

CAMINHOS CONVERGENTES: NOTAS SOBRE A TEORIA DA COALESCÊNCIA EM GENÉTICA DE POPULAÇÕES DE PLANTAS



RENAN MARCELO PORTELA

**CAMINHOS CONVERGENTES:
NOTAS SOBRE A TEORIA DA
COALESCÊNCIA EM GENÉTICA DE
POPULAÇÕES DE PLANTAS**

Todo o conteúdo apresentado neste livro é de responsabilidade do(s) autor(es).

Esta publicação está licenciada sob [CC BY-NC-ND 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Conselho Editorial

Prof. Dr. Ednilson Sergio Ramalho de Souza - UFOPA
(Editor-Chefe)

Prof. Dr. Laecio Nobre de Macedo-UFMA

Prof. Dr. Carlos Erick Brito de Sousa-UFMA

Prof. Dr. Rodolfo Maduro Almeida-UFOPA

Prof. Me. Éfrem Colombo Vasconcelos Ribeiro-IFPA

Prof. Me. Jorge Carlos Silva-ULBRA

“Acreditamos que um mundo melhor se faz com a difusão do conhecimento científico”.

Equipe Home Editora

Renan Marcelo Portela

**CAMINHOS CONVERGENTES:
NOTAS SOBRE A TEORIA DA
COALESCÊNCIA EM GENÉTICA DE
POPULAÇÕES DE PLANTAS**

1ª Edição

Belém-PA
Home Editora
2024

© 2024 Edição brasileira
by Home Editora

© 2024 Texto
by Autor

Todos os direitos reservados

Home Editora
CNPJ: 39.242.488/0002-80
www.homeeditora.com
contato@homeeditora.com
91988165332
Tv. Quintino Bocaiúva, 23011 - Batista
Campos, Belém - PA, 66045-315

Editor-Chefe

Prof. Dr. Ednilson Ramalho

Projeto gráfico

homeeditora.com

Revisão, diagramação e capa

Autor

Bibliotecária

Janaina Karina Alves Trigo Ramos

CRB-8/009166

Produtor editorial

Laiane Borges

Dados Internacionais de Catalogação na publicação (CIP)



C183

Caminhos convergentes: notas sobre a teoria da coalescência em genética de populações de plantas / Renan Marcelo Portela. – Belém: Home, 2024.

Livro em PDF
44p.

ISBN: 978-65-6089-058-9

DOI: 10.46898/home.94c999c8-31c4-42b9-b986-334aea7ba2b7

1. Teoria da Coalescência. I. Portela, Renan Marcelo. II. Título.

CDD 570

Índice para catálogo sistemático

I. Biologia.

SUMÁRIO

APRESENTAÇÃO	6
1 TEORIA DA COALESCÊNCIA: BASES HISTÓRICAS.....	7
2 GENEALOGIAS E ANCESTRAL COMUM MAIS RECENTE.....	17
3 CONCEITOS MATEMÁTICOS E ESTOCÁSTICOS NA COALESCÊNCIA	21
4 ESTIMATIVAS DE PARÂMETROS GENÉTICOS COM BASE NA COALESCÊNCIA	25

APRESENTAÇÃO

"A teoria da coalescência é um pilar fundamental na genética de populações, oferecendo uma lente poderosa através da qual podemos examinar o passado evolutivo.

Aqui será explorado como a teoria da coalescência modela a nossa compreensão da evolução e como ela pode ser aplicada para estimativa de parâmetros genéticos, como fluxo gênico e tamanho efetivo populacional.

O livro está dividido em quatro capítulos, onde cada um deles oferece *insights* sobre diferentes aspectos da teoria da coalescência, desde os aspectos históricos e os cientistas que serviram de base para sua proposição, até os softwares mais utilizados nas inferências.

Este livro foi derivado do resumo das extensas revisões feitas durante seis anos de pós graduação, sendo escrito com o objetivo de ajudar os cientistas que buscam entender os meandros que permeiam essa complexa e fascinante teoria."

CAPÍTULO I

TEORIA DA COALESCÊNCIA: BASES HISTÓRICAS

1 TEORIA DA COALESCÊNCIA: BASES HISTÓRICAS

Este capítulo aborda o desenvolvimento da Teoria da Coalescência de Kingman, situando-a no contexto da genética de populações. Vamos explorar as contribuições chave e os avanços teóricos que moldaram esse campo, culminando nas teorias de John Kingman.

As Bases da Genética e Biologia Populacional

O início da genética de populações remonta a Mendel, cujas leis de hereditariedade foram essenciais para estabelecer a transmissão de traços genéticos (MENDEL, 1866). A integração da genética mendeliana com a teoria da evolução de Darwin (DARWIN, 1859) formou a base da biologia evolutiva. A ponte entre esses campos foi fortalecida com o princípio de Hardy-Weinberg (HARDY, 1908; WEINBERG, 1908), que estabeleceu a constância das frequências alélicas em populações ideais.

O século XX foi um período de extraordinários avanços na genética e na biologia evolutiva, marcado por contribuições fundamentais que moldaram a compreensão moderna da evolução. A síntese evolutiva moderna, um movimento que integrou a genética mendeliana com a teoria da evolução de Darwin, desempenhou um papel central neste desenvolvimento.

Contribuições de R. A. Fisher

Ronald Aylmer Fisher foi uma figura central na integração da genética com a teoria evolutiva. Em sua obra seminal "*The Genetical Theory of Natural Selection*" (1930), Fisher introduziu métodos matemáticos na genética, estabelecendo uma base quantitativa para a teoria evolutiva. Ele desenvolveu o conceito de "*fitness*" genético e explicou como a seleção natural e a herança mendeliana juntas poderiam levar à evolução (FISHER, 1930). A abordagem de Fisher foi

revolucionária, pois ofereceu um modelo matemático para a seleção natural, algo que Darwin não havia conseguido.

O Trabalho de Sewall Wright

Sewall Wright, outro gigante da genética de populações, é conhecido por seu conceito de paisagens adaptativas/ paisagens de aptidão. Esse conceito possibilita visualizar a relação entre genótipos, fenótipos e sucesso reprodutivo, e tem sido uma ferramenta valiosa para entender a complexidade da evolução. Esta abordagem não apenas facilitou a compreensão da dinâmica evolutiva, mas também abriu caminho para novas pesquisas e teorias na genética, como exemplo da coalescência.

Em seu artigo de 1932, Wright descreveu como as populações evoluem através de diferentes combinações de alelos, visualizando a evolução como um movimento através de uma paisagem de "*fitness*" com vários picos e vales (WRIGHT, 1932). Este conceito ajudou a explicar a dinâmica da evolução em populações pequenas e a importância da deriva genética.

As Contribuições de J. B. S. Haldane

John Burdon Sanderson Haldane foi outro pioneiro na aplicação de modelos matemáticos à biologia evolutiva. Em seu livro "The Causes of Evolution" (1932), Haldane analisou o papel das mutações e da seleção natural na evolução. Ele foi um dos primeiros a estimar as taxas de mutação e a discutir a relação entre a seleção natural e a variação genética (HALDANE, 1932). As ideias de Haldane sobre as taxas de mutação e o papel da seleção na evolução abriram novas perspectivas no entendimento da dinâmica evolutiva.

Pesquisadores da Genética de populações

Além de figuras como Ronald Fisher, Sewall Wright, e John Haldane, que foram fundamentais na formação da genética de populações e na síntese evolutiva moderna, há outros cientistas cujas contribuições foram cruciais para o desenvolvimento do campo. Vou destacar alguns deles, juntamente com suas principais teorias e descobertas:

Gustave Malécot: foi um geneticista francês cujo trabalho teve um impacto significativo no campo da genética de populações e na formulação da teoria da coalescência. Suas contribuições podem ser resumidas da seguinte forma:

Trabalho em Genética de Populações e Processos Estocásticos: Malécot foi pioneiro na formulação de modelos de genética populacional como processos estocásticos, particularmente no uso de cadeias de Markov para modelar tais processos. Sua abordagem matemática era notavelmente elegante e precisa, caracterizada por ser compacta e abstrata. A influência de Malécot nesse campo foi fundamental, embora o reconhecimento de suas realizações tenha sido gradual ao longo do tempo. A contribuição de Malécot para a genética de populações e para o desenvolvimento da teoria da coalescência continua a ser valorizada e reconhecida na comunidade científica atual.

