Bárbara Domingues Bitarello

"Seleção balanceadora no genoma humano: relevância biológica e consequências deletérias"

"Balancing selection in the human genome: biological relevance and deleterious consequences"

Diogo Meyer
Orientador

São Paulo
2016

Bárbara Domingues Bitarello

"Seleção balanceadora no genoma humano: relevância biológica e consequências deletérias"

Tese apresentada ao Instituto de Biociências da Universidade de São Paulo, para a obtenção de Título de Doutor em Ciências, na Área de Biologia (Genética).

Orientador: Diogo Meyer

**São Paulo
2016**

Ficha Catalográfica

Domingues Bitarello, Bárbara

"Seleção balanceadora no genoma humano: relevância biológica e consequências deletérias".

299 páginas.

Tese (Doutorado) - Instituto de Biociências da Universidade de São Paulo. Departamento de Genética e Biologia Evolutiva.

1. Evolução Molecular;
2. Evolução Humana;
3. Seleção Balanceadora;
4. Evolução Adaptativa;
5. Genética de Populações;
6. Genômica de Populações;
7. Carga Genética

I. Universidade de São Paulo. Instituto de Biociências. Departamento de Genética e Biologia Evolutiva.

Comissão Julgadora:

Prof. Dr.

Prof. Dr.

Prof. Dr.

Prof. Dr.

Prof. Dr. Diogo Meyer

Resumo

Seleção balanceadora é um processo evolutivo que engloba diversos mecanismos: vantagem do heterozigoto, seleção dependente de frequência, pressões seletivas que variam ao longo do tempo ou do espaço, e alguns casos de pleiotropia. O estudo desses mecanismos em si foi e ainda é um tópico de grande interesse para os biólogos evolutivos, e moldou o estudo da evolução ao longo do último século. Antes de a teoria neutra ter sido proposta, acreditava-se que a seleção balanceadora fosse comum. A descoberta de que muita da diversidade genética observada podia ser explicada por evolução neutra motivou, portanto, uma melhor compreensão da seleção balanceadora como um regime seletivo capaz de manter variantes vantajosas nas populações.

O estudo da seleção balanceadora, em seus primórdios, foi restrito a organismos que podiam ser manipulados em laboratório. Com o advento de métodos que permitiam quantificar a variabilidade genética – tais como a eletroforese de proteínas, sequenciamento em pequena escala e re-sequenciamento genômico de milhares de indivíduos –, a variabilidade genética humana passou a ser ativamente estudada e interpretada. Diversos estudos buscaram por assinaturas de seleção natural – i.e., padrões de variação genômica deixadas por tais regimes seletivos – e avaliaram seu significado comparando-as com o que seria esperado sob um cenário estritamente neutro. A maior parte desses esforços foram concentrados no estudo da seleção positiva, tida como o principal mecanismo responsável pela evolução adaptativa.

Poucos estudos buscaram assinaturas de seleção balanceadora no genoma humano. Isso se deve em parte à escassez de métodos com alto poder para detectar tais assinaturas. Adicionalmente, estudos prévios não analisaram dados em escala genômica, ou se concentraram principalmente nas regiões codificadoras de proteínas. Aqui, nós descrevemos um método simples e com alto poder para detectar assinaturas de seleção balanceadora. Em humanos, esse método supera outros comumente usados para a detecção de tais assinaturas e, em teoria, poderia ser usado para detectá-las em outras espécies, desde que seu poder seja avaliado caso-a-caso através de simulações neutras. Nosso método ("*Non-Central Deviation*", NCD) é apresentado em duas versões: *NCD2*, que requer informação acerca dos polimorfismos da espécie analisada e das substituições entre essa espécie e um grupo externo, e *NCD1*, que requer apenas informação acerca dos polimorfismos da espécie analisada. Embora em humanos *NCD2* supere *NCD1*, este último pode ser utilizado para espécies para as quais não haja informação de um grupo externo.

Quando aplicamos *NCD2* a dados humanos, usando chimpanzé como grupo externo, encontramos mais de 200 genes codificadores de proteínas com forte assinatura de seleção balanceadora, dos quais apenas 1/3 tinha evidência prévia de seleção balanceadora. Encontramos também um enriquecimento para diversas categorias de ontologia gênica, das quais cerca da metade é relacionada à imunidade. Verificamos que dentre os genes com evidências de seleção balanceadora há um excesso de casos de expressão preferencial em tecidos tais como “adrenal” e “pulmão”, e também um excesso de genes com expressão mono-alélica. No geral, vimos que as regiões selecionadas no genoma humano incluem tanto sítios codificadores quanto regulatórios. Não encontramos um excesso de assinaturas de seleção balanceadora em regiões regulatórias, ao contrário do que reportaram outros estudos. Finalmente, encontramos um excesso de polimorfismos não-sinônimos em relação aos sinônimos nos genes selecionados.

Tendo documentado a ocorrência de seleção balanceadora no genoma humano e identificado genes que foram potencialmente alvos deste regime seletivo, nós investigamos as consequências evolutivas desse processo. Nós partimos da hipótese que a seleção balanceadora sobre um sítio reduz a eficiência com a qual a seleção purificadora elimina variantes deletérias em sítios vizinhos. Esse processo é uma consequência do quanto a seleção sobre um loco afeta, através de ligação genética, as frequências de sítios não-neutros adjacentes. Testamos essa hipótese examinando se os genes sob seleção balanceadora apresentam um excesso de variantes deletérias em relação a expectativas derivadas a partir do restante do genoma. Usando três diferentes métricas para determinadas se e/ou o quão deletéria é uma dada variante, identificamos um excesso de variantes deletérias dentro dos genes sob seleção balanceadora, e mostramos que tal padrão não pode ser atribuído a efeitos confundidores. Esse achado mostra que, juntamente com os benefícios associados à variação adaptativa, a seleção balanceadora aumenta o fardo de mutações deletérias no genoma humano.

