



UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE BIOCÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA EVOLUTIVA

Cainã Max Couto da Silva

**Influência da seleção natural em populações nativas de diferentes
ecorregiões americanas**

**Influence of natural selection on native populations from distinct
American ecoregions**

São Paulo

2021

Cainã Max Couto da Silva

**Influência da seleção natural em populações nativas de
diferentes ecorregiões americanas**

**Influence of natural selection on native populations from
distinct American ecoregions**

Tese apresentada ao Instituto de Biociências
da Universidade de São Paulo, para a obten-
ção de Título de Doutor em Ciências, na Área
de Genética e Biologia Evolutiva.

Orientadora: Tábita Hünemeier

São Paulo
2021

Ficha catalográfica elaborada pelo Serviço de Biblioteca do Instituto de Biociências da USP,
com os dados fornecidos pelo (a) autor (a) no formulário:
'<https://biblioteca.ib.usp.br/ficha-catalografica/src/ficha.php>'

Couto-Silva, Cainã Max
Influência da seleção natural em populações
nativas de diferentes ecorregiões americanas /
Cainã Max Couto-Silva ; orientadora Tábita Hünemeier
-- São Paulo, 2021.
124 p.

Tese (Doutorado) -- Instituto de Biociências da
Universidade de São Paulo. Programa de Pós-Graduação
em Genética e Biologia Evolutiva.

1. seleção natural. 2. nativos americanos. 3.
adaptação local. 4. genômica. I. Hünemeier, Tábita,
orient. II. Título.

Bibliotecária responsável pela catalogação:
Elisabete da Cruz Neves - CRB - 8/6228

Dedico este trabalho ao meu filho de quatro patas, meu melhor amigo e fiel companheiro,
que encheu meu coração de amor e ternura, Darwin.

Agradecimentos

À minha orientadora, Tábita Hünemeier, pela orientação, pelas discussões científicas, pela compreensão, e por todas as oportunidades incríveis de aprendizagem dentro e fora do país.

Ao Prof. Dr. David Comas, por me receber de braços abertos para estágio em seu laboratório em Barcelona. E ao pós-doutorando André Flores, que esteve comigo no laboratório de Comas e contribuiu grandemente para o meu desenvolvimento acadêmico.

À todos os pesquisadores de campo que coletaram os dados, principalmente o Prof. Dr. Francisco Maurício Salzano (*in memoriam*), que coletou a maioria das amostras analisadas neste trabalho, e teve um papel fundamental na genética de populações no Brasil.

À toda equipe envolvida no Departamento de Genética e Biologia Evolutiva do Instituto de Biociências da Universidade de São Paulo, pela infraestrutura e serviços prestados.

À CAPES e FAPESP (processo 2017/14916-8) pelas bolsas concedidas e auxílio financeiro para execução desta pesquisa.

À todos os meus colegas de laboratório, Renan Lemes, Marcos Silva, Tiago Silva, Maíra Rodrigues e Gabrielle Rizzato, pelas discussões científicas e todos os bons momentos ao longo do percurso do meu doutorado. Destaco a contribuição especial de Renan Lemes tanto para meu desenvolvimento acadêmico-profissional, como para a elaboração desta tese.

À toda equipe de docentes e discentes do “porão”, local principal onde foi desenvolvida minha pesquisa. Em especial ao André Fonseca, pelas discussões e bons momentos de confraternização.

Aos meus amigos que sempre me apoiaram e fazem parte da minha família, Rafael Santos, Rafael Pinheiro, Diego Silva, Leonardo Jackson, Caio Mendonça, Júlio Afonso e César Henrique.

À minha irmã de consideração, Magna Magalhães, por estar presente em minha vida desde o mestrado, e ter me auxiliado nos momentos mais difíceis.

À Sandra Caballero Salas, pelas conversas maravilhosas, momentos de reflexão, desconstrução, alegria, e por ter me acolhido tão bem em Barcelona.

À minha mãe, Porancy Couto, que sempre batalhou muito para nos manter em casa, e que me ajuda com muito carinho na criação da minha filha.

À minha filha, por me apoiar, compreender, ensinar, e pelo amor incondicional que sinto como pai.

À Talita, mãe do meu filho de quatro patas, Darwin, que esteve comigo do início ao fim do doutorado, me apoiando, apoiando minha família, e proporcionando muitos momentos de alegria e reflexão.

Ao Darwin (*in memoriam*), meu cachorro, meu filho, e meu melhor amigo, que fez dos meus dias, dias mais felizes, e trouxe harmonia para a casa. Sinto muito a sua falta, e quero acreditar que está bem, feliz e correndo em algum lugar além daqui.

À todas as pessoas não citadas aqui, mas que passaram pela minha vida durante o meu doutorado e contribuíram para meu desenvolvimento pessoal.

Sobretudo agradeço às populações indígenas pela participação nesta e em tantas outras pesquisas, contribuindo para o desenvolvimento científico.

1500, o homem branco em Pindorama chegou
Muita riqueza natural foi o que encontrou
Um clima quente, um belo dia e um povo que vivia em harmonia
Ticuna, Caiagangue, Guarani-kwoa, Juruna, Caetés, Xavantes e Tupinambá
Do Oiapoque ao Chuí, no Brasil, testemunhos do maior crime que se viu
Muito ódio, muita maldade, a coroa mandou pra cá a escória da humanidade
Muito sangue, muita matança, esvaindo com toda esperança
Com sentimento de justiça o índio ficou, se levantando bem mais forte contra o opressor
Agora isso é o que importa na sua vida, usando a lança pra curar sua ferida

Direitos iguais e justiça para o povo Tupi e Guarani
E todas as etnias remanescentes daqui

Lista de Abreviaturas

aDNA	ancient DNA
ANC-A	Ancestral A
ANC-B	Ancestral B
AP	Antes do presente
DNA	Deoxyribonucleic Acid
EHH	Extended Haplotype Homozygosity
eQTL	Expression Quantitative Trait Loci
FDR	False Discovery Rate
FUMA GWAS	Functional Mapping and Annotation of Genome-Wide Association Studies
FUNASA	Fundação Nacional da Saúde
Fst	Fixation index
GO	Gene Ontology
GSEA	Gene Set Enrichment Analysis
GWAS	Genome-Wide Association Studies
HGDP	Human Genome Diversity Project
IFC	Ice-free corridor
iHS	Integrated Haplotype Score
KEGG	Kyoto Encyclopedia of Genes and Genomes
LKT	Lewontin-Krakauer Test
LSBL	Locus-Specific Branch Length
MAF	Minor Allele Frequency
ORA	Over-Representation Analysis
PBS	Population Branch Statistics