Estudos sobre Dispersão Geográfica de Genes e Consanguinidade: Malécot realizou estudos importantes sobre como os genes se dispersam geograficamente e sobre a consanguinidade, ou a probabilidade de dois alelos serem idênticos por descendência de um ancestral comum. Esses estudos foram cruciais para a compreensão da estrutura genética das populações e da ancestralidade compartilhada, conceitos que são centrais na teoria da coalescência (MALÉCOT, 1948).

Julian Huxley: Sua obra "Evolution: The Modern Synthesis" (1942), foi pioneira ao articular a união entre a genética mendeliana e a teoria da evolução de Darwin. Huxley argumentou que a evolução ocorre por meio

da seleção natural atuando sobre variações genéticas e ressaltou a importância da genética na evolução (HUXLEY, 1942).

Theodosius Dobzhansky: Dobzhansky, um geneticista americano de origem russa, foi uma figura chave na síntese evolutiva moderna. Sua obra "Genetics and the Origin of Species" (1937) foi fundamental para integrar a genética mendeliana com a teoria da evolução. Ele também é conhecido por suas pesquisas sobre a genética de populações de *Drosophila* e por popularizar o dito "Nada em biologia faz sentido, exceto à luz da evolução" (DOBZHANSKY, 1937).

Lynn Margulis: Conhecida por sua teoria da endossimbiose, que propôs que certas organelas de células eucarióticas, como as mitocôndrias, originaram-se de organismos procarióticos. Embora sua teoria inicialmente enfrentasse ceticismo, ela eventualmente ganhou ampla aceitação e transformou a compreensão da evolução celular (MARGULIS, 1970).

William Donald Hamilton: Hamilton foi um biólogo evolutivo britânico conhecido por suas teorias de seleção de parentesco e altruísmo. Ele desenvolveu a ideia de que os genes que promovem comportamentos altruístas podem ser favorecidos pela seleção natural se beneficiarem parentes próximos, que compartilham muitos dos mesmos genes (HAMILTON, 1964).

Warren Ewens: Ewens é conhecido por sua "Fórmula de Ewens", que descreve a distribuição de frequências alélicas em uma população. Embora seu trabalho seja distinto da teoria da coalescência, ele complementa o entendimento de como as frequências genéticas são distribuídas e como elas mudam ao longo do tempo, aspectos relevantes para a teoria da coalescência (EWENS, 1972).

Motoo Kimura: Kimura foi um geneticista populacional japonês conhecido por introduzir a Teoria Neutra da Evolução Molecular em 1968. Esta teoria propõe que a maioria das variações genéticas observadas em populações são o resultado de mutações neutras - aquelas que não afetam a aptidão do organismo - e deriva genética. Isso foi uma

mudança significativa em relação à ideia de que a evolução é impulsionada principalmente pela seleção natural (KIMURA, 1968).

James Crow: através de suas pesquisas em genética de populações e evolução molecular, forneceram insights importantes sobre os processos aleatórios na evolução genética. Seu trabalho sobre a deriva genética e a teoria neutra da evolução molecular criou um contexto teórico para a coalescência (KIMURA, 1968; CROW & KIMURA, 1970).

Ernst Mayr: Mayr foi um dos principais biólogos evolutivos do século XX e teve um papel importante na síntese evolutiva moderna. Ele é especialmente conhecido por sua obra em especiação e pelo conceito de "populações reprodutivamente isoladas". Mayr argumentou que as barreiras reprodutivas são cruciais para a formação de novas espécies (MAYR, 1942).

Richard R. Hudson: Hudson fez contribuições significativas na área da genética de populações computacional, especialmente no desenvolvimento de algoritmos para simular a coalescência em populações genéticas. Seu trabalho ajudou a tornar a teoria da coalescência uma ferramenta prática para a análise de dados genéticos (Hudson, 1983).

Montgomery Slatkin: Slatkin é outro geneticista que contribuiu para o desenvolvimento da teoria da coalescência, especialmente em relação ao fluxo gênico e à estrutura populacional. Seu trabalho ajudou a entender como a migração entre populações afeta a ancestralidade genética e a coalescência (SLATKIN, 1987).

Nas décadas de 1970 e 1980, o desenvolvimento de novas técnicas de sequenciamento de DNA trouxe uma compreensão mais profunda da complexidade do genoma. Estas técnicas permitiram aos cientistas analisar sequências genéticas em uma escala e com uma precisão sem precedentes. Conseqüentemente, a genética de populações enfrentou a necessidade de adaptar e desenvolver modelos teóricos que pudessem lidar com essa nova riqueza de dados. Os modelos existentes, muitos dos quais eram baseados em suposições simplificadas e ideais, mostraram-

se insuficientes para explicar os padrões complexos de variação genética observados.

John Kingman e a Proposição da Teoria da Coalescência

A Teoria da Coalescência, desenvolvida por John Kingman, representou uma inovação notável na genética de populações e na biologia evolutiva. Esta teoria aborda a ancestralidade genética de uma maneira que incorpora a natureza estocástica dos processos genéticos e oferece uma nova perspectiva sobre a história genética das populações. Vamos detalhar mais profundamente as características essenciais desta teoria:

A Teoria da Coalescência trata a ancestralidade genética como um processo estocástico, o que significa que é governada por leis de probabilidade. Esta abordagem reconhece que a herança genética é influenciada por eventos aleatórios, como a segregação de alelos durante a meiose e a deriva genética.

Em vez de adotar modelos determinísticos que preveem resultados específicos, a teoria usa modelos estocásticos, como cadeias de Markov, para capturar a imprevisibilidade inerente aos processos genéticos. Isso permite uma representação mais realista da dinâmica evolutiva em populações naturais.

A Teoria da Coalescência analisa a história genética retrocedendo no tempo a partir de um conjunto de alelos em uma população atual até o ponto em que esses alelos coalescem em um ancestral comum.

Esta abordagem oferece *insights* significativos sobre como as linhagens genéticas se “fundem” ao longo do tempo, fornecendo informações sobre eventos passados, como flutuações no tamanho da população, fluxo e seleção.

Facilita o entendimento da história evolutiva das espécies estudadas, pois oferecer um *framework* para entender como eventos evolutivos e demográficos que moldaram a diversidade genética ao longo

do tempo. Isso é crucial para estudos que buscam entender a história evolutiva das espécies e os mecanismos subjacentes à evolução.

Atualmente a Teoria da Coalescência é particularmente útil na era do sequenciamento de DNA em larga escala. Ela permite que os cientistas interpretem padrões complexos de diversidade genética e estrutura populacional encontrados nos dados de sequenciamento.

Além de sua aplicação em genética de populações, a Teoria da Coalescência tem implicações em áreas como epidemiologia, para rastrear a disseminação de patógenos, e em conservação, para entender a diversidade genética de espécies ameaçadas.

Alguns fatores que fizeram com que a teoria da coalescência se popularizasse e se consolidasse como uma metodologia essencial podem ser citados:

Capacidade de Processamento: Os computadores da década de 1980 tinham capacidades de processamento significativamente menores em comparação com as tecnologias atuais. Isso restringia a quantidade de dados genéticos que podiam ser analisados e a complexidade dos modelos que podiam ser simulados.

Armazenamento de Dados: Os sistemas de armazenamento de dados da época eram limitados em termos de espaço e eficiência. Grandes conjuntos de dados genéticos, como os produzidos por sequenciamento de DNA, exigiam recursos substanciais de armazenamento, que muitas vezes excediam a capacidade disponível.

Algoritmos e Software: Embora os princípios fundamentais da bioinformática e da análise genética computacional já estivessem sendo desenvolvidos, os algoritmos e softwares da época eram menos avançados. Isso limitava a capacidade de realizar análises complexas, como as necessárias para aplicar plenamente a teoria da coalescência.