De forma geral, nossos achados sugerem que a seleção balanceadora provavelmente mantém variantes genéticas envolvidas em uma miríade de processos biológicos além da imunidade e que ela foi mais comum no genoma humano do que se acreditava anteriormente, afetando entre 1-8% dos genes codificadores de proteínas, bem como diversas regiões não-codificadoras. Adicionalmente, a seleção balanceadora parece ser importante para a evolução humana não apenas por seu efeito sobre a aptidão, mas também por ter sido uma importante força capaz de moldar a diversidade genética observada atualmente em humanos e a susceptibilidade a doenças.

Palavras-chave: evolução molecular, evolução humana, seleção balanceadora, evolução adaptativa, genética de populações, genômica de populações, carga genética

Abstract

Balancing selection is an evolutionary process that encompasses several mechanisms: heterozygote advantage, negative frequency dependent selection, selective pressure that fluctuates in time or in space, and some cases of pleiotropy. The study of these mechanisms *per se* has been and still is a topic of great interest for evolutionary biologists, and has shaped the study of evolution throughout the last century. Before the proposition of the neutral theory of molecular evolution, it was believed that balancing selection was pervasive. The realization that much of the observed genetic diversity could be explained by neutral evolution thus motivated a better understanding of balancing selection as a selective regime capable of maintaining adaptive variants in populations.

The study of balancing selection, in its early stages, was restricted to organisms that could be manipulated in the laboratory. With the advent of methods that allowed quantification of genetic variation – such as protein electrophoresis, small scale sequencing and genome-wide re-sequencing of thousands of individuals – human variation started to be actively studied and interpreted. Several studies have looked for signatures of natural selection – i.e., patterns of genomic variation that selective regimes leave in the genome – and evaluated their significance by comparing them to what would be expected under a strictly neutral scenario. Most of these efforts focused on the study of positive selection, thought of as the prime mechanism responsible for adaptive evolution.

Only a few studies looked for signatures of balancing selection in the human genome. This is partially due to the paucity of powerful methods to detect its signatures. Moreover, previous studies either did not analyze data on genomic scale or focused primarily on protein-coding regions. Here, we describe a powerful and simple method to detect signatures of balancing selection. In humans, it outperforms other methods commonly used to detect such signatures and could in theory be used for other species, provided that its power is evaluated for each species through neutral simulations. Our method ("Non-Central Deviation", NCD) has two versions: *NCD2*, which requires polymorphism information on the ingroup species, as well as divergence information between the ingroup and an outgroup species, and *NCD1*, which only requires the ingroup information. Although *NCD2* is more powerful for humans, *NCD1* can be used for species that lack information from an outgroup.

When applying *NCD2* to human data, using chimpanzee as the outgroup, we found more than 200 protein-coding regions with strong signatures of balancing selection, only 1/3 of which had prior evidence for balancing selection. There was also an enrichment for several gene ontology categories, approximately half of which are related to immunity. We also found that among genes with evidence for balancing selection there was an excess of cases of preferential expression in specific tissues, such as "adrenal" and "lung", and an excess of genes with mono-allelic expression. Overall, we found that selected regions of the genome include both coding and regulatory sites. We failed to find a marked excess of balancing selection in regulatory regions, as reported in previous studies. Finally, we found an excess of nonsynonymous versus synonymous polymorphisms within the selected genes.

Having documented the occurrence of balancing selection in the human genome and identified genes which were potential targets of this selective regime, we next investigated evolutionary consequences of this process. We hypothesized that balancing selection acting on a site reduces the efficiency with which purifying selection purges deleterious variants at nearby sites. This process is a consequence of how the dynamics of selection at one locus, mediated by linkage, can interfere with the frequencies of adjacent non-neutral sites. We tested this hypothesis by examining if the genes under balancing selection show an excess of deleterious variants with respect to expectations derived from the remainder of the genome. Using three different metrics to determine deleteriousness, we identified a significant excess of deleterious variants within balanced genes, and we show that this pattern cannot be attributed to confounding factors. This finding shows that together with the benefits associated with adaptive variation, balancing selection is increasing the burden of deleterious mutations in the human genome.

Overall, our findings suggest that balancing selection likely maintains variation in a myriad of biological processes other than immunity and that it has been more common in the human genome than previously thought, affecting between 1-8% of human protein-coding genes, as well as a number of non-protein coding regions. Moreover, balancing selection appears to be important to human evolution not only because of its influence on fitness, but also because it has been an important force shaping current human genetic diversity and susceptibility to disease.

Keywords: molecular evolution, human evolution, balancing selection, adaptive evolution, population genetics, population genomics, genetic load

Sumário

Prólogo	1
Introdução Geral	4
Seleção Balanceadora: conceito, mecanismos e importância	4
Por que estudar os mecanismos de manutenção da variabilidade genética?	4
Teoria neutra da evolução molecular	8
Mecanismos de manutenção de diversidade adaptativa	12
Evolução adaptativa no genoma humano	25
Assinaturas de seleção balanceadora	26
Seleção balanceadora no genoma humano	33
Carga genética induzida por seleção balanceadora	37
Carga genética	39
Relevância, Questões & Hipóteses	46
Relevância	46
Questões & Hipóteses	49
Bibliografia	52
1 Buscando alvos de seleção balanceadora no genoma humano	61
Considerações Iniciais	61
Introduction	63
Results	66
NCD Method	66
Power of the NCD statistics to detect LTBS	70
Identifying signatures of LTBS in the human genome	78
Reliability of significant and outlier windows	79