PCA	Principal Component Analysis
Rsb	iHS across populations
SGDP	Simons Genome Diversity Project
SNP	Single Nucleotide Polymorphism
XP-EHH	Cross-population Extended haplotype homozygosity

Sumário

1	INTRODUÇÃO GERAL	13
1.1	TEORIA EVOLUTIVA	13
1.1.1	Evolução por meio da Seleção Natural	13
1.1.2	Mecanismos Evolutivos	14
1.1.3	Métodos de Detecção da Seleção Natural	15
1.1.3.1	Métodos baseados no desequilíbrio de ligação	15
1.1.3.2	Métodos baseados na diferenciação populacional	16
1.1.3.3	Adaptação poligênica	17
1.2	ADAPTAÇÃO LOCAL NA AMÉRICA DO SUL	18
2	CONSIDERAÇÕES GERAIS	21
3	RESUMO	23
4	ABSTRACT	25
	REFERÊNCIAS BIBLIOGRÁFICAS	27

CAPÍTULO 1

INTRODUÇÃO GERAL

1.1 TEORIA EVOLUTIVA

1.1.1 Evolução por meio da Seleção Natural

A teoria evolutiva por meio da seleção natural foi proposta inicialmente por Charles Darwin e Alfred Russel Wallace em 1858 (DARWIN; WALLACE, 1858) em uma conferência realizada em Londres, um ano antes da publicação da obra seminal “A Origem das Espécies por meio da Seleção Natural” de Charles Darwin (DARWIN, 1859). Em seu próprio livro, Darwin reconheceu que autores anteriores já haviam feito menção à mesma, embora ele tenha o mérito de reunir um vasto conjunto de evidências que suportam especificamente a teoria evolutiva por meio da seleção natural. Assim sendo, Darwin e Wallace propuseram primeiro que as espécies se modificam ao longo do tempo (evolução), onde os descendentes diferem de seus ancestrais – descendência com modificação –, e tais pequenas mudanças, ao longo de um amplo tempo, podem culminar em grandes mudanças. E, segundo, que esta evolução ocorre por meio da seleção natural, em que indivíduos com características que contribuem para a sobrevivência e reprodução, portanto adaptados ao ambiente, tendem a perpetuar, enquanto os demais tendem a ser eliminados (REECE et al., 2013).

A priori, tanto Darwin quanto Wallace não abordaram diretamente a evolução humana, sendo somente em 1871 que Darwin discutiu publicamente sobre a evolução humana em seu livro “A Descendência do Homem” (DARWIN, 1871). Na época, no entanto, Darwin não compreendia, assumidamente, o mecanismo de herança biológica, e tudo indica que não teve conhecimento dos trabalhos de Gregor Mendel, pesquisador contemporâneo de Darwin (embora o contrário não fosse verdade, pois Mendel conhecia o trabalho de Darwin) (FAIRBANKS, 2020). Mendel publicou seu trabalho que hoje compreende os fundamentos das leis de herança em 1866, ainda que tenha sido reconhecido

somente 34 anos depois, após sua morte. A partir da intersecção da evolução e a genética, numerosos avanços e descobertas foram realizadas principalmente ao longo do século XX, sobretudo com contribuições de grandes nomes como Ronald Fisher, Sewall Wright e John Haldane, culminando na base da genética de populações e genética evolutiva, cerne deste trabalho. A seleção natural, no entanto, não é o único mecanismo da evolução, e segundo a teoria neutra (KIMURA, 1969), tampouco o principal. Portanto, compreender e distinguir os mecanismos que promovem a evolução, além da seleção, é uma etapa crucial nos estudos evolutivos.

1.1.2 Mecanismos Evolutivos

A discussão da evolução dentro do contexto de uma mesma espécie é chamada de microevolução, em contraste com o estudo da evolução entre espécies, denominada macroevolução (JOBILING et al., 2014). Nesta tese, discute-se apenas a microevolução da nossa espécie (*Homo sapiens sapiens*) e, portanto, deste ponto em diante, sempre que for mencionado evolução (mecanismos e descrições), lê-se microevolução.

Trazendo para os dias de hoje, do ponto de vista genético, compreende-se por evolução o processo de mudança na frequência alélica ao longo das gerações (VITTI; GROSSMAN; SABETI, 2013). As vias pelas quais este processo evolutivo ocorre são conhecidas como mecanismos evolutivos, que podem ser separados em quatro principais: 1) deriva genética, 2) migração, 3) mutação e, por fim, 4) a seleção natural (HANCOCK; RIENZO, 2008). A ocorrência de qualquer um destes eventos quebra as premissas de um modelo onde não há evolução, denominado Equilíbrio de Hardy-Weinberg (TEMPLETON, 2006).

A mutação consiste na fonte primária de evolução, ou seja, é ela que introduz alterações diretas no DNA, matéria-prima para a evolução. Uma vez que ela introduz variação no DNA, ela tende a contribuir para a diversidade genética da população (LOSOS et al., 2014). A deriva genética pode ser considerada simplesmente como o acaso do processo evolutivo. Em uma população finita, alguns alelos podem mudar de frequência na população de forma aleatória por uma questão puramente amostral (TEMPLETON, 2006). Em contraste com a mutação, que tende a aumentar a diversidade genética, a deriva tende a diminuir a diversidade genética, e é mais forte em populações menores. A migração consiste no movimento de indivíduos ou gametas de uma população para uma área ocupada por outra população, que resulta em um fluxo gênico, ou seja, na troca genética entre as populações (TEMPLETON, 2006). Na genética de populações, ambos os termos migração e fluxo gênico são utilizados de forma equivalente.

Por fim, há a seleção natural, que pode atuar de forma direcional (positiva/negativa) ou balanceadora. A seleção positiva consiste no aumento da frequência de um ou mais alelos, que por sua vez promovem o aumento da sobrevivência, da fertilidade, ou ambos.

A seleção negativa ocorre quando um alelo se torna prejudicial ao organismo, diminuindo a sobrevivência ou fertilidade, e sua frequência diminui e tende a desaparecer rapidamente dentro de uma população. Sugere-se que a seleção negativa seja a mais comum dos tipos de seleção que ocorre nas populações (POUYET et al., 2018; SALAS, 2019). A seleção balanceadora ocorre quando um genótipo heterozigoto é vantajoso, onde o equilíbrio entre diferentes alelos em um gene propicia um fenótipo adaptativo.

Atualmente, dispõe-se de mais métodos estatísticos para detecção da seleção positiva, em contraste com a seleção negativa ou balanceadora. Parece haver também um maior interesse (ou em consequência disso) na seleção positiva, uma vez que alelos que tenham sido selecionados possuem uma conexão forte com regiões funcionais do genoma que também são de interesse da área médica.