Acesso a Recursos Computacionais: O acesso a recursos computacionais avançados era mais restrito e geralmente limitado a instituições acadêmicas e de pesquisa com financiamento substancial. Isso significava que muitos pesquisadores não tinham acesso às

ferramentas necessárias para aplicar a teoria da coalescência de maneira eficaz.

Desenvolvimento de Métodos Computacionais: A teoria da coalescência, sendo um modelo estocástico complexo, se beneficiaria enormemente de métodos computacionais avançados para sua aplicação. Durante a década de 1980, esses métodos ainda estavam em fases iniciais de desenvolvimento.

Portanto, enquanto a Teoria da Coalescência oferecia um quadro teórico poderoso, a aplicação prática dessa teoria na década de 1980 era limitada pelo estado da tecnologia computacional da época. Com o passar do tempo e o avanço da computação, houve um aumento significativo na capacidade de aplicar a teoria da coalescência e outras técnicas computacionais na genética de populações, levando a novas descobertas e compreensões no campo da biologia evolutiva.

Atualmente, exploramos a análise de grandes volumes de dados, e com o aumento da quantidade de dados genéticos disponíveis, surgiu a necessidade de métodos eficientes para armazenar, gerenciar e analisar essas informações. Isso levou ao desenvolvimento de novos algoritmos e software, bem como ao uso de poder de processamento computacional mais avançado.

A bioinformática emergiu como um campo crucial que combina biologia, informática, matemática e estatística para analisar e interpretar dados biológicos. Ferramentas computacionais tornaram-se indispensáveis para a análise de sequências genéticas, ajudando a identificar genes, prever a estrutura de proteínas e entender as funções genéticas.

O sequenciamento em larga escala necessitava de algoritmos sofisticados para a montagem de sequências, alinhamento de sequências e análise filogenética. A bioinformática desenvolveu métodos como *BLAST* (*Basic Local Alignment Search Tool*) para comparar sequências de proteínas e nucleotídeos, além de algoritmos para a reconstrução de histórias evolutivas e a análise de padrões de variação genética.

Com a teoria da coalescência, tornou-se possível usar simulações computacionais para modelar a história evolutiva de populações. Permitindo aos cientistas testar hipóteses sobre a dinâmica populacional, padrões de migração, efeitos da seleção natural e deriva genética.

CAPÍTULO II

GENEALOGIAS E ANCESTRAL COMUM MAIS RECENTETE

2 GENEALOGIAS E ANCESTRAL COMUM MAIS RECENTE

Esse capítulo do livro aborda a crucial intersecção entre a construção de árvores genealógicas dos alelos e a reconstrução de eventos históricos na genética de populações. Este capítulo mergulha na maneira como os cientistas usam árvores genealógicas para mapear relações evolutivas e como esses métodos podem ser aplicados para reconstruir eventos históricos significativos.

Genealogia até o Ancestral Comum Mais Recente

O conceito de Ancestral Comum Mais Recente (ACMR) é um elemento central tanto na teoria da coalescência quanto na filotaxia, e seu entendimento é crucial para a genética de populações e a biologia evolutiva. Vamos explorar este conceito mais a fundo.

A Teoria da Coalescência, fundamentalmente baseada na genealogia dos alelos, representa uma abordagem distintamente diferente das genealogias taxonômicas tradicionais. Enquanto as genealogias taxonômicas focam nas relações entre espécies ou grupos maiores, a teoria da coalescência concentra-se na ancestralidade de alelos específicos dentro de uma população ou entre populações. Vamos explorar mais detalhadamente estas diferenças:

Em qualquer grupo de organismos, o ACMR é o indivíduo a partir do qual todos os membros do grupo são descendentes. Em termos genéticos, é o ponto em uma árvore genealógica onde as linhagens de dois ou mais indivíduos se fundem.

A teoria da coalescência utiliza modelos estocásticos para rastrear as linhagens genéticas de volta no tempo até o ACMR. Isso envolve analisar como alelos ou genes compartilhados por uma população podem ser traçados até um ancestral comum.

A determinação do ACMR é crucial para compreender a história evolutiva de uma população ou espécie. Ela fornece informações sobre eventos de especiação, migração e outros processos evolutivos.

Quadro 1: Aspectos que geralmente diferenciam ACMR na Filogenética e na Teoria da Coalescência

Aspecto	ACMR na Filotaxia	ACMR na Teoria da Coalescência
Foco	Relações entre espécies ou grandes grupos taxonômicos.	Alelos específicos dentro de uma população ou entre populações próximas.
Escala Temporal	Abrange milhões de anos, focando em ancestrais distantes.	Foca em história evolutiva recente, geralmente abrangendo poucas gerações.
Dados Utilizados	Combinação de dados morfológicos, fósseis e dados genéticos.	Principalmente dados genéticos detalhados, como sequências de DNA.
Aplicação	Usado para construir árvores filogenéticas e mapear a história evolutiva de diferentes espécies.	Usado para entender a história genética e a dinâmica evolutiva de alelos dentro de populações.
Escala de Análise	Macroevolutiva - relações entre espécies e divergência de grandes linhagens.	Microevolutiva - eventos genéticos e dinâmica dentro de populações ou entre populações próximas.
Importância	Fundamental para compreender a origem, diversificação e relações evolutivas entre grupos de organismos.	Crucial para estudar processos microevolutivos como fluxo gênico, seleção, deriva genética e mutação.
Exemplos de Estudo	Estudo da divergência entre monocotiledôneas e dicotiledôneas, origem de grandes famílias de plantas.	Análise da diversidade genética em culturas agrícolas, história evolutiva de espécies de plantas endêmicas.
Representação Gráfica	Árvores filogenéticas mostrando ramificações entre diferentes espécies de plantas ao longo do tempo.	Diagramas de coalescência mostrando como alelos específicos em populações de plantas coalescem ao longo de gerações.
Interpretação	Enfatiza a divergência de espécies de plantas e a especiação ao longo do tempo evolutivo.	Enfatiza a relação e ancestralidade de alelos específicos em plantas em uma escala temporal mais curta.

Este enfoque permite aos cientistas compreender como diferentes cópias de um gene em uma população se relacionam umas com as outras ao longo do tempo.

A coalescência é particularmente útil na análise de processos microevolutivos dentro de populações, como fluxo gênico, seleção, deriva genética e mutação, oferecendo um quadro detalhado da dinâmica evolutiva em uma escala mais granular. Por outro lado, genealogias taxonômicas são ferramentas essenciais para entender a história evolutiva em larga escala, como a origem e diversificação de grandes grupos de organismos ao longo de milhões de anos (PORTELA, 2019).

A escala e foco delas são diferentes, enquanto as genealogias dos alelos são focadas em eventos genéticos dentro de populações ou entre populações próximas, concentrando-se em uma escala temporal mais curta, as genealogias taxonômicas abrangem uma escala temporal e biológica muito mais ampla, lidando com as relações entre espécies e grupos taxonômicos ao longo de milhões de anos.

A construção de genealogias dos alelos geralmente depende de dados genéticos detalhados, como sequências de DNA ou polimorfismos genéticos específicos. Por outro lado, as genealogias taxonômicas podem utilizar uma variedade mais ampla de dados, incluindo características morfológicas, fósseis e, cada vez mais, dados genéticos comparativos.

CAPÍTULO III

CONCEITOS MATEMÁTICOS E ESTOCÁSTICOS NA COALESCÊNCIA

3 CONCEITOS MATEMÁTICOS E ESTOCÁSTICOS NA COALESCÊNCIA

A teoria da coalescência é uma abordagem retrospectiva que tem como objetivo modelar a genealogia dos alelos amostrados, tendo como base o modelo populacional proposto por Fisher (1930) e Wright (1931). A hipótese é de que os alelos amostrados no presente tenham surgido de um único alelo ancestral, dito alelo comum mais recente (NORDBORG, 2001).

Como exemplo, considere um determinado loco amostrado em uma única população, que segue o modelo de Fisher-Wright (FW; Fisher, 1930; Wright, 1931). A ancestralidade dessa amostra é investigada a partir do presente em direção ao passado, até o ponto em que os alelos do referido loco se coalescem em seu ancestral comum mais recente (ACMR). A ocorrência de um evento coalescente é a fusão de dois alelos do tempo (t) em um único alelo qualquer presente no tempo ($t-1$). Em visão prospectiva da história evolutiva das populações, os eventos coalescentes correspondem aos eventos de divergência (FU; LI, 1999).

