prole adquirir o alelo A de um progenitor da geração t . Além disso, por (P7), sabemos que uma prole adquire o alelo A ou se o pai contribuir com um gameta com o alelo A e ele não sofrer mutação, ou se o pai contribuir com um gameta com o alelo a e ele sofrer mutação. A probabilidade do primeiro evento é $(1 - u)f_A(t)$, onde $(1 - u)$ representa a probabilidade de A não sofrer mutação. De forma similar, a probabilidade do segundo evento é $v(1 - f_A(t))$. A partir destas considerações, temos

$$f_A(t + 1) = (1 - u)f_A(t) + v(1 - f_A(t)) = v + (1 - u - v)f_A(t). \quad (46)$$

A equação (46) é uma equação de diferenças de primeira ordem linear da forma (7) e sua solução pode ser obtida de acordo com a fórmula (8). Com efeito,

$$\begin{aligned} f_A(t) &= (1 - u - v)^t f_A(0) + v \left[\frac{(1 - u - v)^t - 1}{(1 - u - v) - 1} \right] \\ &= (1 - u - v)^t f_A(0) + \left[\frac{v - v(1 - u - v)^t}{u + v} \right], \end{aligned}$$

de onde segue que

$$f_A(t) = (1 - u - v)^t \left[f_A(0) - \frac{v}{u + v} \right] + \frac{v}{u + v}. \quad (47)$$

Assim como nos outros modelos, o próximo passo seria analisar o equilíbrio de Hardy-Weinberg, porém, como estamos trabalhando com uma população que admite mutação, é imediato dizer que as frequências não estão em equilíbrio.



5 Conclusão

Como as pesquisas sobre genética de populações e suas múltiplas aplicações têm crescido bastante nos últimos anos, a interdisciplinaridade entre as áreas de Matemática e Biologia torna-se indispensável. Este trabalho, através dos modelos propostos, contribui de forma a salientar a relevância dessa relação.

Nos modelos obtidos, foi possível constatar que, de fato, populações que possuem mutação, seleção ou migração não estão no equilíbrio de Hardy-Weinberg. No modelo básico podemos misturar todo o grupo genético na primeira temporada de acasalamentos que o equilíbrio de Hardy-Weinberg será atingido. Vemos também que este modelo é o ponto de partida para a construção dos outros modelos sem seleção.

É possível perceber que no modelo com populações dioicas, para produzir uma nova geração é necessário que os grupos de alelos de gametas masculinos e femininos sejam misturados, visto que possuem a mesma frequência a partir da primeira geração. Já no último modelo vemos como a mutação afeta as frequências alélicas e conseqüentemente, as frequências genotípicas de uma população. Ressaltamos que o modelo trata-se de apenas uma remodelagem do modelo básico e, mesmo que não reflita o comportamento real de populações, é um ótimo estudo para a compreensão dos elementos de mutação.

Para finalizar, observamos que o presente artigo, embora possua análises de modelos já abordados na literatura, é um grande ponto de partida para a modelagem de genética de populações, uma vez que não existem referências sobre o assunto no nosso idioma.

Referências

- [1] AMABIS, J. M; MARTHO, G. R. **Biologia das populações**: genética, evolução e ecologia. São Paulo: Moderna, 1994.
- [2] BEIGUELMAN, B. **Genética de populações humanas**. Ribeirão Preto: SBG, 2008.
- [3] ELAYDI, S. **An introduction to difference equations**. 3. ed. New York: Springer, 2005.
- [4] HARTL, D. L.; CLARK, A. G. **Principles of population genetics**. 4. ed. Sunderland: Sinauer Associates, 2007.
- [5] LEWIS, R. **Genética humana**: conceitos e aplicações. 5. ed. Rio de Janeiro: Guanabara Koogan, 2004.
- [6] MAGALHÃES, M. N.; LIMA, A. C. P. **Noções de probabilidade e estatística**. 7. ed. São Paulo: EDUSP, 2010.
- [7] OCONE, D. **Discrete and probabilistic models in biology**. Apostila do Curso de Modelos Discretos e Probabilísticos na Biologia da Universidade de Rutgers. New Brunswick, 2014.