



Notas sobre Modelagem Matemática na Genética de Populações com Seleção

Vinícius Freitas de Oliveira e Suzete Maria Silva Afonso

RESUMO: A genética de populações é a ciência que estuda o comportamento das frequências alélicas e genotípicas nas populações e quais os possíveis fenômenos que podem alterá-las ao longo do tempo. A fim de estabelecer uma interdisciplinaridade entre biologia e matemática, estas notas exibem a construção de um modelo matemático para populações com a presença de seleção natural de forma que contemple o comportamento de tais frequências no tempo. Para tanto, são utilizados conceitos de probabilidade, estatística e equações de diferença. O modelo final trata de uma equação de diferenças não linear e o estudo dos resultados conta com o auxílio do método teia de aranha para a análise qualitativa dos resultados. Estas notas foram inspiradas nas notas de aula de Ocone (2014), [7].

Sumário

1	Introdução	1
2	Princípios de Genética de Populações	3
2.1	Frequências Alélicas e Frequências Genotípicas	3
2.2	Acasalamento Aleatório	5
2.3	Mutação, Seleção e Migração	6
2.4	Gerações Não-sobrepostas	6
2.5	Premissa da População Infinita	6
2.6	Equilíbrio de Hardy-Weinberg	7
3	Construção do Modelo	7
4	Análise do Modelo	10
5	Considerações Finais	14

1. Introdução

No final do século XIX, com a necessidade de estudar variações hereditárias nas populações, surgiu a ciência da genética de populações. Essa ciência analisa o comportamento das frequências gênicas nas populações, bem como os fenômenos capazes de alterá-las no decorrer do tempo. Esses fenômenos são compreendidos principalmente pela migração, seleção natural e mutação.

As variações gênicas dependem de características específicas dos indivíduos, o genótipo e o fenótipo, e para entendê-las advimos do sistema estrutural de funcionamento dos organismos vivos, a célula. É na célula que está contida toda a informação genética, passada de forma hereditária. Essa herança é chamada de gene e trata-se de um trecho da molécula de ácido desoxirribonucleico, o DNA (AMABIS e MARTHO, 1994).

Os genes são encontrados em um local dos cromossomos, denominado locus, no plural loci. As diferentes formas e variações dos genes são chamadas de alelos, que têm direta influência na determinação de uma característica (LEWIS, 2004). Os alelos são identificados por letras, podendo ser recessivos ou dominantes. Alelos dominantes, representados por letras maiúsculas, sempre se expressam. Já os alelos recessivos, representados por letras minúsculas, são inibidos na presença de alelos dominantes, ou seja, só possuem expressão quando sozinhos.

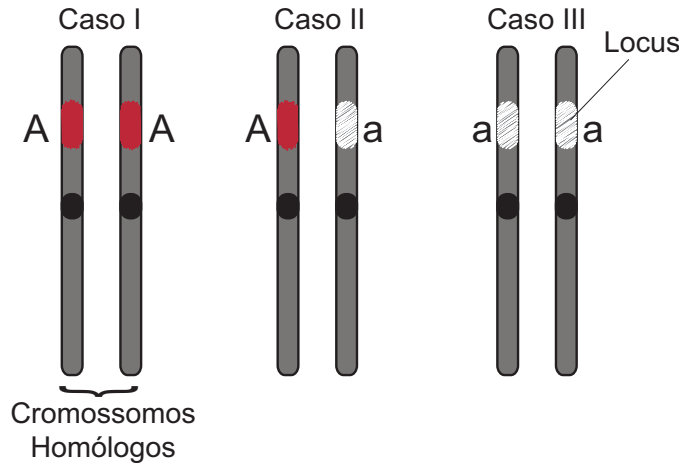
Os cromossomos podem ser subdivididos como autossômicos, que não se diferem entre os sexos, e sexuais, que determinam o sexo do indivíduo. Para as chamadas espécies diploides, os cromossomos autossômicos são agrupados em pares homólogos, que são iguais entre si. Esses cromossomos pareados são chamados de genótipos, que são uma estrutura gráfica dos alelos. São os genótipos que possuem o código para construir uma proteína e determinar uma característica. Essa característica é chamada de fenótipo (LEWIS, 2004).

A diferença entre genótipo e fenótipo é mais complexa, pois além do fenótipo ter dependência direta do genótipo, ele também depende da influência do meio em que se vive (HARTL e CLARK, 2007). Por exemplo, a cor da pele é uma característica, portanto, um fenótipo. O genótipo do indivíduo irá dizer a quantidade da proteína melanina que deverá ser produzida, porém, a exposição ao fator externo, o sol, também terá influência na cor da pele. A Figura 1 mostra cromossomos homólogos com seus diferentes genótipos.

Analisando a Figura 1, pelos casos I e III, observam-se os genótipos **AA** e **aa**, respectivamente, possuindo dois alelos iguais — homozigose; já no segundo caso, observa-se o genótipo **Aa**, com alelos diferentes — heterozigose.