1.1.3 Métodos de Detecção da Seleção Natural

Conforme discutido acima, tanto os alelos produzidos por novas mutações, quanto alelos pré-existentes, podem estar sujeitos à seleção natural. Quando isso acontece, ambos deixam marcas no DNA que nos permite inferir quais alelos foram alvo de seleção em um passado recente ou distante.

Dentre as diversas técnicas para inferência de seleção, algumas das abordagens mais comuns incluem os métodos baseados em sítios segregantes (*e.g.* D de Tajima), no desequilíbrio de ligação (*e.g.* iHS, XP-EHH), na distância genética (*e.g.* Fst, PBS), bem como métodos compostos (VITTI; GROSSMAN; SABETI, 2013). Tais métodos são eficazes, quando utilizados em conjunto e também com simulações demográficas ou permutações, para detecção de seleção positiva principalmente monogênica (NIELSEN et al., 2007). Para detecção de adaptação poligênica, no entanto, outras abordagens são utilizadas, como, por exemplo, análises de enriquecimento gênico e de super-representação (WANG et al., 2013; WATANABE et al., 2017; LIAO et al., 2019).

1.1.3.1 Métodos baseados no desequilíbrio de ligação

Os métodos baseados no desequilíbrio de ligação, usualmente denominados como *selective sweeps* (em português, varreduras seletivas), possuem como princípio o efeito *genetic hitchhiking* (ou “efeito carona”), proposto inicialmente em 1974 (SMITH; HAIGH, 1974). De forma resumida, compreende-se por *genetic hitchhiking* quando um alelo sob forte pressão seletiva, bem como seus alelos neutros “vizinhos”, aumentam de frequência na população em decorrência da seleção. Como resultado, há uma perda da diversidade genética (heterozigosidade) no haplótipo, gerando uma extensão de homozigose do haplótipo. O primeiro método baseado no desequilíbrio de ligação – *Extended Homozygosity Haplotype (EHH) Statistics* – explora essa extensão de homozigose para inferir seleção (SABETI et al., 2002). Este tipo de abordagem é extremamente importante na genética evolutiva,

visto que através das variantes neutras ligadas aos alelos sob seleção, torna-se possível detectar estes últimos, mesmo sem saber previamente sua localização genômica (nesta era da genômica, este evento também é referido como varredura adaptativa) (NOVEMBRE; HAN, 2012; STEPHAN, 2019).

A partir da estatística EHH, outros métodos derivados dela foram implementados ao longo dos anos que a sucederam, como Integrated Homozygosity Statistics (iHS) (VOIGHT et al., 2006), cross-population EHH (XP-EHH) (SABETI et al., 2007), e iHS across populations (Rsb) (TANG; THORNTON; STONEKING, 2007).

Cada um dos métodos mencionados acima possui suas vantagens e desvantagens, e todos utilizam diretamente ou indiretamente uma variação da estatística EHH, diferindo na abordagem da respectiva aplicação. O método iHS, por exemplo, identifica regiões de extensão de homozigose, diferindo os alelos ancestrais e derivados, em uma mesma população (intra-populacional), considerado robusto para *sweeps* incompletos (VITTI; GROSSMAN; SABETI, 2013). Ambos os métodos XP-EHH e Rsb são inter-populacionais, e por meio da comparação com a EHH de uma população próxima, tem maior poder para identificar *sweeps* próximos ou já fixados (SUZUKI, 2010). Normalmente, iHS e XP-EHH são utilizados em conjunto nas análises de seleção.

1.1.3.2 Métodos baseados na diferenciação populacional

A métrica mais simples e difundida para mensurar a diferenciação populacional entre duas populações consiste no índice de fixação (F_{st}). Esta também é uma das métricas mais utilizadas na genética de populações, sendo desenvolvida independentemente por Sewall Wright e Gustave Malécot no final da década de 40 (MALÉCOT, 1948; WRIGHT, 1949). O F_{st} , como originalmente proposto, mensura a distribuição de variação genética entre subpopulações, comparando a diversidade genética dentro das subpopulações à diversidade genética da população total. Pode-se, contudo, adaptar a fórmula para comparação entre duas populações (F_{st} par-a-par, ou *pairwise F_{st}*) para utilizá-la como medida de distância genética (BHATIA et al., 2013; JOBLING et al., 2014).

O princípio dos métodos baseados na distância genética parte do pressuposto de que, quando a seleção está atuando em um locus dentro uma população, mas não dentro de outra população próxima, a frequência alélica entre estas populações deve variar e mostrar-se maior na população sob pressão seletiva (VITTI; GROSSMAN; SABETI, 2013). Partindo deste princípio, surgiu o primeiro teste com o objetivo de detectar sinais de seleção natural utilizando o F_{st} – o Lewontin-Krakauer test (LKT) –, que foi publicado em 1973 (LEWONTIN; KRAKAUER, 1973). A partir de então, principalmente com o avanço da tecnologia e aumento dos dados disponíveis, diversos outros testes foram e continuam surgindo (REES; CASTELLANO; ANDRÉS, 2020).

Dentre os métodos baseados na distância genética, destaca-se aqui os métodos

Locus-Specific Branch Length metric (LSBL) (SHRIVER et al., 2004) e Population Branch Statistics (PBS) (YI et al., 2010). Neles, calcula-se o F_{st} par-a-par entre pelo menos três populações: a população-alvo e uma população próxima (ditas populações-irmãs), bem como uma população externa (outgroup), permitindo deste modo isolar o ramo de diferenciação na população-alvo.

1.1.3.3 Adaptação poligênica

Apesar dos testes para detecção de seleção positiva discutidos acima serem muito eficientes para atingir seus objetivos, cabe ressaltar que os mesmos possuem como foco sítios monogênicos, uma vez que identificam loci independentes entre si e com fortes sinais de seleção (PRITCHARD; RIENZO, 2010). Contudo, compreende-se que a maioria dos fenótipos biológicos são poligênicos, ou seja, vários loci atuam em conjunto a fim de contribuir para um dado fenótipo. Nesse contexto, quando considerados de forma individual, apresentam baixa frequência alélica em cada loci (PRITCHARD; PICKRELL; COOP, 2010; VISSCHER et al., 2017). Exemplos de fenótipos poligênicos incluem altura, peso, pigmentação e fertilidade.