Em uma amostra com tamanho n , têm-se $n-1$ eventos de coalescência caminhando-se em direção ao passado, de forma que o último evento coalescente corresponde ao ancestral comum mais recente da amostra (NORDBORG, 2001; WAKELEY, 2009). Ao considerar uma amostra de tamanho n , tem-se um total de $\frac{n(n-1)}{2}$ possíveis combinações para o primeiro evento de coalescência, sendo o raciocínio análogo aos demais eventos coalescentes (FU; LI, 1999; PORTELA, 2019).

A probabilidade de duas amostras quaisquer coalescerem em alguma geração no passado é de $\frac{1}{2N_e}$. Tem-se ainda que a probabilidade de duas amostras não coalescerem é de $1 - \frac{1}{2N_e}$, sendo N_e o tamanho efetivo populacional. Logo, a probabilidade de ocorrer o primeiro evento coalescente é de $\frac{n(n-1)}{2} * \frac{1}{2N_e}$ considerando um total de n amostras. Isso se dá pelo fato de que em uma população com um tamanho efetivo

populacional constante com $2N_e$ cópias de cada *locos*, existem $2N_e$ "pais em potencial" na geração anterior (ROSA, 2016). Nota-se ainda que, a princípio, o modelo de FW descreve um processo de amostragem binomial de alelos em cada geração.

Quando os eventos coalescentes são modelados de acordo com a distribuição binomial, o tempo de coalescência, que diz respeito aos intervalos entre os eventos coalescentes sucessivos, em cada geração anterior, é modelado de acordo com a distribuição geométrica, de forma que a probabilidade de não ocorrer coalescência nas gerações precedentes de $t-1$ é multiplicada pela probabilidade de coalescência na geração de interesse, sendo $P_c(t) = \left(1 - \frac{1}{2N_e}\right)^{t-1} * \frac{1}{2N_e}$ (HALLIBURTON et al., 2004).

Quando uma amostra de tamanho n é muito grande, a probabilidade de coalescer é muito pequena, e os eventos coalescentes são estatisticamente modelados seguindo a distribuição de Poisson, tendo a taxa de um evento de coalescência ($\lambda = 1$) para cada amostra em um intervalo de tempo (t_1, t_2, t_3 , até t_n). Isso ocorre devido a distribuição binomial se aproximar de uma distribuição de Poisson, quando a probabilidade de sucesso é pequena e n muito grande. Para esse caso, os tempos de coalescência são modelados usando a distribuição exponencial, onde o tempo esperado de coalescência é dado por $P_c(t) = \frac{1}{2N_e} e^{-\frac{t-1}{2N_e}}$ (HALLIBURTON et al., 2004).

As diferentes combinações possíveis para se obter eventos de coalescência determinam o número de configurações possíveis (também chamado de número de genealogias) envolvendo as amostras das populações atuais. Dessa maneira, o processo de inferência de certa genealogia, segundo a teoria da coalescência, baseia-se em diferentes modelos probabilísticos (ROSA, 2016).

O tempo necessário para ocorrer um evento coalescente é proporcional ao tamanho efetivo populacional e amostral, de modo em que os primeiros eventos coalescentes devem acontecer, em média, mais rapidamente no passado, enquanto que os coalescentes mais antigos irão

demorar cada vez mais à medida em que o ACMR se aproxima (WAKELEY, 2009).

Kingman ampliou a observação de Wright levando dois cromossomos escolhidos aleatoriamente em uma população de tamanho N sobre $2N$ gerações até se encontrarem no seu ancestral comum mais recente. Suas descobertas permitiram o uso de uma amostra aleatória de indivíduos para inferir parâmetros para toda a população, demonstrando, teoricamente, que é possível calcular a probabilidade de certa genealogia para qualquer número de indivíduos (NORDBORG, 2001).

CAPÍTULO IV

ESTIMATIVAS DE PARAMÊTROS GENÉTICOS COM BASE NA COALESCÊNCIA

4 ESTIMATIVAS DE PARÂMETROS GENÉTICOS COM BASE NA COALESCÊNCIA

Historicamente, métodos estatísticos baseados em frequências alélicas foram usados para estimar os parâmetros genéticos populacionais a partir de conjuntos de dados (ROSA, 2016). Esses métodos foram aprimorados e se expandiram em diversas variantes que foram aceitos e debatidos por muitos pesquisadores, citando-se as estatísticas *F* de Wright (1951, 1978), as de Nei (1973) e suas extensões (NATH; GRIFFITHS, 1996), inferência bayesiana STRUCTURE (PRITCHARD et al., 2000), componentes principais (CRUZ et al., 2011), agrupamentos hierárquicos clássicos (ODONG et al., 2011), entre outros. De modo geral, a maioria desses métodos utiliza as frequências alélicas para capturar níveis de diferenciação. Entretanto, isso pode ser o resultado de eventos de fluxo gênico recentes, de modo a não considerar os eventos migratórios ocorridos no passado (WILSON; RANNALA, 2003; ROSA, 2016), dificultando o entendimento das relações e do passado evolutivo das populações.

As estatísticas *F* de Wright ainda são um marco inicial de qualquer estudo genético populacional, especialmente em estudos visando à conservação. Entretanto, ao longo dos anos, as limitações têm se tornado evidentes (NEIGEL, 2002). Muitos desses métodos citados usam suposições restritivas, por exemplo, não permitindo considerar mutação. As estatísticas *F*, assim como os métodos análogos ao F_{ST} , são frequentemente empregados em pares de populações, podendo levar a estimativas tendenciosas dos parâmetros (BEERLI, 2004; SLATKIN, 2005) e a reutilização de dados nesses métodos de pares é indesejável de um ponto de vista estatístico (BEERLI, 2009).

Em 1982, Sir John Frank Charles Kingman desenvolveu a teoria da coalescência (Kingman, 1982 a, b, c). Os trabalhos de Kingman passaram a ter enorme representatividade para os estudos que envolvem a genética de populações e evolução, abrindo portas para novos métodos em genética de populações, levando-se em consideração que as

inferências baseadas na teoria da coalescência não eram praticáveis até meados dos anos 1990, devido a restrições computacionais (BEERLI, 2009).

Tamanho Efetivo Populacional

O tamanho efetivo populacional (N_e) é um dos vários “conceito-chave” em genética de populações, sendo inicialmente introduzido por Wright (1930) em seu trabalho “*Evolution in Mendelian Populations*”. Sua finalidade é fornecer uma maneira de calcular a taxa de mudança evolutiva causada pela amostragem aleatória de frequências alélicas em uma população finita (deriva genética).

O conceito de tamanho efetivo populacional refere-se ao número de indivíduos que se reproduzem e deixam descendentes, desta maneira, transmitindo os genes para a próxima geração. Esse conceito é fundamental para fins evolutivos, conservação genética e melhoramento de espécies pois é um dos principais fatores determinantes na perda de variabilidade em populações ameaçadas de extinção.

Quanto menor o tamanho efetivo de uma população, maiores são os efeitos da deriva genética, como aumento da variância entre populações e diminuição da diversidade dentro de populações, colaborando para fixação ou perda de alelos (RIDLEY, 2006). Uma população com um pequeno N_e experimentará uma deriva genética mais rápida e uma perda mais rápida de variabilidade genética em comparação com uma população maior.

O tamanho efetivo conceitualmente equivale-se ao de uma população idealizada em relação ao modelo populacional de Fisher (1930) e Wright (1931), considerando que reduz a variabilidade genética na mesma taxa que a população do estudo (CABALLERO, 1994; CHARLESWORTH, 2009; SEBBENN; SEOANE, 2005). Dessa forma, o tamanho efetivo populacional teórico pode diferir do tamanho censo de uma população, sendo, na maioria das vezes, menor (WRIGHT, 1938).

Isso pode diferir devido aos sistemas de acasalamento e variação no sucesso reprodutivo entre os indivíduos.