O estudo dos genótipos e de suas frequências genotípicas é muito importante, pois nos permite prever as suas distribuições em proles provenientes de acasalamentos de um grupo de indivíduos da mesma espécie. As aplicações dessas previsões são numerosas e muito úteis na medicina, no direito, na biotecnologia, na sociologia, na antropologia, nas políticas sociais, dentre outras áreas. Por exemplo, a compreensão das distribuições genotípicas no decorrer do tempo pode ajudar na percepção de doenças hereditárias, levando a uma possível cura, ou no mapeamento genético da susceptibilidade de doenças, incluindo o câncer, ou na melhoria no desempenho de animais domésticos e plantas cultivadas, ou ainda na organização de programas de cruzamentos para a conservação de espécies ameaçadas de extinção (HARTL e CLARK, 2007).

Figura 1: Cromossomos Homólogos e Genótipos



FONTE: Os Autores (2018)

Aqui, temos como objetivo realizar aplicações matemáticas na genética de populações por meio da construção e análise de um modelo matemático para populações com seleção. Cabe mencionar que essa construção pode também ser verificada em Ocone (2014), [7]. A modelagem será promovida por meio de conceitos de probabilidade, estatística e equações de diferenças e o modelo será examinado mediante a análise de estabilidade e o método teia de aranha.

Para a elaboração da figura presente no texto foi utilizado o software Adobe Illustrator^R e para a construção dos gráficos o software AutoCAD^R.

2. Princípios de Genética de Populações

O grande propósito dos modelos de genética de populações é descobrir qual é a probabilidade que um indivíduo portador de um genótipo G tenha um descendente com esse mesmo genótipo. Essa probabilidade dependerá de diversas premissas sobre quão provável é que indivíduos de diferentes genótipos acasalem entre si. Os itens abordados nesta seção tratam de princípios por trás dessas premissas.

2.1. FREQUÊNCIAS ALÉLICAS E FREQUÊNCIAS GENOTÍPICAS. Considere uma população qualquer e um dos possíveis genótipos, $G_1 \cdots G_k$. A frequência $f_{G_1 \cdots G_k}$ é definida pela seguinte fórmula:

$$f_{G_1 \cdots G_k} \triangleq \frac{\text{número de indivíduos com genótipo } G_1 \cdots G_k}{\text{tamanho da população}}. \quad (1)$$

Quando o tempo for considerado, $f_{G_1 \cdots G_k}(t)$ 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:

$$f_A \triangleq \frac{\text{número de } A\text{'s na coleção alélica de } l}{\text{tamanho da coleção alélica de } l}, \quad (2)$$

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

Exemplo 2.1. 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: $AA, AA, Aa, AA, aa, aa, aa, Aa, Aa, AA, AA, aa, Aa, Aa, AA$.

Se essas amostras representam toda a população, então as frequências genotípicas valem: $f_{AA} = \frac{6}{15} = \frac{2}{5}$, $f_{Aa} = \frac{5}{15} = \frac{1}{3}$ e $f_{aa} = \frac{4}{15}$, valores obtidos com a fórmula (1). Com a fórmula (2), vemos que as frequências alélicas valem: $f_A = \frac{17}{30}$ e $f_a = \frac{13}{30}$.

Veja que para um locus com dois alelos, A e a , os possíveis genótipos são AA , Aa e aa , assim, é fácil notar que: $f_A + f_a = 1$ e $f_{AA} + f_{Aa} + f_{aa} = 1$. Além disso, existe uma relação entre as frequências genotípicas e alélicas. É o que veremos na próxima proposição.

Proposição 2.2. Seja \mathcal{P} uma população de uma espécie diploide com tamanho N e seja l um locus nessa população que admite dois alelos, A e a . Então,

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

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 $2N$. Contudo, para contarmos a quantidade de A 's no grupo de alelos do locus l , usaremos as frequências genotípicas. Pela definição de f_{AA} (veja (1)), 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 (2) 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 obtida de forma análoga. □

Embora as frequências alélicas possam ser encontradas através das frequências genotípicas, sem pormenorizar, a recíproca não é válida, pois as frequências genotípicas não dependem apenas da quantidade de alelos, mas de como os alelos estão distribuídos entre os indivíduos.

2.2. ACASALAMENTO ALEATÓRIO. De acordo com Magalhães e Lima (2010), uma variável aleatória trata-se de um valor de interesse proveniente de um experimento aleatório e é exatamente esse conceito que arquiteta a definição de acasalamento aleatório.

Definição 2.3. *O acasalamento aleatório é a criação de um novo indivíduo pela união de dois gametas, um óvulo e um espermatozoide, escolhidos por experimentação aleatória na população. Em suma, cada gameta, seja masculino ou feminino, é cedido de maneira aleatória por seus pais.*

No caso de espécies monoicas, espécies que apresentam órgãos sexuais dos dois sexos, ambos gametas virão de um mesmo grupo de acasalamento, havendo possibilidade de reposição, ou seja, os dois gametas podem ser de um mesmo pai ou não. Quando os gametas vem de um mesmo pai, ocorre a chamada autofecundação. Para uma população com tamanho N , a probabilidade de ocorrer uma autofecundação é de $1/N$. Consequentemente, para uma população muito grande, essa probabilidade é extremamente pequena. O lema abaixo estabelece um princípio simples para calcular a probabilidade genotípica da prole através da frequência de ocorrência.