Non-random distribution across the genome	81
The biological pathways influenced by LTBS	81
Overlap of significant windows across populations	83
The putative function of balanced SNPs	83
The top candidate genes	85
Discussion	93
NCD Method	93
Pervasiveness of LTBS in the human genome	93
Protein-coding and intergenic targets	94
The frequency of the balanced allele(s)	95
The candidate genes	95
Conclusions	97
Materials and Methods	98
Simulations	98
Power analyses	99
Human population genetic data and filtering	100
Identifying signatures of LTBS	101
Enrichment Analyses	104
References	107
Supplementary Text	114
S1 Text: Additional analyses for significant and outlier windows and genes	114
S2 Text: A set of significant genes	116
S3 Text: Manual verification of reliability of SNPs contained in four of the outlier genes	116
Supplementary Tables	120
Supplementary Figures	144
2 Acúmulo de mutações deletérias em genes que foram alvos de seleção balanceadora de longo prazo em humanos	165
Considerações Iniciais	165
Introduction	167
Methods	173
Population datasets	173

Targets of balancing selection	174
Annotation	175
Quantifying genetic load	176
Re-sampling control SNPs	181
Results	185
The site frequency spectrum of balanced genes	185
Measures of deleteriousness correlate negatively with allelic frequency	186
Extreme values for HLA SNPs	193
Increased nonsynonymous to synonymous SNPs in balanced genes	195
Increased proportion of damaging to synonymous SNPs in balanced genes	197
Increased C-scores in balanced genes	198
Discussion and Conclusions	203
Increased genetic load in balanced genes	203
The challenges of quantifying genetic load	205
Sheltered load and hitch-hiking	207
References	209
Considerações Finais e Perspectivas	214
Seleção balanceadora no genoma humano	214
Desenvolvimento e avaliação de um novo método para a detecção de assinatura de seleção balanceadora	214
Prevalência de SBLP no genoma humano	216
Partilhamento entre continentes	218
Características das regiões candidatas	219
Variação deletéria em regiões e genes com assinaturas de SBLP	225
Perspectivas	228
Conciliando assinaturas de seleção e fenótipos	228
Potencial das estatísticas NCD em futuros estudos	230
Bibliografia	232
Apêndices	234
Apêndice A.1.	235

Apêndice A.2.	247
Apêndice A.3.	261
Apêndice A.4.	277

Prólogo

Existem diversas fontes de evidência da ação da seleção natural no genoma humano. A seleção natural pode ser direcional, aumentando ou diminuindo a frequência de variantes vantajosas ou deletérias (seleção positiva ou negativa, respectivamente) ou balanceadora. A seleção positiva vem sendo amplamente investigada há pelo menos duas décadas, sob a forma de “*scans* genômicos” e é vista como o mecanismo principal da evolução adaptativa. Estima-se que entre 2-14% do genoma humano foram alvo desse regime seletivo, em diversas escalas de tempo.

Diferentemente do que ocorre para a seleção positiva – a qual diminui a diversidade genética – a seleção balanceadora mantém a diversidade genética nas populações e, apesar de sua relevância, poucos estudos até hoje exploraram seus alvos no genoma humano.

Aqui, eu me propus a explorar a importância da seleção balanceadora ao longo evolução humana: seus alvos no genoma, as propriedades destes alvos e as possíveis consequências deletérias causadas na vizinhança de polimorfismos balanceados.

Primeiramente, busquei fazer um levantamento o mais completo possível dos alvos de seleção balanceadora no genoma humano. Para isso, desenvolvemos uma nova ferramenta estatística (NCD, *Non-Central Deviation*), otimizada para a detecção de assinaturas de seleção balanceadora em humanos. Essa ferramenta foi utilizada em dados genômicos de quatro populações a fim de mapear a atuação desse regime seletivo em humanos: seus alvos, suas propriedades biológicas e genômicas, e as diferenças entre diferentes populações e continentes (Capítulo 1).

Em segundo lugar, buscamos testar a hipótese de que a seleção balanceadora, mantendo polimorfismos por longas escalas de tempo (milhões de anos), teria um efeito deletério sobre sítios próximos ao(s) sítio(s) selecionados. Essa hipótese é oriunda de uma ampla literatura acerca do acúmulo de mutações deletérias em regiões vizinhas a alvos de seleção positiva, de carga genética em humanos, e também do fato de que muitos genes sob seleção balanceadora parecem estar associados a doenças complexas. Além disso, há evidências de tal acúmulo ocorre ao redor dos genes HLA que foram alvos de seleção balanceadora. A fim de explorar essas questões, o primeiro capítulo da tese trata da detecção dos alvos da seleção balanceadora no genoma humano, e o segundo no estudo dos efeitos da seleção balanceadora sobre regiões vizinhas do genoma.

Além disso, apresento uma introdução geral aos temas dos dois capítulos, e uma discussão final sobre as implicações dos achados dos dois estudos. Assim, a tese está dividida em:

- **Introdução Geral:** uma introdução acerca **(a)** do regime seletivo conhecido como seleção balanceadora – definição, importância histórica para a genética de populações, mecanismos através dos quais atua, propriedades do regime seletivo e importância evolutiva – e **(b)** do conceito de carga genética e dos efeitos deletérios que a seleção natural deixa em regiões ligadas geneticamente aos alvos de seleção.
- **Primeiro capítulo:** em que apresento um novo método estatístico (“NCD”), focado especificamente na detecção de assinaturas genômicas deixadas por regimes de seleção balanceadora que perduram por milhões de anos, bem como os resultados de um *scan* genômico feito com o método descrito usando dados reais de genética de populações humanas.

-
- **Segundo capítulo:** em que apresento uma investigação acerca do acúmulo de mutações deletérias em regiões que foram alvo de seleção balanceadora em humanos, conforme detectadas no *scan* genômico apresentado no primeiro capítulo.
 - **Considerações Finais e Perspectivas:** em que discuto os principais achados dos dois capítulos principais da tese, bem como as perspectivas para trabalhos futuros que podem decorrer de nossas contribuições.