Uma população idealizada, neste contexto, é uma que segue certas premissas simplificadoras, como acasalamento aleatório, ausência de seleção, mutação, migração e tamanho populacional constante ao longo das gerações. Estas premissas raramente são cumpridas em populações naturais, mas oferecem um ponto de referência para análises teóricas.

O N_e é amplamente utilizado em modelos evolutivos e análises estatísticas para prever o comportamento de traços genéticos em populações ao longo do tempo, particularmente em estudos que utilizam a teoria da coalescência.

Na literatura, são descritos diferentes tipos de estimadores para o tamanho efetivo, podendo eles serem baseados em dados fenotípicos ou em marcadores moleculares, sendo usado como medida de representatividade genética a partir determinada amostra (VENCOVSKY; CROSSA, 2003).

As estimativas de N_e são cruciais para entender como a deriva genética afeta determinada a evolução. Populações com N_e pequeno são mais suscetíveis à fixação aleatória de alelos e à perda de diversidade genética, o que pode ter implicações significativas para a perpetuação, sobrevivência e adaptabilidade da população.

Flutuações no tamanho populacional ao longo do tempo podem reduzir o N_e em relação ao tamanho médio da população.

A distribuição desigual de prole entre indivíduos, a existência de subpopulações isoladas e os sistemas de acasalamento não aleatórios (como a endogamia) também afetam o N_e , frequentemente reduzindo-o em comparação com o tamanho real da população (FRANKHAM et al., 2014).

O fluxo gênico é um termo genérico que diz respeito à migração de genes de uma população para outra, onde inclui-se todos os mecanismos possíveis que resultam na movimentação de alelos entre populações (NEIGEL, 1997). É juntamente com as medidas de diversidade genética e

o grau de estruturação genética espacial das populações, um dos parâmetros genéticos populacionais mais importantes para fins de conservação genética e manejo, pois o fluxo pode restringir o processo evolutivo (SLATKIN, 1989; TAMBARUSSI et al., 2015).

Considerando o processo evolutivo como uma mudança nas frequências alélicas em uma população, ao longo do tempo (WILSON; BOSSERT, 1971), e que essas mudanças evolucionárias ocorrem por meio de quatro fenômenos diferentes, frequentemente chamados de “forças evolutivas” ou “mecanismos evolutivos”, os quais são a mutação, a seleção, a deriva e o fluxo gênico, onde esses mecanismos, agindo separadamente ou em conjunto, afetam a mudança de frequências alélicas em uma população de geração em geração.

Fatores como tamanho populacional pequeno e a presença de endogamia reduzem a variabilidade genética dentro de populações e acentuam os efeitos da deriva genética (TAMBARUSSI et al., 2015), contribuindo no sentido de aumentar a diferenciação, divergência e heterogeneidade genética entre as populações. O fluxo gênico age justamente no sentido de se contrapor aos efeitos da deriva genética, permitindo o aumento da variabilidade existente nas populações, espalhando novos genes e combinações de genes em populações (SLATKIN, 1989), e, ao mesmo tempo, a redução da variabilidade entre populações devido ao número de indivíduos que migram, por geração, de uma população para outra (WRIGHT, 1951). No entanto, caso o fluxo gênico não ocorra, é possível que as populações acumulem diferenças genéticas em uma escala temporal, podendo se tornarem espécies distintas (FUTUYAMA, 1992; RIDLEY, 2006; HARTL, 2010).

Há descrito, genericamente, quatro modelos corriqueiramente utilizados para explicar como o fluxo gênico atua nas metapopulações (FUTUYAMA, 1992). Os modelos são:

- I. Modelo de n-ilhas: a migração ocorre ao acaso dentre um grupo de pequenas subpopulações;

- II. Modelo continente-ilha: existe, efetivamente, um movimento unidirecional de uma subpopulação grande para uma menor, quase isolada;
- III. Modelo de “alpondras” (*stepping-stone*): cada subpopulação recebe migrantes somente das subpopulações vizinhas;
- IV. Modelo de isolamento por distância: o fluxo gênico ocorre preferencialmente entre indivíduos mais próximos espacialmente, em subpopulações de distribuição contínua.

O processo de estimação das taxas de fluxo gênico é uma atividade bastante complexa, pois estas taxas não dependem apenas do processo de dispersão, mas, também, do sucesso na reprodução do indivíduo migrante e do tamanho da população que recebe o migrante (ENNOS, 1994), além de fatores como sistema reprodutivo da espécie e as relações ecológicas entre as plantas e seus polinizadores e dispersores (DICK et al., 2003; ELLSTRAND, 2014; RIDLEY, 2006).

O estudo do fluxo gênico pode ser realizado com base em duas estratégias: (i) métodos diretos, e (ii) métodos indiretos (SLATKIN, 1985; 1987; BEERLI, 1998). Os métodos diretos utilizam observações diretas ou estimativas de distância de dispersão e de sucesso reprodutivo, as quais são baseadas em observações do movimento dos vetores de pólen e sementes (LATTA et al., 1998) e na análise da paternidade (JONES; ARDREN, 2003). Já os métodos indiretos utilizam dados genéticos de amostras de indivíduos, oriundos de várias subpopulações, para realizar inferência das taxas de migração (BEERLI, 1998).

Os métodos diretos podem ajudar a determinar o padrão de migração dos indivíduos durante o estudo e podem fornecer informações sobre um histórico muito recente. Pressupõe-se que os poucos indivíduos rastreados com tais métodos sejam escolhidos aleatoriamente e que seus movimentos não sofram viés, e os dados coletados para este caso podem fornecer informações interessantes sobre o padrão de migração de uma população específica (BEERLI, 1998).

Entretanto, os métodos diretos possuem várias limitações, sendo pouco viável em populações arbóreas naturais, isso porque taxas de

migração pequenas podem ser indetectáveis e a precisão e a confiabilidade das estimativas dos parâmetros são baixas quando o estudo é baseado em poucos indivíduos (BEERLI, 1998). Se o estudo for de curta duração e não for repetido, não é possível saber se o padrão de migração observado foi acidental ou geral, tendo em vista que a observação da dispersão está necessariamente limitada no tempo e no espaço. Dessa forma, as estimativas indicam o fluxo gênico de um tempo particular, não sendo capaz de captar os eventos ocorridos no passado histórico das populações (BEERLI, 1998).

Os métodos indiretos, utilizando dados genéticos de amostras de indivíduos, são muito utilizados para populações vegetais, fazendo uso de modelos probabilísticos. Os dados genéticos podem vir de diferentes tipos de marcadores moleculares, sendo esses dados a base para inferência nos padrões de fluxo gênico (GRIFFITHS et al., 2004).

Como vantagem do método indireto, não há necessidade de rastrear os indivíduos ao longo do tempo. Para este método, as estimativas são médias ao longo do tempo evolutivo e revelam padrões de migração gerais (não individuais) (BEERLI, 1998). Esse método pode ser aplicado a qualquer organismo; contudo, apresenta como desvantagem a necessidade de assumir que os marcadores são seletivamente neutros, de modo que a similaridade genética entre diferentes subpopulações seja resultado da migração e não de pressões de seleção (BEERLI, 1998). Ainda, os marcadores precisam mostrar elevado polimorfismo para que seja possível avaliar as diferenças entre subpopulações. Um marcador com uma taxa de mutação muito lenta não revelará eventos de migração recentes, porém, pode ter algumas informações sobre eventos de migração no passado, quando comparado com populações geograficamente mais distantes (BEERLI, 1998).

Várias abordagens usando dados de marcadores genéticos para a inferência de taxas de migração são descritas, citando-se os estimadores baseados em frequências alélicas e estatística F de Wright, análogos como descrito por Michalakis e Excoffier, 1996); estimadores de máxima verossimilhança, baseados em frequências alélicas (RANNALA;

HARTIGAN, 1996; TUFTO et al. 1996); estimadores baseados nas genealogias dos indivíduos amostrados (KINGMAN, 1982), com taxas de migração estimadas usando procedimentos de Bahlo e Griffiths (1998); Beerli e Felsenstein, (1998); Wakeley, (1998).