Lema 2.4. *Seja S uma população com reprodução por acasalamento aleatório e A um alelo. Se p_A^S é a probabilidade do gameta possuir o alelo A e f_A^S é a frequência desse alelo na população S , então:*

$$p_A^S = f_A^S \quad (3)$$

O Lema 2.4 traz um dos resultados mais importantes para modelos baseados em acasalamentos aleatórios e sua prova pode ser encontrada em Ocone (2014). A equação (3) é uma consequência da ideia de escolha aleatória do parceiro e da natureza da reprodução sexual.

Para finalizarmos as considerações a respeito do acasalamento aleatório, o exemplo abaixo traz uma aplicação importante do Lema 2.4.

Exemplo 2.5. *Considere um acasalamento aleatório ocorrendo em uma população de genótipos AA , Aa e aa em uma população monoica (com um locus e dois alelos A e a), com frequências alélicas f_A e f_a . Sejam p_A^1 e $p_a^1 = 1 - p_A^1$ as probabilidades dos progenitores masculinos passarem os alelos A e a , respectivamente, para a prole no acasalamento aleatório. E sejam p_A^2 e $p_a^2 = 1 - p_A^2$ as probabilidades dos progenitores femininos transmitirem os alelos A e a , respectivamente, para a prole no acasalamento aleatório. Uma vez que ambos os progenitores são escolhidos do mesmo grupo de acasalamento (estamos considerando uma espécie monoica), segue que:*

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

A probabilidade da prole, de um acasalamento aleatório, possuir genótipo AA é equivalente à probabilidade dela receber o alelo A de cada progenitor (análogo para aa), já que os progenitores são escolhidos de forma independente. Sendo assim, tem-se

$$P(\text{prole ser } AA) = p_A^1 \cdot p_A^2 = (f_A)^2 = \left(f_{AA} + \frac{f_{Aa}}{2} \right)^2 \quad (4)$$

e

$$P(\text{prole ser } aa) = p_a^1 \cdot p_a^2 = (f_a)^2 = \left(f_{aa} + \frac{f_{Aa}}{2} \right)^2. \quad (5)$$

Por fim, para calcular a probabilidade da prole obter um genótipo Aa , leva-se em conta a contribuição de ambos os alelos, A e a . Desta forma, obtém-se

$$P(\text{prole ser } Aa) = p_A^1 p_a^2 + p_a^1 p_A^2 = 2f_A f_a = 2 \left(f_{AA} + \frac{f_{Aa}}{2} \right) \left(f_{aa} + \frac{f_{Aa}}{2} \right). \quad (6)$$

2.3. MUTAÇÃO, SELEÇÃO E MIGRAÇÃO. A mutação é uma mudança aleatória na sequência do DNA de um único gene. Essa variação ocorre em um alelo de um gameta proveniente do grupo de acasalamentos por um erro no processo de reprodução ou por consequências da radiação (BEIGUELMAN, 2008).

A seleção natural, intitulada por Darwin, é um princípio em que indivíduos de uma população estão sujeitos, devido à luta pela vida, a probabilidades de sobrevivência antes de se tornarem possíveis reprodutores, por diversos fatores, incluindo a influência genotípica. Em outras palavras, a seleção natural acontece quando diferentes genótipos têm diferentes chances de sobrevivência ou sucesso reprodutivo (HARTL e CLARK, 2007).

Já a migração, segundo Hartl e Clark (2007), ocorre quando indivíduos entram e saem, em massa e de forma desorganizada, de uma população. As chamadas cidades universitárias dispõem dessa característica, pois muitos jovens chegam na cidade para estudar, fazem parte do grupo de reprodutores por um período de tempo muito curto, se formam e vão embora.

2.4. GERAÇÕES NÃO-SOBREPOSTAS. O conceito de gerações não-sobrepostas exprime que os indivíduos de uma geração qualquer t acasalam entre si para produzir a geração $t + 1$ e a partir disso não acasalam mais. Da mesma forma, a geração $t + 1$ acasala entre si e, analogamente, não acasalam mais, e assim por diante (OCONE, 2014). Esse princípio é muito útil para modelos de populações com acasalamentos sazonais, como é o caso de muitas espécies de pássaros.

2.5. PREMISSE DA POPULAÇÃO INFINITA. O próximo princípio é a suposição de uma população infinita, ou seja, a suposição de que o limite do tamanho da população tende ao infinito. Embora essa premissa seja extremamente simples, os seus resultados e a sua aplicação tem uma enorme relevância.

Considere uma população de filhos construída por acasalamento aleatório de uma determinada população de pais. Seja $f_{G_1 \dots G_k}$ a frequência de um genótipo na população de descendentes e seja $p_{G_1 \dots G_k}$ a probabilidade de que o genótipo é produzido em um único acasalamento aleatório.