No anexo incluí publicações que resultaram de colaborações realizadas no período do doutorado. Três delas (Apêndices A.1, A.2, A.4) trazem como tema principal os genes HLA – exemplos clássicos de alvos de seleção balanceadora –, sendo dois deles focados nas assinaturas de seleção balanceadora desses genes (A.2 e A.4), e o outro focado em problemas metodológicos com sequenciamento de nova geração para essa região do genoma (A.1). Finalmente, o trabalho no anexo A.3 trata da evolução adaptativa de forma mais ampla, com foco em testes de seleção positiva.

Introdução Geral

Seleção Balanceadora: conceito, mecanismos e importância

ENTENDER como surgiram e como foram mantidas as variações fenotípicas observadas nas populações naturais, bem como suas possíveis implicações funcionais, é um dos objetivos centrais da genética de populações (Kimura, 1983; Dobzhansky, 1937). Não é surpreendente, portanto, que a elucidação acerca dos mecanismos através dos quais a variabilidade genética ao nível molecular é mantida nas populações seja uma "grande questão": talvez o problema mais importante do campo da biologia evolutiva e da genética de populações (Kimura, 1983), marcado por acirrados debates (Figura 1).

Por que estudar os mecanismos de manutenção da variabilidade genética?

O "vigor do híbrido", ou heterose, é observado há séculos. Mendel percebeu que as ervilhas híbridas de seus estudos tinham altura média maior do que as das linhagens parentais (Crow, 1987) e Darwin escreveu um livro inteiro sobre

Figura 1: **Linha do tempo do estudo da seleção balanceadora** Adaptada a partir de Gloss e Whiteman (2016) resumindo as principais contribuições teóricas e empíricas para a compreensão da importância da seleção balanceadora para a manutenção de variação genética. SBLP, seleção balanceadora de longo prazo (ver texto).

vigor do híbrido em plantas (Darwin, 1876)¹. Apesar de ter sido observado frequentemente por criadores de plantas e animais, esse fenômeno só pôde ser corretamente interpretado após a re-descoberta das Leis de Mendel no início do século 20 (revisado em Crow, 1987), quando estabeleceu-se de forma definitiva que a variação genética é um dos fatores determinantes da variação fenotípica.

Em 1922, Fisher menciona que, embora o vigor do híbrido fosse um fato, não era clara a razão biológica pela qual um heterozigoto qualquer seria mais apto que seu homozigoto correspondente. Fisher foi o primeiro a demonstrar como um equilíbrio de frequências entre dois alelos pode ser mantido em um loco sob vantagem de heterozigoto (ver página 13). Ele propôs, ainda, que tal fenômeno deveria ser comum na natureza, e capaz de explicar tanto o vigor do híbrido quanto os efeitos deletérios às vezes observados em animais domesticados submetidos a endogamia.

Na medida em que começaram a surgir, desde a primeira metade do século 20, métodos capazes de acessar a variação genética diretamente – uma propriedade que, começava-se então a perceber, era aparentemente ubíqua em populações naturais –, começou também a surgir um interesse crescente em se explicar os padrões de variação genética (revisado em Bamshad e Wooding, 2003; Gloss e Whiteman, 2016). Mesmo com métodos que para o geneticista de populações contemporâneo parecem bastante limitados (pois sempre forneciam subestimativas dos níveis reais de variabilidade genética das populações), variações genéticas já eram observadas desde meados do século 20, e sua persistência foi atribuída à ação da seleção balanceadora (Dobzhansky, 1937).

O termo “seleção balanceadora” agregava todo e qualquer processo evolu-

¹Nesse livro, Darwin faz experimentos de auto-polinização e polinização cruzada em mais de 60 espécies de plantas e conclui que (na maior parte dos casos) a performance da prole resultante de auto-polinização é, para diversos traços, inferior.

tivo que, de forma adaptativa, mantinha variação genética nas populações. Por exemplo, Dobzhansky (1937) observou polimorfismos na orientação de longos trechos de DNA em cromossomos de *Drosophila* – as chamadas “inversões cromossômicas” usando técnicas de coloração de cromossomos, muito antes de o sequenciamento de DNA ser possível – e atribuiu à seleção balanceadora a manutenção de tais polimorfismos na natureza (Figura 1).

Nessa época, no início do século 20, o interesse pela seleção balanceadora tinha múltiplas origens. Criadores de animais e plantas queriam maximizar a produtividade/performance, e o vigor do híbrido era observado, porém não totalmente compreendido. Dobzhansky, assim como Muller, estava preocupado com o potencial da espécie para continuar evoluindo. Entretanto, Muller tinha outra preocupação mais “urgente”: o impacto do aumento da taxa de mutação causada por radiação nas gerações futuras (revisado em Crow, 1987). Essas três preocupações giravam em torno do quanto a variabilidade genética na população depende da sobredominância², i.e. em *loci* nos quais o fenótipo do heterozigoto está além dos limites dos fenótipos dos dois homozigotos correspondentes³.

Nos anos 60, Lewontin e Hubby (1966) revelaram à comunidade científica que os níveis de variação genética em alozimas de *Drosophila* eram muito mais altos do que o que se estimava ser a variabilidade genética nas populações até então analisadas (Figura 1). Suas descobertas foram baseadas em um novo método de detecção de polimorfismos – a eletroforese de proteínas –, que permitia quantificar ainda mais variações do que as técnicas de coloração de cromosso-

²Esse termo, usado frequentemente como sinônimo para vantagem do heterozigoto, foi empregado originalmente para explicar a heterose em plantas (Hedrick, 2012).