Alguns estimadores são misturas dos grupos citados (BEERLI, 1998), como no caso de Slatkin e Maddison (1989), que desenvolveram um método que utiliza resultados da teoria da coalescência e logo em seguida apresenta uma tabela de interpolação produzida pela simulação da coalescência com migração, na qual o número mínimo de eventos de migração encontrados na melhor genealogia está relacionado a uma taxa de migração $4N_e m$.

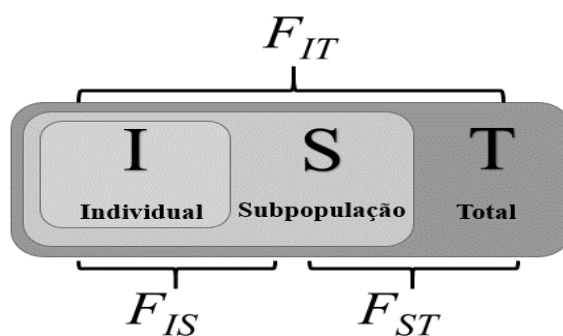
A maioria desses estimadores foram desenvolvidos partindo de pressupostos simplificados, assumindo que a migração e as taxas de mutação são constantes ao longo do tempo (BEERLI, 1998). Além disso, quase todos os métodos usam uma população de tamanho constante. Os métodos que permitem supor que as subpopulações não são constantes estão pouco desenvolvidos e são escassos, citando-se Bahlo e Griffiths (1998). Entretanto, a disponibilidade de métodos para estimar parâmetros populacionais sob condições não constantes aumentará o desenvolvimento de novas ferramentas estatísticas e computacionais, e a implementação de um novo método geralmente são lentas. Além disso, métodos mais complexos normalmente possuem maior quantidade de parâmetros, que demandarão mais dados para serem estimados, tornando muitas vezes impraticável o processo de estimação (BEERLI, 1998).

Escolher uma abordagem para a estimação de fluxo gênico não é uma tarefa simples, pois, cada um desses métodos tem seu próprio conjunto de suposições, as quais podem ser incompatíveis com o estudo. Além disso, os resultados obtidos com dados reais fornecem uma orientação pouco confiável para a escolha de um método, porque normalmente não sabemos os valores reais dos parâmetros subjacentes (BEERLI, 2009).

Grande parte dos modelos de estimativa de fluxo gênico baseia-se em populações de Fisher-Wright (FISHER, 1930; WRIGHT, 1931). O modelo populacional de Fisher-Wright com migração é relativamente simples e de fácil entendimento, tendo propriedades que facilitam as manipulações matemáticas (BEERLI, 2009). O modelo consiste em uma subpopulação, com um número constante de indivíduos, onde em cada geração os indivíduos produzem um grande número de gametas os quais podem permanecer em uma subpopulação ou migrarem para outra subpopulação. Dessa forma, os novos indivíduos são formados pela “escolha” aleatória de dois gametas em uma subpopulação, onde os novos indivíduos substituem seus genitores (HARTL, 2010).

Wright (1951) descreveu uma estrutura que utiliza as estatísticas F , para uma população subdividida, propondo três coeficientes: F_{IS} , F_{IT} e F_{ST} . Usando o modelo de alelos infinitos, F pode ser entendido como uma probabilidade de que dois alelos escolhidos aleatoriamente sejam idênticos por descendência, e os F 's são proporções onde as letras subscritas são os níveis hierárquicos, como no caso o 'I' é indivíduo, 'S' é subpopulação e 'T' é o total de subpopulações.

Figura 1: Esquema referente a estatísticas F e as relações entre os níveis hierárquicos populacionais.



Fonte: Adaptado de Portela (2019)

Beerli (1998), utilizando conceitos de Michalakis e Excoffier (1996), explica que o F_{ST} pode ser descrito de forma genérica como sendo:

$$F_{ST} = \frac{g(\sigma_w) - g(\sigma_b)}{g(\sigma_b)}$$

Onde, σ_w é a variação dentro da população, σ_b é a variação entre populações e $g(x)$ são funções com correção para os diferentes estimadores análogos ao F_{ST} com variância dentro de uma população (σ_w) ou entre populações (σ_b). Dessa maneira, o F_{ST} pode ser substituído por diferentes estimadores específicos, citando-se alguns como o θ (WEIR, 1996), N_{ST} (LYNCH; CREASE, 1990), $\langle F_{ST} \rangle$ (HUDSON et al., 1992), Φ_{ST} (EXCOFFIER; SMOUSE 1994), ρ_{ST} (ROUSSET, 1996), G_{ST} (NEI, 1973) e R_{ST} (SLATKIN, 1993).

Essas variâncias são proporcionais aos tempos médios de coalescência em uma subpopulação em relação a toda a população. Sendo o F_{ST} interpretado como uma medida da diferenciação entre subpopulações, onde valores próximos de zero indicam que a população não está estruturada (FRANKHAM et al., 2002).

O F_{ST} é de forma corriqueira transformado em uma medida mais direta para migração. Wright (1951) mostrou que, para um modelo de população em ilhas, com um número infinito de subpopulações e nenhuma mutação, o F_{ST} pode ser escrito como sendo:

$$F_{ST} = \frac{1}{1 + 4N_e m} = N_e m = \frac{1}{4} \left(\frac{1}{F_{ST}} - 1 \right)$$

em que, N_e é o tamanho efetivo de uma população Fisher-Wright e m é a taxa de migrantes por geração. Beerli (1998) ressalta que, mesmo a população não se comportando exatamente como uma população de Fisher-Wright, ainda é possível usar N_e para fazer comparações com outras populações. Relaxar a suposição de ter um número infinito de subpopulações é relativamente simples e foi descrito anteriormente, citando-se Li (1976), onde demonstra:

$$F_{ST} = \frac{1}{1 + 4N_e m \frac{d^2}{(d-1)^2}}$$

em que d é o número de subpopulações, d é certamente diferente do número de populações amostradas. Ressalta-se que os resultados para $N_e m$ podem ser muito diferentes dependendo de qual versão do análogo ao F_{ST} que foi utilizada (CHARLESWORTH, 1998).

O modelo de população de n -ilhas faz uso de apenas dois parâmetros, sendo o tamanho efetivo da população (N_e) e a taxa de imigração por geração m . Assume-se que os tamanhos das subpopulações são os mesmos e que a taxa de migração é a mesma entre todas as subpopulações. Entretanto, essas suposições são frequentemente violadas em estudos de populações naturais, para as quais não conhecemos nem os padrões reais de migração nem os tamanhos populacionais (BEERLI, 2009).

Beerli (1998), fazendo uso dos conceitos propostos na teoria da coalescência, expõe que é possível construir a genealogia dos indivíduos amostrados e que, em uma única população, este processo depende apenas do tamanho efetivo populacional.

Kingman (1982) mostrou que a probabilidade de ocorrer um evento de coalescência no passado, entre duas cópias genéticas escolhidas aleatoriamente a partir de uma amostra de tamanho n , no intervalo de tempo (t), que é medido em gerações, escalado pela taxa de mutação (μ) é:

$$\text{Prob}(t | N_e, \mu) = \exp\left(-t \frac{n(n-1)}{4N_e}\right) \frac{2}{4N_e} \quad (3.4)$$

Pode-se ainda estimar $\text{Prob}(g | N_e, \mu)$, sendo a probabilidade de toda uma genealogia (g) iniciando com n amostras de alelos ou sequências e, voltando no tempo, multiplicando as probabilidades para cada intervalo de tempo entre os “nós” nesta genealogia. Dessa forma, seria possível examinar todas as possíveis genealogias e encontrar a genealogia ou grupo de genealogias para as quais a probabilidade dada aos parâmetros da população é a maior.