A premissa da população infinita consiste em impor que, para todo genótipo, seja válida a seguinte identidade¹:

$$f_{G_1 \dots G_k} = p_{G_1 \dots G_k}. \quad (7)$$

Na realidade prática, a utilização da premissa de população infinita é útil para uma população tão grande de tal forma que a identidade (7) represente o que ocorre quando a população tende ao infinito, tornando-se uma boa aproximação.

2.6. EQUILÍBRIO DE HARDY-WEINBERG. Godfrey Harold Hardy e Wilhem Weinberg chegaram quase simultaneamente a conclusões sobre a genética de populações. Essas conclusões passaram a ser conhecidas como a lei do equilíbrio de Hardy-Weinberg.

Definição 2.6. *As frequências genotípicas f_{AA} , f_{Aa} e f_{aa} , com $f_{AA} + f_{Aa} + f_{aa} = 1$, estão no equilíbrio de Hardy-Weinberg se existir $p \in \mathbb{R}$, com $0 \leq p \leq 1$, de tal modo que:*

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

O equilíbrio de Hardy-Weinberg é muito importante pois sua ausência diz que a população pode sofrer mutação, migração ou seleção, adulterando os resultados das variações genéticas. Daí vem a importância prática de saber se a população está ou não nesse equilíbrio.

3. Construção do Modelo

Para a construção do modelo ², consideremos as seguintes premissas:

1. Acasalamento aleatório;;
2. Gerações não-sobrepostas;
3. População infinita;
4. Espécies monoicas;
5. População sem migração e mutação;
6. População com seleção.

¹ OCONE (2014) traz detalhes técnicos para explicar a validade desta identidade.

² A construção desse modelo também é encontrada em Ocone (2014).

Também consideraremos que os acasalamentos ocorrem sazonalmente, ou seja, a geração t acasala para produzir a geração $t + 1$ (depois não acasala mais por 2), assim, o próximo acasalamento só ocorrerá quando a geração $t + 1$ for madura sexualmente para produzir a geração $t + 2$, e assim por diante.

Para iniciarmos a modelagem, consideremos uma população com um locus e dois alelos, A e a . Definiremos dois tipos de frequências para uma mesma geração. Denotamos as frequências alélicas e genotípicas na geração t , no momento do nascimento, por $f_A(t)$ e $f_a(t)$ e $f_{AA}(t)$, $f_{Aa}(t)$ e $f_{aa}(t)$, respectivamente. Por exemplo, a frequência f_{AA} é a probabilidade de um indivíduo nascido na geração t possuir o genótipo AA .

O segundo grupo de frequências são aquelas que representam a geração t no momento da maturidade sexual, usaremos as notações $p_A(t)$ e $p_a(t)$ para as frequências alélicas e $p_{AA}(t)$, $p_{Aa}(t)$ e $p_{aa}(t)$ para as frequências genotípicas. Observemos que a frequência no momento da maturidade sexual depende da sobrevivência dos indivíduos. Assim, para o genótipo AA , por exemplo, segue que³ $p_{AA}(t) = P(U_{AA}|S)$, onde U_{AA} é o evento do indivíduo nascido na geração t possuir genótipo AA , e S é o evento do indivíduo nascido na geração t sobreviver. A mesma derivação vale para $p_{Aa}(t)$ e $p_{aa}(t)$.

A premissa 6 diz que na população há presença de seleção e sabemos que ela ocorre quando os genótipos afetam a sobrevivência. Isso pode acontecer em duas situações: quando os genótipos possuem diferentes probabilidades de reprodução ou quando os genótipos possuem diferentes taxas de sobrevivência. Essas taxas de sobrevivência, ou coeficientes de seleção, serão denotadas por w_{AA} , w_{Aa} e w_{aa} e são iguais de geração em geração, ou seja, os coeficientes de seleção são independentes de indivíduo para indivíduo e independentes da geração. Para simplificarmos a compreensão, diremos que w_{AA} é a probabilidade de um indivíduo sobreviver visto que possui o genótipo AA , de forma análoga w_{Aa} para o genótipo Aa , e w_{aa} para o genótipo aa .

Nas próximas linhas aplicaremos o Teorema de Bayes, que será enunciado abaixo. Sua demonstração pode ser encontrada em Magalhães e Lima (2010).