³Essa é uma definição mais antiga de vantagem do heterozigoto (Crow, 1987). Outras serão mencionadas ao longo do texto.

mos utilizadas por Dobzhansky uma década antes. A seleção balanceadora – e, particularmente, a seleção que varia ao longo do tempo – se consagrou então como uma explicação bastante popular para a manutenção de polimorfismos na natureza (revisado em Bamshad e Wooding, 2003; Gloss e Whiteman, 2016).

É importante mencionar que o mecanismo principal de seleção balanceadora invocado para explicar os níveis de variação em *Drosophila* era a heterogeneidade espacial nas pressões seletivas (Levene, 1953), decorrente do fato de tais espécies ocuparem habitats diversos: uma determinada variante não seria a mais adaptativa em todos os habitats, assim levando à manutenção de polimorfismos na população (Figura 1). Mais tarde, Dempster (1955) mostrou que pressões seletivas que variam ao longo do tempo também podem manter variações genéticas (revisado em Gloss e Whiteman, 2016).

Nas décadas seguintes, na medida em que modelos matemáticos foram sendo desenvolvidos no sentido de descrever tais regimes seletivos, evidências empíricas de sua ocorrência na natureza começaram a surgir (i.e, Allison, 1954). Nesse período, acreditava-se amplamente em uma teoria neo-Darwinista bastante “selecionista”, i.e, focada em seleção natural como o principal mecanismo capaz de alterar frequências alélicas e fenotípicas: as variantes seriam em sua grande maioria deletérias, e uma proporção menor seria vantajosa (Figura 2).

Teoria neutra da evolução molecular

Com a sensibilidade cada vez maior dos métodos moleculares em detectar os níveis de variabilidade (i.e, Lewontin e Hubby, 1966), constatou-se a abundância de polimorfismos⁴ em populações naturais, o que levou a um grande desen-

⁴A presença de variantes fenotípicas discretas em uma população é chamada de polimorfismo. Os polimorfismos “visíveis” são, contudo, uma subestimativa da diversidade genética subjacente (Charlesworth e Charlesworth, 2010). Ao longo do texto, polimorfismos são diferen-

Figura 2: Modelos selecionista, neutro e quase neutro de evolução molecular. Figura adaptada a partir de Bromham e Penny (2003) e Bernardi (2007). Em 1859, Darwin publica o livro "A origem das espécies", no qual descreve suas ideias sobre seleção natural. Darwin acredita que possam haver mutações neutras, porém a maioria das mutações são deletérias e poucas são vantajosas. Na primeira metade do século 20 a seleção natural é conciliada com as bases do mecanismo molecular de herança (Neo-Darwinismo). Neste período, o foco selecionista aborda apenas mutações deletérias e vantajosas. Em 1968 Kimura publica a primeira versão do modelo neutro de evolução molecular (e uma atualização importante em 1983) e em 1973 Ohta propõe o modelo quase neutro de evolução molecular. Esses dois últimos autores mostram que a deriva genética pode manter variantes neutras ou quase neutras nas populações. Ver também Figura 1.

volvimento de modelos de seleção balanceadora (Gloss e Whiteman, 2016). A Figura 1 resume muitas dessas contribuições.

Entretanto, esse entusiasmo foi contrabalanceado pela então recente “teoria neutra da evolução molecular”. Em seu livro divisor de águas para o campo da biologia evolutiva⁵, Kimura apresenta à comunidade científica a ideia que a principal causa das mudanças evolutivas no nível molecular – i.e, mudanças

tes alelos alelos existentes na população para um dado loco.

⁵“The Neutral Theory of Molecular Evolution (1983)” é a versão consultada aqui e amplamente usada em genética de populações, embora a teoria tenha sido proposta pela primeira vez antes (Kimura, 1968).

no material genético *per se* – não é a seleção positiva darwiniana, mas a fixação aleatória de variantes neutras ou quase neutras (Figura 2).

Desde a primeira proposição da teoria neutra em 1968, Kimura enfrentou muitas críticas, em grande parte devido ao fato de a biologia evolutiva ter sido dominada por mais de meio século pela visão darwinista de que organismos se tornam progressivamente adaptados a seus ambientes pelo acúmulo de variantes benéficas (Figura 2). Além disso, a teoria recebeu críticas de cunho técnico tais como a dificuldade de conciliar suas previsões com alguns aspectos importantes dos dados gerados: variância elevada em taxas de evolução molecular, viés no uso de códons, evidência de seleção em estudo de genes específicos, entre outros. Essas críticas forçaram mudanças na teoria, que foi sendo revisada ao longo do tempo (Kimura, 1968; Kimura, 1983; Kimura, 1991). Um outro desdobramento interessante foi a proposição da teoria “quase neutra” que, ao contrário da teoria neutra original, que tratava apenas de polimorfismos neutros e os muito deletérios, incorporou também os polimorfismos fracamente deletérios (Ohta, 1973) e os benéficos (Ohta, 1995; Ohta e Gillespie, 1996; Figura 2).

Segundo a teoria neutra, variantes neutras, introduzidas por mutações, poderiam subir de frequência estocasticamente. Kimura (1968,1983) apontou que, para contribuir para a adaptação, as mutações precisam ser mais do que benéficas: elas precisam escapar a perda por deriva, especialmente quando raras. Levando ambos os fatores em conta, Kimura concluiu que as mutações de efeito intermediário são as que mais provavelmente contribuem para a adaptação (revisado em Orr, 2005). Adicionalmente, concluiu que a maior parte da variabilidade molecular intraespecífica, como por exemplo aquela manifestada pelos polimorfismos de proteínas, é essencialmente neutra, de forma que a maior

parte dos alelos polimórficos são mantidos na espécie devido a um balanço entre mutação e extinção aleatória de alelos (Kimura, 1983).