Alguns dos métodos para detecção da interação entre genes, que possivelmente atuam em conjunto em um mesmo fenótipo (*i.e.* epistasia), atuam através da verificação de interação estatística entre variantes genéticas intra- ou inter-populacional, quer seja por métodos convencionais de correlação que utilizam regressão linear ou logística, respectivamente (WAN et al., 2010; UEKI; CORDELL, 2012), ou mesmo estatísticas de correlação desenvolvidas especificamente para estudos genéticos (CLIMER et al., 2014; CLIMER; TEMPLETON; ZHANG, 2014). Contudo, adaptação poligênica não pode nem deve ser deduzida apenas por meio da identificação de sinais de epistasia. Diferentemente da detecção de sinais de seleção positiva monogênica, ainda não existem métodos convencionais estabelecidos para detecção de seleção poligênica (PRITCHARD; PICKRELL; COOP, 2010) e, ainda, boa parte da aplicação dos métodos desenvolvidos até a presente data têm sido alvos de críticas (SOHAIL et al., 2019; REFOYO-MARTÍNEZ et al., 2020). Métodos para detecção de adaptação poligênica frequentemente têm utilizado bancos de dados funcionais e fenotípicos, como GWAS Catalog (BUNIELLO et al., 2019), GO (ASHBURNER et al., 2000; The Gene Ontology Consortium, 2019), KEGG (KANEHISA; GOTO, 2000; KANEHISA et al., 2019), entre outros. Outro método baseado em redes (subgrupos), que também adota informações das categorias biológicas presentes em banco de dados públicos, foi desenvolvido a fim de verificar de forma mais sensível sinais de adaptação poligênica (GOUY; DAUB; EXCOFFIER, 2017).

Em contraste com outros métodos que requerem dados fenotípicos ou ambientais associados à amostra (*e.g.* HANCOCK et al., 2010; GÜNTHER; COOP, 2013), ou mesmo demográficos (*e.g.* RACIMO; BERG; PICKRELL, 2018), os métodos que utilizam

categorias biológicas, como a análise de enriquecimento gênico (GSEA, do inglês, *gene set enrichment analysis*) (SUBRAMANIAN et al., 2005) ou como a análise de super representação (ORA, do inglês, *over representation analysis*) (KHATRI; SIROTA; BUTTE, 2012). Por meio dessa abordagem, atribui-se um valor a cada gene (*e.g.* Fst, ou índices de teste de seleção), e testa-se para verificar se um conjunto de genes com os valores mais altos estão super representados em vias ou fenótipos de banco de dados, considerando o pool gênico da amostra (BARGHI; HERMISSON; SCHLÖTTERER, 2020).

Dada a reconhecida escassez de estudos de adaptação poligênica em nativos americanos (MENDES et al., 2020), e que tais métodos têm sido frequentemente aplicados em artigos recentes de adaptação local (BERGEY et al., 2018; HARRISON et al., 2019; HSIEH et al., 2017; LOPEZ et al., 2019; REYNOLDS et al., 2019), optamos por também aplicar métodos de adaptação poligênica no presente estudo.

1.2 ADAPTAÇÃO LOCAL NA AMÉRICA DO SUL

O povoamento da América é um tema amplamente estudado e, desde o início, tem sido alvo de intenso debate na comunidade científica. As diversas áreas da ciência como arqueologia, paleontologia, paleoantropologia, paleoclimatologia, bem como a genética evolutiva, frequentemente se contrapõem. Logo, há um esforço recorrente da comunidade em conciliar os achados dessas diferentes áreas.

Ao passo que a ciência evolui, como, por exemplo, com o advento de novas metodologias, aquisição e disponibilização de dados, principalmente dados de DNA antigo (aDNA, do inglês, *ancient DNA*), nossa compreensão sobre o povoamento torna-se um pouco mais clara, ainda que provavelmente estejamos longe de um consenso geral. Para discorrer sobre este tema, portanto, é importante salientar que este é um campo com mudanças recorrentes, bem como concepções diferentes entre os próprios pesquisadores da área. Sabendo disso, no texto a seguir adotamos como consenso o delineamento realizado por Sutter em um extensivo artigo de revisão (SUTTER, 2020), apresentado aqui de forma a atingir dois objetivos principais: i) introduzir de forma simples e direta os principais pontos para compreensão do povoamento da América, com enfoque na América do Sul, e ii) discorrer sobre suas consequências para os estudos de adaptação local.

A hipótese mais plausível para o povoamento das Américas atribui sua entrada pela província da Beríngia, uma região que conectava a Sibéria ao Alasca e noroeste do Canadá, hoje submersa (Estreito de Bering), mas que ficou exposta durante o último máximo glacial (entre 17 e 24 mil anos atrás), formando uma ponte que não apenas possibilitou a passagem dos seres humanos e outros animais, como provavelmente também serviu de abrigo durante esse período tão hostil (HOFFECKER et al., 2016; SUTTER, 2020).

Este abrigo se explica pelas características climáticas e vegetativas propícias para o refúgio humano, apontadas por estudos paleoambientais (WOOLLER et al., 2018), e também porque duas grandes geleiras na entrada da América bloqueavam a passagem para o continente. Estima-se que os povos ali presentes permaneceram nesta região por aproximadamente 5 a 8 mil anos AP (antes do presente) (FAGUNDES et al., 2008). Neste período, provavelmente acumularam mutações que hoje estão presentes em boa parte dos nativo-americanos, como, por exemplo, novos haplogrupos mitocondriais (TAMM et al., 2007), ou novas variantes com possível maior valor adaptativo, como identificado no gene *FADS* (AMORIM et al., 2017).

Partindo do noroeste asiático e da Sibéria, os primeiros nativos-americanos então se diversificaram de uma população ancestral homogênea, separando-se por volta de 22 a 18 mil anos AP em dois ramos principais: o grupo dos beringianos antigos (AB, do inglês, *Ancient Beringians*) (MORENO-MAYAR et al., 2018a; MORENO-MAYAR et al., 2018b), e o grupo dos nativos-americanos ancestrais, que por sua vez se dividiram em nativos-americanos do Sul, também denominados como ANC-A (Ancestral A), e nos nativos-americanos do Norte, ou ANC-B (Ancestral-B) (RAGHAVAN et al., 2015; MORENO-MAYAR et al., 2018b; POSTH et al., 2018; SCHEIB et al., 2018).

Os primeiros estudos com múltiplas amostras de aDNA da América do Sul com alta cobertura de sequenciamento surgiram em 2018 (POSTH et al., 2018; MORENO-MAYAR et al., 2018b). Apesar de algumas diferenças, ambos concordam que o povoamento da América do Sul ocorreu de forma extremamente rápida, como já anteriormente proposto (LLAMAS et al., 2016). As rotas mais prováveis para migração adentro da América do Sul percorridas pelos nativos-americanos foram pela costa do Pacífico Norte (NPC, do inglês, *North Pacific coast*) e o corredor livre de gelo (IFC, do inglês, *ice-free corridor*), formada entre as duas grandes geleiras (POTTER et al., 2018), sendo que esta última tornou-se viável como rota apenas por volta de 13 mil anos AP (PEDERSEN et al., 2016).