Teorema 3.1 (Teorema de Bayes). *Considere que os eventos A_1, A_2, \dots, A_k formem uma partição do espaço amostral Ω e considere B um outro evento tal que $P(A_i)$ e $P(B|A_i)$, $i = 1, 2, \dots, k$, sejam conhecidas. Para qualquer $j = 1, 2, \dots, k$, tem-se*

$$P(A_j|B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(B|A_j) \cdot P(A_j)}{P(B)}$$

³ $P(U_{AA}|S)$ equivale a probabilidade condicional do evento U_{AA} visto que o evento S ocorreu. Magalhães e Lima (2010) trazem uma abordagem mais completa sobre as probabilidades condicionais.

em que

$$P(B) = \sum_{i=1}^k P(A_i) \cdot P(B|A_i).$$

Por conseguinte, pelo Teorema de Bayes, tem-se:

$$p_{AA}(t) = \frac{P(S|U_{AA})P(U_{AA})}{P(S)}.$$

Além disso,

$$\begin{aligned} P(S) &= P(S|U_{AA})P(U_{AA}) + P(S|U_{Aa})P(U_{Aa}) + P(S|U_{aa})P(U_{aa}) = \\ &= w_{AA}f_{AA}(t) + w_{Aa}f_{Aa}(t) + w_{aa}f_{aa}(t), \end{aligned}$$

de onde segue que

$$p_{AA}(t) = \frac{w_{AA}f_{AA}(t)}{w_{AA}f_{AA}(t) + w_{Aa}f_{Aa}(t) + w_{aa}f_{aa}(t)}. \quad (8)$$

O próximo objetivo é simplificar a equação (8) em função das frequências alélicas. Sabemos que a probabilidade de um pai da geração t , escolhido aleatoriamente, passar o alelo A para a prole é $p_A(t)$, e pela Proposição 2.2, segue que $p_A(t) = p_{AA}(t) + (1/2)p_{Aa}$. Consequentemente, pelas premissas 1 e 2, para todo $t \geq 0$, tem-se:

$$f_A(t+1) = p_A(t) \quad \text{e} \quad f_a(t+1) = (1 - p_A(t)), \quad (9)$$

e pelas equações (4), (5) e (6),

$$f_{AA}(t+1) = p_A^2(t), \quad f_{Aa}(t+1) = 2p_A(t)(1 - p_A(t)) \quad \text{e} \quad f_{aa}(t+1) = (1 - p_A(t))^2. \quad (10)$$

Quando $t \geq 1$ as frequências acima estarão no equilíbrio de Hardy-Weinberg com

$$p = p_A(t-1) = f_A(t).$$

Portanto, substituindo (9) e (10) na Equação (8), temos:

$$p_{AA}(t) = \frac{w_{AA}f_A^2(t)}{w_{AA}f_A^2(t) + w_{Aa}2f_A(t)(1 - f_A(t)) + w_{aa}(1 - f_A(t))^2}, \quad (11)$$

similarmente,

$$p_{Aa}(t) = \frac{w_{Aa}2f_A(t)(1 - f_A(t))}{w_{AA}f_A^2(t) + w_{Aa}2f_A(t)(1 - f_A(t)) + w_{aa}(1 - f_A(t))^2}. \quad (12)$$

Para completarmos o nosso objetivo usamos o fato de $f_A(t+1) = p_A(t) = p_{AA}(t) + (1/2)p_{Aa}(t)$ e somamos a Equação (11) com a metade da Equação (12) e finalmente obtemos:

$$f_A(t+1) = \frac{w_{AA}f_A^2(t) + w_{Aa}f_A(t)(1-f_A(t))}{w_{AA}f_A^2(t) + w_{Aa}2f_A(t)(1-f_A(t)) + w_{aa}(1-f_A(t))^2}. \quad (13)$$

Embora (13) seja válida apenas para todo $t \geq 1$, se assumirmos que a geração 0 está no equilíbrio de Hardy-Weinberg em sua infância, ou seja, que $f_{AA}(0) = f_A^2(0)$, $f_{Aa}(0) = 2f_A(0)(1-f_A(0))$ e $f_{aa}(0) = (1-f_A(0))^2$, então a Equação (13) também será válida para $t = 0$, e para prosseguirmos, assumiremos que isso ocorre, o que não afetará a análise qualitativa do limite de $f_A(t)$ quando t tende ao infinito.

Até então poderíamos considerar a Equação (13) como o modelo final, porém é conveniente simplificarmos a escrita com uma função chamada *função fitness*. Biologicamente, essa função avalia o quão adaptado está o indivíduo no ambiente em que vive e a tradução matemática é mostrada a seguir:

$$W(p) = p^2w_{AA} + 2p(1-p)w_{Aa} + (1-p)^2w_{aa}, \quad \text{com } 0 \leq p \leq 1. \quad (14)$$

Portanto, para todo $t \geq 0$, aplicando (14) no modelo (13), obtemos a versão final, conforme segue abaixo.

$$f_A(t+1) = \frac{w_{AA}f_A^2(t) + w_{Aa}f_A(t)(1-f_A(t))}{W(f_A(t))}. \quad (15)$$

No início desta seção dissemos que a seleção ocorre em duas situações, porém o modelo foi construído tendo em vista que a seleção ocorre apenas quando genótipos possuem diferentes probabilidades de sobrevivência. Felizmente o modelo (15) pode ser generalizado para a outra situação, quando genótipos possuem diferentes probabilidades de reprodução. Para isso, é preciso apenas manter a premissa de acasalamentos aleatórios.

4. Análise do Modelo

Para a análise gráfica da equação (15) usaremos o método teia de aranha (*cobweb*), que pode ser analisado pelo leitor em Elaydi (2005).