Ao longo dos anos 60 e 70, por influência das ideias de Kimura, os geneticistas evolutivos ficaram cada vez mais convencidos de que muita – se não a maior parte – da evolução molecular reflete a fixação de mutações neutras ou quase neutras, e não benéficas. A teoria parecia ser capaz de explicar, através de processos estocásticos, a maior parte da variação observada dentro de populações. Nesse período, o estudo teórico dos mecanismos de evolução adaptativa – seleção positiva e balanceadora – diminuiu consideravelmente (Orr, 2005) (Figuras 1 e 2).

Contudo, é interessante observar que a teoria neutra não se opõe à noção de que a evolução de forma e função possam ser guiadas por seleção darwiniana, mas destaca um outro aspecto do processo evolutivo ao enfatizar o papel crucial que as pressões mutacionais e a deriva genética possuem no nível molecular. Kimura (1983) define a teoria neutra como:

“(...)the theory that at the molecular level evolutionary changes and polymorphisms are mainly due to mutations that are nearly enough neutral with respect to natural selection that their behavior and fate are mainly determined by mutation and random drift.” (Kimura 1983; primeiro capítulo).

A teoria olha os polimorfismos ao nível da proteína e do DNA como fases transitórias da evolução molecular e rejeita a noção de que a maior parte desses polimorfismos seja adaptativo e mantido na espécie devido a alguma forma de seleção balanceadora (Figura 2). A teoria neutra prevê, portanto, que a maior parte das variantes genéticas são neutras; ela não rejeita, por outro lado, que as

variantes genéticas *funcionais* - aquelas que afetam os fenótipos, e que dentro de seu modelo, representam uma minoria da variação existente na natureza - poderiam ser mantidas por seleção balanceadora (Kimura, 1983; Gloss e Whitman, 2016).

Com esse histórico – não exaustivo – sobre o tema da manutenção de diversidade nas populações, busquei apresentar os conceitos e distinção entre *polimorfismos neutros*, mantidos com certa probabilidade simplesmente pela combinação dos efeitos de mutação, deriva, “efeito carona”⁶ e migração, e *polimorfismos adaptativos*, que são mantidos por um ou mais mecanismos de seleção balanceadora⁷. A seguir, detalharei cada um desses mecanismos.

Mecanismos de manutenção de diversidade adaptativa

Dentro do pensamento evolutivo clássico não se previa um nível de variação que implicaria na existência de muitos casos de polimorfismos balanceados (Figura 1), i.e, aqueles mantidos em frequências intermediárias. Essa visão foi contestada pela descoberta de polimorfismos balanceados na natureza. Tal descoberta foi uma importante contribuição que ocorreu nos primórdios da genética de populações e que ajudou muito para uma melhor compreensão da evolução (Charlesworth e Charlesworth, 2010).

Níveis de variação genética, sabe-se hoje, são influenciados por fatores demográficos, tais como flutuações no tamanho populacional, estruturação, mis-

⁶*Genetic hitch-hiking*, processo através do qual mutações neutras – ou, em alguns casos, deletérias – mudam de frequência em uma população devido ao efeito de ligação genética com uma mutação selecionada (revisado em Cutter e Payseur, 2013). Esse tópico será abordado em maior detalhe na seção “Carga genética induzida por seleção balanceadora” e também no Capítulo 2.

⁷Embora a seleção positiva também atue sobre as mutações vantajosas ou adaptativas, ela tende a fixar tais variantes vantajosas na população, e portanto reduz, em vez de manter, a variabilidade genética.

cigenação e migração (Tishkoff e Williams, 2002)⁸. Por outro lado, padrões de diversidade variam ao longo do genoma por diversos motivos, incluindo taxas de mutação, taxas de recombinação, conversão gênica⁹ e processos seletivos (Tishkoff e Williams, 2002). A classe de processos seletivos que levam a um aumento de diversidade genética vantajosa é conhecida como “seleção balanceadora” (Andrés, 2011; Key et al., 2014). Refere-se a tais polimorfismos como “polimorfismos balanceados” (Charlesworth e Charlesworth, 2010).

Seleção balanceadora é um termo que engloba diversos mecanismos, os quais serão discutidos a seguir.

Aptidões constantes

A modelagem de regimes de seleção balanceadora frequentemente é feita usando-se as aptidões dos genótipos sem que se especifique as causas subjacentes para as diferenças de fenótipo¹⁰. Embora seja uma estratégia útil, e na prática muitas das assinaturas genômicas deixadas pelos diferentes mecanismos de seleção balanceadora sejam as mesmas, é importante buscar entender quais aspectos biológicos de um organismo são capazes de determinar sua aptidão (Charlesworth e Charlesworth, 2010).

O modelo mais básico de seleção natural assume aptidões constantes dos genótipos (Figura 3A). Para um sistema diploide e bi-alélico, as aptidões dos genótipos A_1A_1 , A_1A_2 e A_2A_2 são dadas por w_{11} , w_{12} , e w_{22} , respectivamente. Em sistemas diploides, é comum utilizar o conceito de aptidões marginais dos

⁸Entretanto, por haver estocasticidade, em média a demografia afeta o genoma como um todo, e não apenas regiões específicas.

⁹Um tipo específico de recombinação, que resulta em uma troca não-recíproca de material genético, em que uma fita de DNA é usada para modificar a sequência de outra (Tishkoff e Williams, 2002).

¹⁰Por exemplo, ver a seção de métodos do Capítulo 1.

alelos A_1 e A_2 para nos referirmos a uma média da aptidão de cada alelo considerando todos os genótipos em que ele aparece (i.e, homo e heterozigoto). No cenário que estamos tratando aqui, embora w_{11} , w_{12} e w_{22} não mudem ao longo do tempo, as aptidões marginais continuam dependendo das frequências alélicas. Ou seja, é possível que as duas aptidões marginais se igualem, e que as frequências alélicas parem de mudar, atingindo um *equilíbrio estável de frequências*. Chamamos a frequência de cada alelo, no equilíbrio, de *frequência de equilíbrio*¹¹. Interessantemente, os valores absolutos das aptidões dos genótipos são irrelevantes: apenas os valores relativos dos genótipos entram nas equações de seleção. Pode-se, portanto, definir um genótipo (e.g. o heterozigoto) como sendo o "padrão", e expressar as aptidões dos outros genótipos em relação a este, como apresentado a seguir (Charlesworth e Charlesworth, 2010).