Esta estrutura proposta por Kingman (1982) pode ser expandida de modo a incluir todos os possíveis parâmetros da população por meio dessa abordagem, de modo que a estrutura original da coalescência pode ser expandida incorporando-se outros parâmetros populacionais (HUDSON, 1990), tais como crescimento populacional (GRIFFITHS; TAVARÉ 1994; KUHNER et al. 1998), taxas de migração (NATH; GRIFFITHS 1996; BAHLO; GRIFFITHS 1998; BEERLI; FELSENSTEIN 1998), taxas de recombinação (GRIFFITHS; MARJORAM 1996), e seleção (KRONE; NEUHAUSER 1997; NEUHAUSER; KRONE 1997). Inclui-se ainda modelos com migração e fluxo gênico assimétrico, podendo ser descrito na forma um sistema de duas populações (BEERLI, 1998) da seguinte maneira:

$$\text{Prob}(t | N_e^{(1)}, N_e^{(2)}, m_{21}, m_{12}, \mu) = \exp\left(-t \left[\frac{n_1(n_1 - 1)}{4N_e^{(1)}} + t \frac{n_2(n_2 - 1)}{4N_e^{(2)}} + n_1 m_1 + n_2 m_2 \right]\right) \beta$$

em que t é o intervalo de tempo; n é o tamanho da amostra; μ é a taxa de mutação; N_e é o tamanho efetivo populacional; β é a probabilidade de um coalescente na população i , com probabilidade $\frac{2}{4N_e^{(i)}}$ ou um evento de migração da população j para i , com probabilidade m_{ji} .

As metodologias mais recentes de fluxo gênico via coalescência, a partir de marcadores microssatélites, baseiam-se, principalmente, nos métodos de máxima verossimilhança e inferência bayesiana (KUHNER; SMITH, 2007).

Nos últimos anos, programas computacionais que possibilitam estimar parâmetros usando dados genéticos sob vários modelos de coalescência foram desenvolvidos (BEERLI; FELSENSTEIN, 1999, 2001; BAHLO; GRIFFITHS, 2000; WILSON; RANNALLA, 2003; De IORIO; GRIFFITHS, 2004; HEY; NIELSEN 2004; BEERLI, 2006; EWING; RODRIGO, 2006; KUHNER, 2006). A utilização dessas abordagens possibilita estimar parâmetros relacionados à genética de populações por

meio de uma abordagem que considera o passado evolutivo e as relações genealógicas dos alelos.

Quadro 2: Exemplos de *softwares* usualmente utilizados nas estimativas de parâmetros genéticos com base na teoria da coalescência.

Software/ pacote	Autores	Exemplos estimativas
BEAST	Drummond et al., 2012	Pacotes de inferência Bayesiana via Cadeias de Markov Monte Carlo (MCMC) para uma ampla gama de modelos de coalescência, incluindo o uso de sequências temporais amostradas.
MIGRATE-N	Beerli e Felsenstein, 2001	Estima tamanhos efetivos populacionais e taxas de migração entre populações, utilizando um modelo de coalescência.
Msprime	Kelleher et al., 2016	Um simulador de coalescência eficiente para grandes amostras e modelos genômicos complexos.
LAMARC	Kuhner, 2006	Estima tamanhos populacionais efetivos, taxas de crescimento e migração, utilizando um algoritmo de Máxima Verossimilhança baseado em coalescência.
IMa2	Hey, 2010	Análise de dados de DNA, RNA e proteínas usando um modelo Bayesiano baseado na teoria da coalescência.
GENECLASS2	Piry et al., 2004	Identificação de indivíduos ou populações utilizando métodos baseados em algoritmos de coalescência e outras técnicas genéticas.
Arlequin	Excoffier e Lischer, 2010	Ferramenta de análise de dados genéticos populacionais que oferece métodos estatísticos baseados em diferentes pressupostos.
CoaSim	Mailund et al.	Dados genéticos sob o modelo de coalescência.
SIMCOAL2	Laval e Excoffier, 2004	Simulação de dados genéticos sob modelos complexos de coalescência.
BayesAss	Wilson e Rannala, 2003	Inferência Bayesiana de fluxo gênico recente utilizando dados de genótipos <i>multilocus</i> .

REFERÊNCIAS

- BAHLO, M.; GRIFFITHS, R. C. Inference from gene trees in a subdivided population. *Theoretical Population Biology*, v. 57, n. 2, p. 79-95, 2000.
- BAHLO M.; GRIFFITHS R. C. Genetree. Program and documentation distributed by the authors. Department of Mathematics, Monash University, Sydney, Australia, 1998.
- BEERLI, P. Effect of unsampled populations on the estimation of population sizes and migration rates between sampled populations. *Molecular Ecology*, v.13, p. 827-836, 2004.
- BEERLI, P. Estimation of migration rates and population sizes in geographically structured populations. In *Advances in Molecular Ecology*, edited by G. Carvalho, of NATO Science Series A: Life Sciences. Amsterdam: IOS Press. v. 306, p. 39-53, 1998.
- BEERLI, P. How to use MIGRATE or why are Markov chain Monte Carlo programs difficult to use? *Population genetics for animal conservation*. Cambridge University Press, Cambridge, United Kingdom, p. 39-77, 2009.
- BEERLI, P. MIGRATE documentation (version 3.6). Technical Report. 2013. Disponível em: <http://popgen.sc.fsu.edu/Migrate/Migrate-n.html>. Acesso em: 14 de junho de 2018.
- BEERLI, P.; FELSENSTEIN, J. Maximum likelihood estimation of migration rates and population numbers of two populations using a coalescent approach. *Genetics*, v. 152, p. 763-773, 1999.
- BEERLI, P.; FELSENSTEIN, J. MIGRATE-N: Maximum likelihood estimation of migration rates and population numbers of two populations using a coalescent approach. 2001.
- CHARLESWORTH, B. Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, v. 10, n. 3, p. 195-205, 2009.
- CROW, J. F.; KIMURA, M. *An Introduction to Population Genetics Theory*. New York: Harper & Row, 1970.
- DARWIN, C. *On the Origin of Species*. London: John Murray, 1859.
- DOBZHANSKY, T. *Genetics and the Origin of Species*. New York: Columbia University Press, 1937.

DRUMMOND, A. J.; et al. BEAST: Bayesian evolutionary analysis by sampling trees. 2012.

ENNOS, R. A. Estimating the relative rates of pollen and seed migration among plant populations. *Heredity*, v. 72, p. 250-259, 1994.

EWENS, W. J. The Sampling Theory of Selectively Neutral Alleles. *Theoretical Population Biology*, v. 3, n. 1, p. 87-112, 1972.

EXCOFFIER, L.; SMOUSE, P. E. Using allele frequencies and geographic subdivision to reconstruct gene trees within a species: molecular variance parsimony. *Genetics*, v. 136, n. 1, p. 343-359, 1994.

EXCOFFIER, L.; LISCHER, H. E. L. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows, 2010.

FISHER, R. A. *The Genetical Theory of Natural Selection*. Oxford: Clarendon Press, 1930.

FRANKHAM, R. Effective population size/adult population size ratios in wildlife: a review. *Genetical Research*, v. 66, n. 2, p. 95-107, 1995.

FRANKHAM, R.; BRADSHAW, C.J.A.; BROOK, B.W. Genetics in conservation management: Revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. *Biological Conservation*. v. 170, p. 56-63, 2014.

FUTUYMA, D. J. *Biologia evolutiva*. Ribeirão Preto: Sociedade Brasileira de Genética, p. 631, 1992.

HALDANE, J. B. S. *The Causes of Evolution*. London: Longmans, Green and Co., 1932.

HALLIBURTON, R. *Introduction to Population Genetics*. Pearson Education, 2004.

HAMILTON, W. D. The Genetical Evolution of Social Behaviour. I & II. *Journal of Theoretical Biology*, v. 7, n. 1, p. 1-16; 17-52, 1964.

HARDY, G. H. Mendelian proportions in a mixed population. *Science*, v. 28, n. 706, p. 49-50, 1908.

HARTL, D. L. *Principles of Population Genetics*. Sunderland, MA: Sinauer Associates, 2010.

HEY, J. IMA2: Isolation with migration models for more than two populations, 2010.