Inicialmente, observemos que os coeficientes de seleção são estritamente positivos, visto que são probabilidades, conseqüentemente a função fitness também assume valores positivos, $W(p) > 0$. Para simplificarmos a notação do modelo, denotaremos $f_A(t)$ por $f(t)$ e definiremos a função $\phi(p)$ por

$$\phi(p) = \frac{p^2w_{AA} + p(1-p)w_{Aa}}{W(p)}, \quad 0 \leq p \leq 1.$$

Assim fica mais fácil enxergar a equação (15) como uma equação de diferenças autônoma de primeira ordem na seguinte forma $f(t+1) = \phi(f(t))$. O primeiro passo para a aplicação da teia de aranha é calcular os pontos fixos da função ϕ . Com efeito, os pontos $p \in [0, 1]$ que satisfazem $\phi(p) = p$ são:

$$p_1^* = 0, \quad p_2^* = 1 \quad \text{e} \quad p_3^* = \frac{w_{Aa} - w_{aa}}{-w_{AA} + 2w_{Aa} - w_{aa}}.$$

Os dois primeiros pontos fixos possuem uma análise preliminar muito importante. Note que quando $p = 0$ o alelo A entrará em extinção, pois se isso ocorrer em alguma geração, as próximas gerações continuarão em abstinência de A pois não há mutação para mudar esse quadro. De modo semelhante veja que quando $p = 1$ o alelo a entrará em extinção.

Além disso, como p pertence ao intervalo $[0, 1]$, para que p_3^* esteja entre 0 e 1 é necessário que ou $w_{Aa} > w_{aa}$ e $w_{Aa} > w_{AA}$ ou $w_{Aa} < w_{aa}$ e $w_{Aa} < w_{AA}$. De fato, quando o numerador e o denominador de p_3^* são positivos, segue que:

$$0 < \frac{w_{Aa} - w_{aa}}{2w_{Aa} - w_{AA} - w_{aa}} < 1 \Rightarrow \begin{cases} w_{Aa} - w_{aa} > 0 \Rightarrow w_{Aa} > w_{aa} \\ w_{Aa} - w_{aa} < 2w_{Aa} - w_{AA} - w_{aa} \Rightarrow w_{Aa} > w_{AA}, \end{cases}$$

e caso o numerador e o denominador de p_3^* sejam negativos, segue que:

$$0 < \frac{w_{Aa} - w_{aa}}{2w_{Aa} - w_{AA} - w_{aa}} < 1 \Rightarrow \begin{cases} w_{Aa} - w_{aa} < 0 \Rightarrow w_{Aa} < w_{aa} \\ w_{Aa} - w_{aa} > 2w_{Aa} - w_{AA} - w_{aa} \Rightarrow w_{Aa} < w_{AA}. \end{cases}$$

O segundo passo para aplicarmos a teia de aranha é encontrarmos a derivada de ϕ no ponto p . Com efeito,

$$\phi'(p) = \frac{p^2 w_{AA} w_{Aa} + 2p(1-p)w_{AA}w_{aa} + (1-p)^2 w_{Aa}w_{aa}}{W^2(p)}.$$

Agora, dividiremos nosso estudo em quatro casos, que são eles: dominância do alelo a , dominância do alelo A , dominância dos heterozigotos e dominância dos homozigotos.

• Caso I: Dominância do Alelo a

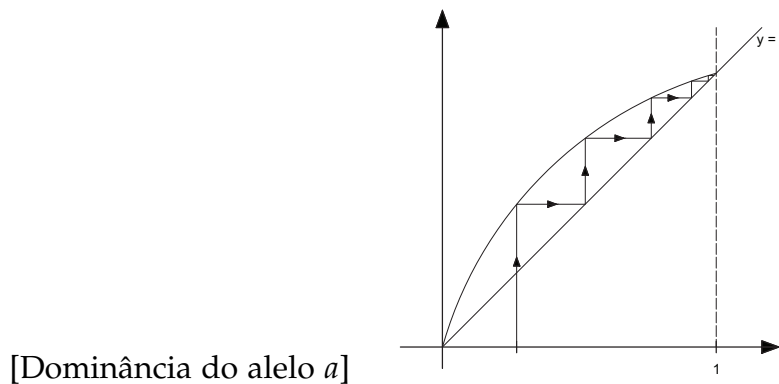
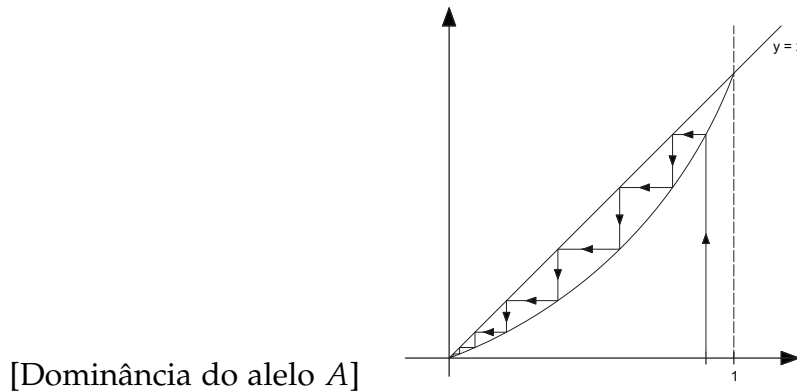
Para o primeiro caso consideremos que $w_{AA} < w_{Aa} < w_{aa}$. Veja que o ponto fixo p_3^* não está entre 0 e 1 e assim não faz parte do estudo, consequentemente teremos apenas dois pontos fixos de ϕ em $0 \leq p \leq 1$ que são $p_1^* = 0$ e $p_2^* = 1$. Podemos concluir que o gráfico da ϕ deve estar inteiramente acima ou inteiramente abaixo da reta diagonal $y = x$ no intervalo $0 < p < 1$.