Vantagem do heterozigoto Seguindo o modelo anterior, e usando frequências relativas dos genótipos, um cenário de vantagem do heterozigoto contendo dois alelos A_1 e A_2 com frequências p e q pode ser modelado da seguinte forma¹²:

$$\begin{array}{ccc} A_1A_1 & A_2A_2 & A_1A_2 \\ 1 - t & 1 & 1 - s \end{array}$$

, onde t e s são os coeficientes seletivos dos alelos A_1 e A_2 , respectivamente.

Quando as duas aptidões marginais, de A_1 e A_2 , são iguais ($qt = ps$), tem-se um *equilíbrio polimórfico*, e a *frequência de equilíbrio* é dada por:

$$p_{eq} = \frac{t}{t+s}; q_{eq} = 1 - p_{eq}$$

¹¹No caso de locos bi-alélicos, a frequência de equilíbrio pode ser definida como a frequência do alelo menos frequente.

¹²A primeira demonstração e discussão de como um polimorfismo pode ser mantido por seleção, de forma bastante semelhante à apresentada aqui, foi feita no trabalho intitulado "On the dominance ratio", de Fisher (1922). Ver a Figura 1.

Figura 3: Uma possível estratégia para a identificação de instâncias de seleção balanceadora consiste em medir diferenças de aptidão entre classes genótípicas. Em (A), temos um cenário de sobredominância ou vantagem do heterozigoto, onde as aptidões dos genótipos são constantes mas a do heterozigoto é sempre mais alta que a de ambos os homozigotos. Nesse exemplo, os homozigotos têm aptidões distintas (sobredominância assimétrica). Em (B), temos um cenário de aptidões variáveis ao longo do espaço (por exemplo, diferentes habitats ocupados por uma espécie) ou no tempo (quando um genótipo homozigoto tem maior aptidão em um dado momento, e reduz em gerações subsequentes). Figura adaptada de Key et al. (2014).

Através de vantagem do heterozigoto (também comumente chamada de sobredominância), polimorfismos podem ser mantidos em uma população devido à maior aptidão do genótipo heterozigoto em relação aos dois genótipos homozigotos, o que leva a um balanço de frequências entre as duas variantes (Andrés, 2011; Fijarczyk e Babik, 2015; Key et al., 2014). A frequência de equilíbrio será 0.5 quando $s = t$, um cenário improvável, conhecido como sobredominância simétrica (Figura 4A), e $\neq 0.5$ quando $s \neq t$ (Figura 4B). Desde que a condição $qt = ps$ seja atendida, a frequência de equilíbrio é atingida independentemente de sua frequência inicial (Figura 4), i.e., o equilíbrio é *estável*. No Capítulo 1, propomos uma estatística sumária que utiliza muitas dessas propriedades

para detectar assinaturas de seleção balanceadora em dados genômicos (ver, por exemplo, Figuras 1 e 6 no Capítulo 1).

Um desdobramento interessante do modelo acima é que, supondo-se que um alelo A_2 , inicialmente raro, entra em uma população inicialmente fixada para A_1 (por migração ou por mutação), então a proporção de genótipos A_2A_2 produzida por reprodução aleatória (q^2) é muito baixa comparada com a de heterozigotos ¹³. Como genótipos A_2A_2 são inicialmente muito raros, a única condição compatível com um aumento de frequência de A_2 é que A_1A_2 tenha aptidão mais alta que A_1A_1 , *mesmo que* indivíduos A_2A_2 tenham aptidão muito reduzida. Assim, A_2 só vai aumentar de frequência até um certo ponto, uma vez que A_2A_2 é deletério.

Com aptidão constante dos genótipos (Figura 3A), a seleção balanceadora atuando sobre um só loco sempre maximiza a aptidão média de uma população com reprodução aleatória, ainda que, no caso de vantagem do heterozigoto, homozigotos com aptidões mais baixas sejam gerados por segregação mendeliana a cada geração (Charlesworth e Charlesworth, 2010).

Devido a essa propriedade, Sellis et al. (2011) demonstraram, teoricamente, que é provável que esta seja a trajetória da maior parte das mutações adaptativas em diploides. Esse estado “balanceado” seria, por definição, muito frequente e de curta duração, e não deixaria assinaturas passíveis de serem captadas pelos métodos focados em assinaturas de seleção balanceadora de longa duração. Por outro lado, teria a vantagem de manter diversidade adaptativa nas populações – mesmo que a curto prazo – o que compensaria a perda de aptidão causada pelos homozigotos. Esse fenômeno poderia ser responsável por manter diversidade adaptativa nas populações, que por sua vez poderia ser usada

¹³Pois $p \approx 1$ e $2pq \approx 2q$

Figura 4: Independente da frequência inicial de cada alelo em um sistema bi-alélico com vantagem de heterozigoto, um equilíbrio pode ser atingido (desde que o alelo sobreviva a perda por deriva). Se as aptidões relativas dos dois genótipos homozigotos em relação ao heterozigoto são idênticas, o equilíbrio é atingido em uma frequência de 0.5 para cada alelo (A). Se um dos genótipos homozigotos tem aptidão relativa maior do que o outro, a frequência de equilíbrio atingida será diferente 0.5 (B, onde a frequência de equilíbrio é de 0.3 para um alelo e 0.7 para o outro, consequentemente). No Capítulo 1, um exemplo real de sobredominância assimétrica é apresentado (Página 68).

prontamente pela seleção natural em casos de mudança repentina de pressão seletiva¹⁴. Embora seja uma proposição muito interessante, faltam evidências de empíricas para apoiá-la ou rejeitá-la.