- HEY, J.; NIELSEN, R. Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences*, v. 104, n. 8, p. 2785-2790, 2007.
- HUDSON, R. R. Testing the Constant-Rate Neutral Allele Model with Protein Sequence Data. *Evolution*, v. 37, n. 1, p. 203-217, 1983.
- HUXLEY, J. *Evolution: The Modern Synthesis*. London: George Allen & Unwin, 1942.
- JONES, A. G.; ARDREN, W. R. Methods of parentage analysis in natural populations. *Molecular Ecology*, v. 12, n. 10, p. 2511-2523, 2003.
- KELLEHER, J.; et al. Msprime: a fast coalescent simulator, 2016.
- KIMURA, M. Evolutionary Rate at the Molecular Level. *Nature*, v. 217, p. 624-626, 1968.
- KINGMAN, J. F. C. On the genealogy of large populations. *Journal of Applied Probability*, v. 19, n. A, p. 27-43, 1982.
- KINGMAN, J. F. C. The coalescent. *Stochastic Processes and their Applications*, v. 13, n. 3, p. 235-248, 1982.
- KUHNER, M. K. LAMARC: likelihood analysis with metropolis algorithm using random coalescence. 2006.
- KUHNER, M. K.; SMITH, L. P. Comparing likelihood and Bayesian methods for population genetic analysis using microsatellite data. *Molecular Ecology*, v. 16, n. 22, p. 4879-4891, 2007.
- LAVAL, G.; EXCOFFIER, L. SIMCOAL2: a coalescent simulator, 2004.
- LATTA, R. G.; et al. Seed and pollen flow in fragmented landscapes. *Bioscience*, v. 48, n. 2, p. 104-111, 1998.
- LYNCH, M.; CREASE, T. J. The analysis of population survey data on DNA sequence variation. *Molecular Biology and Evolution*, v. 7, n. 2, p. 377-394, 1990.
- MAILUND, T.; et al. CoaSim: A flexible environment for simulating genetic data under coalescent models. *BMC Bioinformatics*, v. 6, n. 1, p. 252, 2005.
- MALÉCOT, G. *Les Mathématiques de l'Hérédité*. Paris: Masson et Cie, 1948.
- MARGULIS, L. *Origin of Eukaryotic Cells*. Yale University Press, 1970.

MAYR, E. Systematics and the Origin of Species from the Viewpoint of a Zoologist. Harvard University Press, 1942.

MENDEL, G. Experiments on Plant Hybridization. Verhandlungen des naturforschenden Vereines in Brünn, Bd. IV für das Jahr, 1865, Abhandlungen, 1866.

MICHALAKIS, Y.; EXCOFFIER, L. A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. *Genetics*, v. 142, n. 3, p. 1061-1064, 1996.

NEI, M. Molecular Evolutionary Genetics. New York: Columbia University Press, 1973.

NUNNEY, L.; ELAM, D. R. The effective size of fluctuating populations. *Genetics*, v. 125, n. 2, p. 367-376, 1990.

NORDBORG, M. Coalescent Theory. In: BALDING, D. J.; BISHOP, M.; CANNINGS, C. (Eds.). Handbook of Statistical Genetics. John Wiley & Sons, 2001.

PIRY, S.; et al. GENECLASS2: Genetic assignment methods using markers of different types, 2004

RANNALA, B.; HARTIGAN, J. A. Estimation of gene flow and effective population sizes in island populations using a coalescent approach. *Genetics*, v. 142, n. 3, p. 1061-1064, 1996.

RIDLEY, H. N. Dispersal of Plants Throughout the World. Ashford, Kent: L. Reeve & Co., 2006.

ROSA, G. J. M. Statistical and Machine Learning Methods in Animal Sciences. Springer, 2016.

ROUSSET, F. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*, v. 145, n. 4, p. 1219-1228, 1996.

SEBBENN, A. M. SEOANE, C. E. S. Estimativa de tamanho efetivo de endogamia por marcadores genéticos. *R. Árvore*, Viçosa-MG, v.29, n.1, p.1-7, 2005.

SLATKIN, M. A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, v. 139, n. 1, p. 457-462, 1993.

SLATKIN, M. Gene flow and the geographic structure of natural populations. *Science*, v. 236, n. 4803, p. 787-792, 1987.

SLATKIN, M.; MADDISON, W. P. A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics*, v. 123, n. 3, p. 603-613, 1989.

TAMBARUSSI, E. V.; BOSHIER, D.; VENCOVSKY, R.; FREITAS, M. L. M.; SEBBENN, A. M. Paternity analysis reveals significant isolation and near neighbour pollen dispersal in small *Cariniana legalis* Mart. Kuntze populations in the Brazilian Atlantic Forest. *Ecology and Evolution*, Oxford, v. 5, p. 4735-5147, 2015.

TUFTO, J.; et al. A model for polygenic variation and stabilizing selection. *Genetics*, v. 142, n. 3, p. 1061-1064, 1996.

VENCOVSKY, R.; CROSSA, J. Measurements of representativeness used in genetic resources conservation and plant breeding. *Crop Science*, v. 43, p. 1912-1921, 2003.

WAKELEY, J. *Coalescent Theory: An Introduction*. Roberts and Company Publishers, 2009.

WEINBERG, W. Über den Nachweis der Vererbung beim Menschen. *Jahreshefte des Vereins für vaterländische Naturkunde in Württemberg*, v. 64, p. 368-382, 1908.

WEIR, B. S. *Genetic Data Analysis II*. Sunderland, MA: Sinauer Associates, 1996.

WILSON, E. O.; BOSSERT, W. H. *A primer of population biology*. Sinauer, Stamford, Connecticut, USA, 1971.

WILSON, G. A.; RANNALA, B. BayesAss: Bayesian inference of recent migration rates using multilocus genotypes, 2003.

WILSON, I. J.; RANNALA, B. Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*, v. 163, n. 3, p. 1177-1191, 2003.

WRIGHT, S. Evolution in Mendelian Populations. *Genetics*, v. 16, n. 2, p. 97-159, 1931.

WRIGHT, S. Size of population and breeding structure in relation to evolution. *Science*, v. 87, n. 2263, p. 430-2264, 1938.

WRIGHT, S. The genetical structure of populations. *Annual of Eugenics*. v. 15, p. 313-354, 1951.

WRIGHT, S. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. In: *Proceedings of the Sixth International Congress of Genetics*. v. 1, p. 356-366, 1932.

SOBRE O AUTOR

Engenheiro Florestal (2016) pela Universidade Estadual do Centro Oeste, Mestre (2017) e Doutor (2023) em Ciências Florestais com ênfase em proteção de plantas, genética e melhoramento florestal pela Universidade Estadual do Centro Oeste, graduado em Ciências Biológicas (2021) UNICSUL/Positivo e especialista em Biotecnologia e Bioprocessos (2019) pela Universidade Estadual de Maringá.

Atua como Pesquisador Corporativo na Klabin S. A., suas publicações abrangem principalmente as áreas Fitopatologia, Proteção de Plantas, Biometria, Biologia molecular, Genética de Populações, Melhoramento Genético e Resistência Genética a Doenças de Plantas.

CAMINHOS CONVERGENTES: NOTAS SOBRE A TEORIA DA COALESCÊNCIA EM GENÉTICA DE POPULAÇÕES DE PLANTAS

A teoria da coalescência é um pilar fundamental na genética de populações, oferecendo uma lente poderosa através da qual podemos examinar o passado evolutivo.

Aqui será explorado como a teoria da coalescência modela a nossa compreensão da evolução e como ela pode ser aplicada para estimativa de parâmetros genéticos, como fluxo gênico e tamanho efetivo populacional.

O livro está dividido em quatro capítulos, onde cada um deles oferece insights sobre diferentes aspectos da teoria da coalescência, desde os aspectos históricos e os cientistas que serviram de base para sua proposição, até os softwares mais utilizados nas inferências.

Este livro foi derivado do resumo das extensas revisões feitas durante seis anos de pós graduação, sendo escrito com o objetivo de ajudar os cientistas que buscam entender os meandros que permeiam essa complexa e fascinante teoria.”

Autor

Home Editora
CNPJ: 39.242.488/0002-80
www.homeeditora.com
contato@homeeditora.com
91988165332
Tv. Quintino Bocaiúva, 23011 - Batista
Campos, Belém - PA, 66045-315