A inclinação da reta tangente à função $y = \phi(p)$ em p_1^* é dada pela derivada da ϕ no ponto p_1^* , ou seja, $\phi'(0) = w_{Aa}/w_{aa}$. A hipótese inicial,

$w_{Aa} < w_{aa}$, implica que $\phi'(0) < 1$, conseqüentemente o gráfico da ϕ estará inteiramente abaixo da diagonal $y = x$, conforme mostra a Figura 2(a).

Aplicando o método da teia de aranha, conforme mostra a Figura 2(a), concluímos que $\lim_{t \rightarrow \infty} f(t) = 0$, o que significa que o alelo A entrará em extinção e o alelo a dominará o grupo de alelos.

Figura 2: Dominância dos alelos



FONTE: Os Autores (2018)

• **Caso II: Dominância do Alelo A**

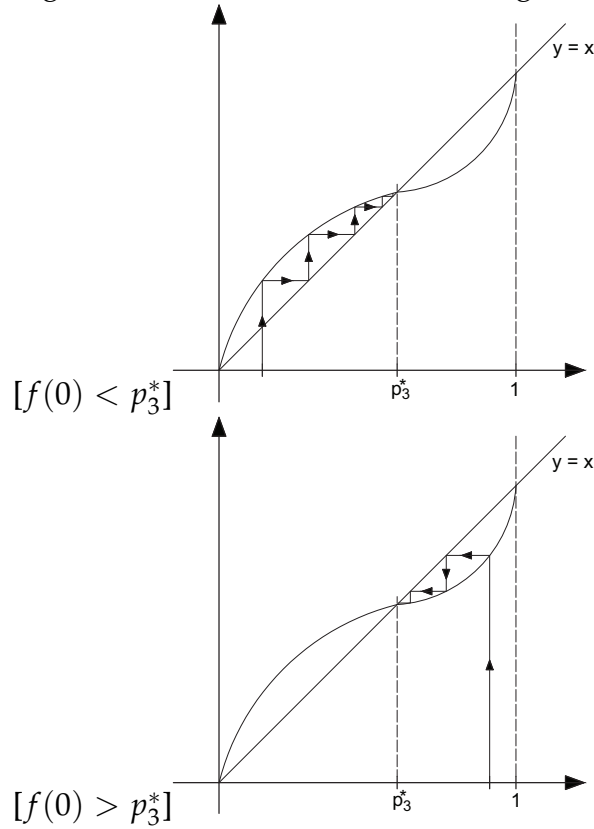
Para este caso consideremos que $w_{aa} < w_{Aa} < w_{AA}$. Novamente, o ponto fixo p_3^* não fará parte do estudo, então os pontos fixos da ϕ são p_1^* e p_2^* . Dessa vez, temos que $\phi'(0) > 1$, em virtude da hipótese. Conseqüentemente, o gráfico da ϕ estará inteiramente acima da diagonal $y = x$, conforme mostra a Figura 2(b).

Aplicando o método da teia de aranha para esse caso, conforme mostra a Figura 2(b), concluímos que $\lim_{t \rightarrow \infty} f(t) = 1$, o que significa que o alelo a entrará em extinção e o alelo A dominará o grupo de alelos.

• **Caso III: Dominância dos Heterozigotos**

Agora, consideremos $w_{Aa} > w_{AA}$ e $w_{Aa} > w_{aa}$. Veja que para esse caso os três pontos fixos da ϕ pertencem ao intervalo $[0, 1]$. Porém, o gráfico da ϕ não estará inteiramente acima ou inteiramente abaixo da diagonal $y = x$, como foi garantido nos casos anteriores.

Figura 3: Dominância dos Heterozigotos



FONTE: Os Autores (2018)

Como $\phi'(0) > 1$, o gráfico da ϕ passará acima da diagonal $y = x$ no intervalo $0 < p < p_3^*$. Consequentemente, p_3^* será assintoticamente estável (Figura 3(a)) e pela análise de estabilidade (veja em ELAYDI (2005)) segue que $\phi'(p_3^*) < 1$, portanto o gráfico da ϕ manter-se-á abaixo da diagonal $y = x$ até chegar ao ponto fixo p_2^* . As Figuras 3(a) e 3(b) mostram essa ocorrência.