Apesar dos avanços no estudo do povoamento da América, ainda há muito a ser estudado para maior compreensão deste tema, principalmente na América do Sul. Posth et al. (2018) reconheceram na conclusão do próprio artigo, por exemplo, que uma das limitações do trabalho foi a ausência de dados (aDNA) proveniente de populações amazônicas. Assim como as terras altas andinas, a Amazônia faz parte das principais ecorregiões americanas (ANTONELLI et al., 2018), e abrigam ainda hoje povos nativos-americanos, que, em conjunto com os dados de aDNA, contribuem tanto para o estudo do povoamento e demografia dos nativos-americanos, quanto para os estudos de adaptação local a ambientes que poderiam ser considerados como inóspitos.

Desde a saída da África até a entrada na América, as populações passaram por numerosos eventos gargalos de garrafa seguidos de efeitos fundadores, haja visto o evidente decréscimo da diversidade genética ao longo deste percurso (PRUGNOLLE; MANICA;

BALLOUX, 2005). Estima-se que a população ancestral dos nativos-americanos passou por um gargalo de garrafa profundo (FAGUNDES et al., 2008), e o mesmo provavelmente ocorreu em todas as divisões que se sucederam dentro da América. Portanto, torna-se mais do que fundamental analisar de forma cautelosa as análises de seleção natural em populações nativo-americanas, não só para descartar os vieses potenciais da deriva genética, como também o fluxo gênico entre populações de ecorregiões diferentes.

Desta forma, tivemos como objetivo investigar sinais de seleção natural em populações nativas de duas das principais ecorregiões Americanas: Amazônia e Andes, visto que ambas as regiões consistem em “laboratórios naturais” para estudo de adaptação local. Para isso, discorreremos sobre as características de cada uma destas regiões, e aplicamos os métodos de detecção de seleção natural aqui mencionados, em conjunto outras abordagens mais específicas (e.g. metanálise, simulação demográfica).

CAPÍTULO 2

CONSIDERAÇÕES GERAIS

Neste trabalho, analisamos a influência da seleção natural em populações nativas americanas de duas ecorregiões principais: Floresta Amazônica e Andes.

Por meio de múltiplos testes de seleção positiva, análise de enriquecimento gênico, anotação dos genes e vias candidatas e metanálises, hipotetizamos que os genes e vias candidatas à seleção na Amazônia tiveram como pressão seletiva o estilo de vida caçador-coletor, sendo a procura de novidades, o período de escassez alimentícia e a resposta a de patógenos presentes neste ambiente os principais fatores seletivos. Neste contexto, alguns dos traços que outrora foram importantes para a subsistência destas populações podem hoje apresentar um efeito adverso dada a mudança do estilo de vida local, como a introdução de alimentos industrializados e o modo de vida sedentário. Por exemplo, as rotas para o sistema de recompensa do cérebro, que poderiam favorecer maior disposição à procura de novidades (exposição ao risco), podem hoje estar relacionados com os elevados índices de alcoolismo nestas populações. De forma similar, genes que possibilitaram a subsistência em períodos de escassez alimentícia dado à sazonalidade das florestas tropicais, podem estar relacionados com o aumento da obesidade e doenças relacionadas à síndrome metabólica. Em relação ao sistema imunológico, detectamos genes relacionados à seleção de linfócitos T, e genes relacionados à resposta sorológica ao agente transmissor da doença de Chagas, *Trypanosoma cruzi* (e.g. *PPP3CA* e *DYNC111*). Dada à pequena incidência da doença de Chagas nas populações amazônicas, quando comparado à populações andinas, é possível que os nativos amazônicos tenham desenvolvido um mecanismo de defesa ao protozoário *Trypanosoma cruzi*. Torna-se interessante, portanto, a realização de estudos funcionais com os genes aqui identificados para avaliar a taxa de infecção do protozoário.

Nas populações andinas, identificamos cinco genes candidatos principais – *DUOX2*, *SP100*, *CLC*, *TMEM38B*, *AS3MT* – pelo método PBS, corroborados por sua vez pelo método XP-EHH. Nossos resultados apontaram um mecanismo complexo de adaptação

nos nativos andinos, que envolveu os genes *CLC*, *SP100* e *DUOX2*, não encontrados em estudos prévios de adaptação ao ambiente andino. Considerando a diversidade de rotas evidenciadas relacionadas a adaptação exclusiva ao altiplano andino, inferimos que os nativos andinos se adaptaram a esse ambiente de maneiras diferentes e usando estratégias moleculares distintas das de outros povos que vivem em grandes altitudes.

Em conclusão, nosso trabalho identificou novos genes candidatos tanto na região da floresta amazônica quanto nas terras altas andinas, que muito provavelmente tiveram um papel essencial para adaptação local das populações nativas americanas na Amazônia e nos Andes. Os resultados desta tese abrem portas para outras pesquisas importantes, incluindo estudos funcionais, e levantam uma preocupação no que se refere à saúde pública indígena.

3 Resumo

O continente Americano foi povoado há aproximadamente 15.000 anos, e rapidamente os primeiros nativos americanos se dispersaram da América do Norte à América do Sul, passando por uma variedade de ecorregiões distintas, entre elas, a floresta tropical amazônica e as terras altas andinas. Ambas as regiões apresentam desafios para a subsistência humana. A floresta amazônica é úmida e apresenta baixa penetração solar, elevados índices de patógenos e agentes transmissores, bem como períodos de escassez de alimento humano, ao passo que as terras altas andinas são caracterizadas pela baixa concentração disponível de oxigênio, frio intenso e maior intensidade de radiação UV. Desta forma, as populações nativas nessas ecorregiões poderiam possuir variantes genéticas que favoreceram sua subsistência nestes ambientes, frente aos desafios mencionados. A fim de avaliar a influência da seleção natural nestes ambientes, utilizamos 285 indivíduos nativos sul americanos (222 terras baixas e 63 terras altas). Para detecção de seleção positiva utilizamos métodos baseados em haplótipos, como iHS e XP-EHH, ou baseados na diferenciação populacional via F_{st} , como PBS. Adicionalmente, utilizamos os resultados dos testes de seleção para aplicar testes de enriquecimento gênico, análise de eQTL in silico, anotação funcional dos genes candidatos, simulações demográficas, e realizamos uma metanálise dos resultados já publicados referentes às populações nativas de floresta tropical de diversos continentes. Como resultado, na região amazônica, identificamos genes e vias de sinalização candidatas relacionadas ao metabolismo energético, vias cardiovasculares, defesa imunológica e comportamento *Novelty seeking*. Na região andina, identificamos genes candidatos com funções essenciais ao metabolismo em situações de estresse promovida pela hipóxia, mostrando que os nativos andinos parecem apresentar vias alternativas de adaptação ao altiplano quando comparados com outras populações de altitude dos demais continentes. O presente trabalho apresenta dados inéditos relativos à adaptação dos nativos americanos a duas importantes ecorregiões no continente, evidenciando que rotas metabólicas antes importantes para a exploração e sobrevivência ao ambiente, hoje apresentam grande impacto no perfil epidemiológico dessas populações.

Palavras-chave: seleção natural, nativos americanos, adaptação local, genômica.

4 Abstract

The American continent was peopled approximately 15,000 years ago. The first Native Americans quickly spread out from North to South America, passing through various distinct ecoregions, including the Amazon rainforest and the Andean highlands. Both regions present challenges to human livelihood. The Amazon rainforest is humid with low solar penetration, high pathogenicity, and periods of scarcity of human food. On the other hand, Andean highlands are known for their low oxygen concentration, intense cold, and high UV radiation intensity. Therefore, native populations from such environments could have genetic variants that favored their subsistence in these environments. To assess the influence of natural selection in these ecoregions, we used 285 Native South American individuals (222 from lowlands and 63 from highlands). To detect positive selection, we applied either haplotype-based methods, such as iHS and XP-EHH, or methods based on population differentiation via F_{st} , such as PBS. Furthermore, we used the results of the selection tests to apply gene enrichment tests, *in silico* eQTL analysis, functional annotation of the candidate genes, demographic simulations, and we carried out a meta-analysis from already published data on native populations from tropical forests of different continents. As a result, in the Amazon region, we identified candidate genes and signaling pathways related to energy metabolism, cardiovascular pathways, immune defense, and novelty seeking behavior. In the Andean region, we have identified candidate genes with essential functions to metabolism in stress situations promoted by hypoxia, showing that Andean natives seem to have alternative ways for adapting to the altiplano when compared to other altitude populations from distinct continents. The present work reveals unprecedented data related to the adaptation of Native Americans to two leading ecoregions in the continent, showing that metabolic routes that were previously important to the environment exploration and survival, today play a big role in the epidemiological profile of these populations.

Keywords: natural selection, native americans, local adaptation, genomics.

REFERÊNCIAS BIBLIOGRÁFICAS

- AMORIM, C. E. G. et al. Genetic signature of natural selection in first Americans. *Proceedings of the National Academy of Sciences*, v. 114, n. 9, p. 2195–2199, fev. 2017. ISSN 0027-8424, 1091-6490. Disponível em: <<http://www.pnas.org/lookup/doi/10.1073/pnas.1620541114>>.
- ANTONELLI, A. et al. Amazonia is the primary source of Neotropical biodiversity. *Proceedings of the National Academy of Sciences*, v. 115, n. 23, p. 6034–6039, jun. 2018. ISSN 0027-8424, 1091-6490. Disponível em: <<http://www.pnas.org/lookup/doi/10.1073/pnas.1713819115>>.
- ASHBURNER, M. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics*, v. 25, n. 1, p. 25–29, maio 2000. ISSN 1061-4036.
- BARGHI, N.; HERMISSON, J.; SCHLÖTTERER, C. Polygenic adaptation: a unifying framework to understand positive selection. *Nature Reviews. Genetics*, jun. 2020. ISSN 1471-0064.
- BERGEY, C. M. et al. Polygenic adaptation and convergent evolution on growth and cardiac genetic pathways in African and Asian rainforest hunter-gatherers. *Proceedings of the National Academy of Sciences of the United States of America*, v. 115, n. 48, p. E11256–E11263, 2018. ISSN 1091-6490.
- BHATIA, G. et al. Estimating and interpreting FST: the impact of rare variants. *Genome Research*, v. 23, n. 9, p. 1514–1521, set. 2013. ISSN 1549-5469.
- BUNIELLO, A. et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research*, v. 47, n. D1, p. D1005–D1012, 2019. ISSN 1362-4962.
- CLIMER, S.; TEMPLETON, A. R.; ZHANG, W. Allele-specific network reveals combinatorial interaction that transcends small effects in psoriasis GWAS. *PLoS computational biology*, v. 10, n. 9, p. e1003766, set. 2014. ISSN 1553-7358.
- CLIMER, S. et al. A custom correlation coefficient (CCC) approach for fast identification of multi-SNP association patterns in genome-wide SNPs data. *Genetic Epidemiology*, v. 38, n. 7, p. 610–621, nov. 2014. ISSN 1098-2272.
- DARWIN, C. *The Origin of Species: By Means of Natural Selection or the Preservation of Favoured Races in the Struggle for Life*. London: John Murray, 1859. Disponível em: <http://darwin-online.org.uk/converted/pdf/1859_Origin_F373.pdf>.
- DARWIN, C. *The Descent of Man, and Selection in Relation to Sex*. London: John Murray, 1871. Disponível em: <http://darwin-online.org.uk/converted/pdf/1871_Descent_F937.1.pdf>.
- DARWIN, C.; WALLACE, A. On the Tendency of Species to form Varieties; and on the Perpetuation of Varieties and Species by Natural Means of Selection. *Zoological Journal of the Linnean Society*, v. 3, n. 9, p. 45–62, ago. 1858. ISSN 19459475. Disponível em:

<<https://academic.oup.com/zoolinnea/article-lookup/doi/10.1111/j.1096-3642.1858.tb02500.x>>.

FAGUNDES, N. J. R. et al. Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. *American Journal of Human Genetics*, v. 82, n. 3, p. 583–592, mar. 2008. ISSN 1537-6605.

FAIRBANKS, D. J. Mendel and Darwin: untangling a persistent enigma. *Heredity*, v. 124, n. 2, p. 263–273, fev. 2020. ISSN 1365-2540. Number: 2 Publisher: Nature Publishing Group. Disponível em: <<https://www.nature.com/articles/s41437-019-0289-9>>.

GOUY, A.; DAUB, J. T.; EXCOFFIER, L. Detecting gene subnetworks under selection in biological pathways. *Nucleic Acids Research*, v. 45, n. 16, p. e149, set. 2017. ISSN 1362-4962.

GÜNTHER, T.; COOP, G. Robust Identification of Local Adaptation from Allele Frequencies. *Genetics*, v. 195, n. 1, p. 205–220, set. 2013. ISSN 0016-6731, 1943-2631. Disponível em: <<http://www.genetics.org/lookup/doi/10.1534/genetics.113.152462>>.

HANCOCK, A. M.; RIENZO, A. D. Detecting the Genetic Signature of Natural Selection in Human Populations: Models, Methods, and Data. *Annual Review of Anthropology*, v. 37, n. 1, p. 197–217, out. 2008. ISSN 0084-6570, 1545-4290. Disponível em: <<http://www.annualreviews.org/doi/10.1146/annurev.anthro.37.081407.085141>>.