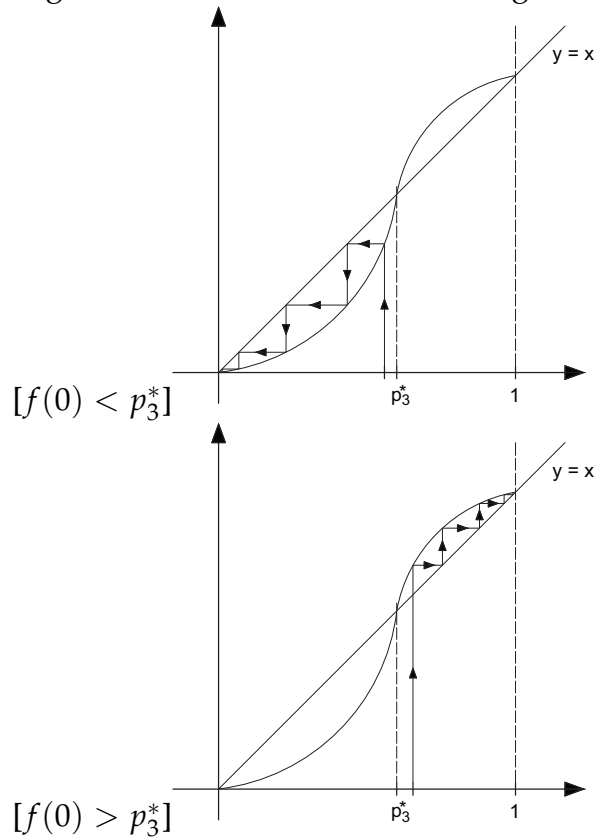
Aplicando a teia aranha, tanto para $f(0) < p_3^*$ (Figura 3(a)) ou para $f(0) > p_3^*$ (Figura 3(b)), segue que: $\lim_{t \rightarrow \infty} f(t) = p_3^*$, que implica a dominância dos heterozigotos, Aa .

• **Caso IV: Dominância dos Homozigotos**

Assumiremos para o último caso que $w_{Aa} < w_{AA}$ e $w_{Aa} < w_{aa}$ ocorrem, e como no Caso III, a função ϕ possui os três pontos fixos no intervalo $[0, 1]$. Também de forma similar, veja que, como $\phi'(0) < 1$, o gráfico da ϕ estará abaixo da diagonal $y = x$ até o ponto fixo p_3^* e acima da diagonal $y = x$ a partir deste ponto.

Pela teia de aranha (Figura 4), segue que: $\lim_{t \rightarrow \infty} f(t) = 0$, quando $f(0) < p_3^*$ e $\lim_{t \rightarrow \infty} f(t) = 1$, quando $f(0) > p_3^*$.

Figura 4: Dominância dos Homozigotos



FONTE: Os Autores (2018)

Finalmente, este caso nos mostrou que $w_{Aa} < w_{AA}$ e $w_{Aa} < w_{aa}$ acarretam a dominância dos homozigotos, AA e aa .

5. Considerações Finais

Levando em conta as numerosas aplicações da genética de populações e a relevância que elas exercem na sociedade, torna-se imprescindível a interdisciplinaridade entre as áreas da biologia e da matemática. Com este trabalho, através da construção do modelo matemático, foi possível perceber como essas ciências se relacionam harmonicamente.

Com o modelo construído vimos a grande utilidade da ferramenta teia de aranha. Mesmo nos deparando com uma equação de diferenças não linear, não foi necessário encontrar uma solução para obtermos conclusões a respeito das frequências no decorrer do tempo.

Embora os resultados obtidos nos quatro casos estudados pareçam óbvios, a análise do modelo fornece-nos também resultados quantitativos. No caso em que os heterozigotos dominaram foi possível observar qual é exatamente o valor do limite das frequências. Já no caso da dominância dos homozigotos pudemos encontrar a linha que divide as regiões de dominância dos alelos. Portanto, o modelo não nos promove apenas uma análise qualitativa dos seus resultados, por meio do método da teia de aranha, mas também nos direciona a valores exatos de dominância.

Por fim, observamos que este trabalho, embora não possua a análise de um modelo inédito, tem grande utilidade científica, principalmente por não existirem referências sobre o assunto no nosso idioma, podendo ser um agente fomentador para essa linha de pesquisa no país.

Referências

1. AMABIS, J. M; MARTHO, G. R. **Biologia das Populações: Genética, Evolução e Ecologia**. 1. ed. São Paulo: Editora 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: Editora Guanabara Koogan S.A., 2004.
6. MAGALHÃES, M. N.; LIMA, A. C. P. **Noções de Probabilidade e Estatística**. 7. ed. São Paulo: Editora da Universidade de São Paulo, 2010.
7. OCONE, D. **Discrete and Probabilistic Models in Biology**. Apostila do Curso de Modelos Discretos e Probabilísticos na Biologia da Universidade de Rutgers, 2014. [1](#), [3](#)