HANCOCK, A. M. et al. Human adaptations to diet, subsistence, and ecoregion are due to subtle shifts in allele frequency. *Proceedings of the National Academy of Sciences*, v. 107, n. Supplement_2, p. 8924–8930, maio 2010. ISSN 0027-8424, 1091-6490. Disponível em: <<http://www.pnas.org/cgi/doi/10.1073/pnas.0914625107>>.

HARRISON, G. F. et al. Natural selection contributed to immunological differences between hunter-gatherers and agriculturalists. *Nature Ecology & Evolution*, v. 3, n. 8, p. 1253–1264, 2019. ISSN 2397-334X.

HOFFECKER, J. F. et al. Beringia and the global dispersal of modern humans: Beringia and the Global Dispersal of Modern Humans. *Evolutionary Anthropology: Issues, News, and Reviews*, v. 25, n. 2, p. 64–78, mar. 2016. ISSN 10601538. Disponível em: <<http://doi.wiley.com/10.1002/evan.21478>>.

HSIEH, P. et al. Exome Sequencing Provides Evidence of Polygenic Adaptation to a Fat-Rich Animal Diet in Indigenous Siberian Populations. *Molecular Biology and Evolution*, v. 34, n. 11, p. 2913–2926, nov. 2017. ISSN 1537-1719.

JOBLING, M. A. et al. *Human Evolutionary Genetics*. 2. ed. New York: Garland Science, 2014. ISBN 978-0-8153-4148-2.

KANEHISA, M.; GOTO, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, v. 28, n. 1, p. 27–30, jan. 2000. ISSN 0305-1048.

KANEHISA, M. et al. New approach for understanding genome variations in KEGG. *Nucleic Acids Research*, v. 47, n. D1, p. D590–D595, 2019. ISSN 1362-4962.

KHATRI, P.; SIROTA, M.; BUTTE, A. J. Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS computational biology*, v. 8, n. 2, p. e1002375, 2012. ISSN 1553-7358.

- KIMURA, M. The rate of molecular evolution considered from the standpoint of population genetics. *Proceedings of the National Academy of Sciences of the United States of America*, v. 63, n. 4, p. 1181–1188, ago. 1969. ISSN 0027-8424.
- LEWONTIN, R. C.; KRAKAUER, J. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, v. 74, n. 1, p. 175–195, maio 1973. ISSN 0016-6731.
- LIAO, Y. et al. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Research*, v. 47, n. W1, p. W199–W205, jul. 2019. ISSN 0305-1048, 1362-4962. Disponível em: <<https://academic.oup.com/nar/article/47/W1/W199/5494758>>.
- LLAMAS, B. et al. Ancient mitochondrial DNA provides high-resolution time scale of the peopling of the Americas. *Science Advances*, v. 2, n. 4, p. e1501385, abr. 2016. ISSN 2375-2548.
- LOPEZ, M. et al. Genomic Evidence for Local Adaptation of Hunter-Gatherers to the African Rainforest. *Current biology: CB*, v. 29, n. 17, p. 2926–2935.e4, 2019. ISSN 1879-0445.
- LOSOS, J. B. et al. *The Princeton Guide to Evolution*. 1. ed. Princeton, New Jersey: Princeton University Press, 2014. ISBN 978-0-691-14977-6.
- MALÉCOT, G. *Les mathématiques de l'hérédité*. Paris: Masson & Cie, 1948.
- MENDES, M. et al. The history behind the mosaic of the Americas. *Current Opinion in Genetics & Development*, v. 62, p. 72–77, jun. 2020. ISSN 0959437X. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0959437X20301076>>.
- MORENO-MAYAR, J. V. et al. Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. *Nature*, v. 553, n. 7687, p. 203–207, jan. 2018. ISSN 0028-0836, 1476-4687. Disponível em: <<http://www.nature.com/articles/nature25173>>.
- MORENO-MAYAR, J. V. et al. Early human dispersals within the Americas. *Science (New York, N.Y.)*, v. 362, n. 6419, 2018. ISSN 1095-9203.
- NIELSEN, R. et al. Recent and ongoing selection in the human genome. *Nature Reviews Genetics*, v. 8, n. 11, p. 857–868, nov. 2007. ISSN 1471-0056, 1471-0064. Disponível em: <<http://www.nature.com/articles/nrg2187>>.
- NOVEMBRE, J.; HAN, E. Human population structure and the adaptive response to pathogen-induced selection pressures. *Philosophical Transactions of the Royal Society B: Biological Sciences*, v. 367, n. 1590, p. 878–886, mar. 2012. ISSN 0962-8436, 1471-2970. Disponível em: <<https://royalsocietypublishing.org/doi/10.1098/rstb.2011.0305>>.
- PEDERSEN, M. W. et al. Postglacial viability and colonization in North America's ice-free corridor. *Nature*, v. 537, n. 7618, p. 45–49, 2016. ISSN 1476-4687.
- POSTH, C. et al. Reconstructing the Deep Population History of Central and South America. *Cell*, v. 175, n. 5, p. 1185–1197.e22, 2018. ISSN 1097-4172.
- POTTER, B. A. et al. Current evidence allows multiple models for the peopling of the Americas. *Science Advances*, v. 4, n. 8, p. eaat5473, ago. 2018. ISSN 2375-2548. Disponível em: <<https://advances.sciencemag.org/lookup/doi/10.1126/sciadv.aat5473>>.

POUYET, F. et al. Background selection and biased gene conversion affect more than 95% of the human genome and bias demographic inferences. *eLife*, v. 7, 2018. ISSN 2050-084X.

PRITCHARD, J. K.; PICKRELL, J. K.; COOP, G. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Current biology: CB*, v. 20, n. 4, p. R208–215, fev. 2010. ISSN 1879-0445.

PRITCHARD, J. K.; RIENZO, A. D. Adaptation – not by sweeps alone. *Nature Reviews Genetics*, v. 11, n. 10, p. 665–667, out. 2010. ISSN 1471-0056, 1471-0064. Disponível em: <<http://www.nature.com/articles/nrg2880>>.

PRUGNOLLE, F.; MANICA, A.; BALLOUX, F. Geography predicts neutral genetic diversity of human populations. *Current biology: CB*, v. 15, n. 5, p. R159–160, mar. 2005. ISSN 0960-9822.

RACIMO, F.; BERG, J. J.; PICKRELL, J. K. Detecting Polygenic Adaptation in Admixture Graphs. *Genetics*, v. 208, n. 4, p. 1565–1584, abr. 2018. ISSN 0016-6731, 1943-2631. Disponível em: <<http://www.genetics.org/lookup/doi/10.1534/genetics.117.300489>>.

RAGHAVAN, M. et al. POPULATION GENETICS. Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science (New York, N.Y.)*, v. 349, n. 6250, p. aab3884, ago. 2015. ISSN 1095-9203.

REECE, J. B. et al. *Campbell Biology*. 10. ed. Glenview,: Pearson, 2013.

REES, J. S.; CASTELLANO, S.; ANDRÉS, A. M. The Genomics of Human Local Adaptation. *Trends in genetics: TIG*, v. 36, n. 6, p. 415–428, jun. 2020. ISSN 0168-9525.

REFOYO-MARTÍNEZ, A. et al. *How robust are cross-population signatures of polygenic adaptation in humans?* [S.l.], 2020. Disponível em: <<http://biorxiv.org/lookup/doi/10.1101/2020.07.13.200030>>.

REYNOLDS, A. W. et al. Comparing signals of natural selection between three Indigenous North American populations. *Proceedings of the National Academy of Sciences of the United States of America*, v. 116, n. 19, p. 9312–9317, 2019. ISSN 1091-6490.

SABETI, P. C. et al. Detecting recent positive selection in the human genome from haplotype structure. *Nature*, v. 419, n. 6909, p. 832–837, out. 2002. ISSN 0028-0836, 1476-4687. Disponível em: <<http://www.nature.com/articles/nature01140>>.

SABETI, P. C. et al. Genome-wide detection and characterization of positive selection in human populations. *Nature*, v. 449, n. 7164, p. 913–918, out. 2007. ISSN 1476-4687.

SALAS, A. The natural selection that shapes our genomes. *Forensic Science International. Genetics*, v. 39, p. 57–60, 2019. ISSN 1878-0326.

SCHEIB, C. L. et al. Ancient human parallel lineages within North America contributed to a coastal expansion. *Science*, v. 360, n. 6392, p. 1024–1027, jun. 2018. ISSN 0036-8075, 1095-9203. Disponível em: <<https://www.sciencemag.org/lookup/doi/10.1126/science.aa6851>>.

- SHRIVER, M. D. et al. The genomic distribution of population substructure in four populations using 8,525 autosomal SNPs. *Human Genomics*, v. 1, n. 4, p. 274–286, maio 2004. ISSN 1473-9542.
- SMITH, J. M.; HAIGH, J. The hitch-hiking effect of a favourable gene. *Genetical Research*, v. 23, n. 1, p. 23–35, fev. 1974. ISSN 0016-6723, 1469-5073. Disponível em: <https://www.cambridge.org/core/product/identificator/S0016672300014634/type/journal_article>.
- SOHAIL, M. et al. Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. *eLife*, v. 8, 2019. ISSN 2050-084X.
- STEPHAN, W. Selective Sweeps. *Genetics*, v. 211, n. 1, p. 5–13, jan. 2019. ISSN 0016-6731, 1943-2631. Disponível em: <<http://www.genetics.org/lookup/doi/10.1534/genetics.118.301319>>.
- SUBRAMANIAN, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, v. 102, n. 43, p. 15545–15550, out. 2005. ISSN 0027-8424.
- SUTTER, R. C. The Pre-Columbian Peopling and Population Dispersals of South America. *Journal of Archaeological Research*, mar. 2020. ISSN 1059-0161, 1573-7756. Disponível em: <<http://link.springer.com/10.1007/s10814-020-09146-w>>.
- SUZUKI, Y. Statistical methods for detecting natural selection from genomic data. *Genes & Genetic Systems*, v. 85, n. 6, p. 359–376, 2010. ISSN 1880-5779, 1341-7568. Disponível em: <<http://joi.jlc.jst.go.jp/JST.JSTAGE/ggs/85.359?from=CrossRef>>.
- TAMM, E. et al. Beringian standstill and spread of Native American founders. *PloS One*, v. 2, n. 9, p. e829, set. 2007. ISSN 1932-6203.
- TANG, K.; THORNTON, K. R.; STONEKING, M. A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS biology*, v. 5, n. 7, p. e171, jul. 2007. ISSN 1545-7885.
- TEMPLETON, A. R. *Population Genetics and Microevolutionary Theory*. 1. ed. New Jersey: John Wiley & Sons, 2006. ISBN 978-0-471-40951-9.
- The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Research*, v. 47, n. D1, p. D330–D338, 2019. ISSN 1362-4962.
- UEKI, M.; CORDELL, H. J. Improved statistics for genome-wide interaction analysis. *PLoS genetics*, v. 8, n. 4, p. e1002625, 2012. ISSN 1553-7404.
- VISSCHER, P. M. et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. *The American Journal of Human Genetics*, v. 101, n. 1, p. 5–22, jul. 2017. ISSN 00029297. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0002929717302409>>.
- VITTI, J. J.; GROSSMAN, S. R.; SABETI, P. C. Detecting Natural Selection in Genomic Data. *Annual Review of Genetics*, v. 47, n. 1, p. 97–120, nov. 2013. ISSN 0066-4197, 1545-2948. Disponível em: <<http://www.annualreviews.org/doi/10.1146/annurev-genet-111212-133526>>.

VOIGHT, B. F. et al. A map of recent positive selection in the human genome. *PLoS biology*, v. 4, n. 3, p. e72, mar. 2006. ISSN 1545-7885.

WAN, X. et al. BOOST: A fast approach to detecting gene-gene interactions in genome-wide case-control studies. *American Journal of Human Genetics*, v. 87, n. 3, p. 325–340, set. 2010. ISSN 1537-6605.

WANG, J. et al. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Research*, v. 41, n. Web Server issue, p. W77–83, jul. 2013. ISSN 1362-4962.

WATANABE, K. et al. Functional mapping and annotation of genetic associations with FUMA. *Nature Communications*, v. 8, n. 1, p. 1826, 2017. ISSN 2041-1723.

WOOLLER, M. J. et al. A new terrestrial palaeoenvironmental record from the Bering Land Bridge and context for human dispersal. *Royal Society Open Science*, v. 5, n. 6, p. 180145, jun. 2018. ISSN 2054-5703, 2054-5703. Disponível em: <<https://royalsocietypublishing.org/doi/10.1098/rsos.180145>>.

WRIGHT, S. The genetical structure of populations. *Annals of Eugenics*, v. 15, n. 4, p. 323–354, 1949.

YI, X. et al. Sequencing of 50 Human Exomes Reveals Adaptation to High Altitude. *Science*, v. 329, n. 5987, p. 75–78, jul. 2010. ISSN 0036-8075, 1095-9203. Disponível em: <<https://www.sciencemag.org/lookup/doi/10.1126/science.1190371>>.