Aptidão média Seria lógico pensar que, em um cenário de vantagem do heterozigoto, a inevitável geração de homozigotos a cada geração por segregação mendeliana levaria a uma redução da aptidão média da população. Entretanto,

¹⁴Esse fenômeno, que vem sendo cada vez mais estudado, é chamado de *selection on standing variation*.

esse não é o caso: a aptidão *média* de uma população (i.e, considerando todos os genótipos e suas frequências) com reprodução aleatória e vantagem do heterozigoto *atinge o seu máximo no equilíbrio*. Por isso, diz-se que a frequência de equilíbrio em um cenário de vantagem do heterozigoto é *aquela que maximiza a aptidão média da população* (Charlesworth e Charlesworth, 2010; Andrés, 2011). Portanto, embora a presença de homozigotos com baixa aptidão no caso da anemia falciforme, por exemplo, seja muito prejudicial para o indivíduo, a aptidão da população como um todo é mais alta quando indivíduos resistentes à malária são mantidos na população (Charlesworth e Charlesworth, 2010; Wright, 1937).

Seleção antagonista Pressões seletivas *opostas* entre diferentes contextos – ambientes, sexo do indivíduo, componentes de aptidão e estágios de desenvolvimento – podem gerar seleção antagonista ao nível genético-populacional (Connallon e Clark, 2013; Prout, 2000). Alelos selecionados de forma antagonista são aqueles que aumentam a aptidão em um contexto, mas diminuem-na em outro. Quando os contextos são componentes individuais de aptidão, tem-se um cenário conhecido como "pleiotropia antagonista", em que um mesmo loco afeta mais de um caráter, e para um deles um alelo é adaptativo, e para o outro, é deletério (Connallon e Clark, 2013).

Suponhamos dois componentes de aptidão: fertilidade e sobrevivência. Se os efeitos do homozigoto para uma dada mutação atuam em direções opostas nos dois caracteres, e o alelo favorável em relação a cada caráter é dominante sobre o alelo deletério, o resultado será um cenário essencialmente indistinguível daquele de vantagem de heterozigoto, em que o loco pode evoluir para ter frequências intermediárias através de seleção balanceadora. Tal cenário foi

descrito pela primeira vez em um trabalho sobre evolução do envelhecimento (Williams, 1957), em que se propôs que se um alelo causa alta fertilidade na juventude e envelhecimento precoce, o segundo efeito seria compensado pelo primeiro (revisado em Charlesworth, 2000).

Entretanto, a caminhada até o equilíbrio seria lenta, requerindo dezenas de milhares de gerações, mesmo com coeficientes seletivos não muito baixos (Connallon e Clark, 2013), o que implica que: (1) alelos próximos do equilíbrio devem ser relativamente antigos (ver discussão sobre escalas de tempo, abaixo); e (2) que as populações devem em geral estar longe do equilíbrio para loci sujeitos a esse regime (Connallon e Clark, 2013).

Exemplos de vantagem do heterozigoto Como na prática é muito difícil estabelecer inequivocamente qual (ou quais) mecanismo(s) de seleção balanceadora é responsável pela manutenção de um polimorfismo balanceado durante longas escalas de tempo, existem poucos exemplos não controversos de vantagem do heterozigoto nessa escala de tempo (Charlesworth e Charlesworth, 2010).

Os genes do MHC (HLA em humanos¹⁵) são um possível exemplo de vantagem do heterozigoto. Heterozigotos para alelos de genes codificadores de moléculas apresentadoras de antígenos teriam a capacidade de responder a um repertório maior de antígenos, e, portanto, responder melhor a infecções (Doherty e Zinkernagel, 1975). Existe, entretanto, uma vasta literatura que mostra as dificuldades de se diferenciar os mecanismos atuantes sobre os genes MHC e suas contribuições relativas¹⁶. O cenário mais provável é que diversos mecanismos contribuíram, em diferentes pontos do espaço e do tempo, para a evolução

¹⁵*Major histocompatibility complex* e *Human leukocyte antigen*, respectivamente.

¹⁶Ver Introdução e Discussão do artigo no Apêndice A.4 para uma discussão detalhada sobre este tópico.

desse sistema, e possivelmente concomitantemente. Por outro lado, é improvável que a vantagem do heterozigoto, apenas, seja capaz de explicar o número de alelos encontrados nos genes do MHC (De Boer et al., 2004).

Em escalas de tempo mais recentes (milhares de anos), o exemplo mais clássico é o da β - hemoglobina defeituosa que acarreta em anemia falciforme ao portador da mutação, e sua relação com proteção à malária¹⁷. Muitos outros exemplos de vantagem do heterozigoto são conhecidos com base em mensurações de aptidão de diferentes genótipos, por exemplo em animais domesticados – tais como porcos, cachorros, cavalos, gatos e ovelhas. Em muitos casos, traços selecionados por seleção artificial e que resultam em fenótipos desejáveis, geram como consequência fenótipos indesejados nos homozigotos recessivos (Hedrick, 2012). Um exemplo extremo é o da ausência de cauda no gato da raça Manx, que é o fenótipo heterozigoto (selecionado); em homozigose, a mutação é letal (Hedrick, 2012).

Aptidões não constantes

Na prática, a premissa de aptidões que são constantes ao longo do tempo é pouco realista: a variabilidade genética pode ser mantida ativamente mesmo na ausência de qualquer forma de vantagem do heterozigoto (Figura 3B).

Dependência de frequência negativa Assumimos novamente os genótipos A_1A_1 , A_1A_2 e A_2A_2 , e que A_1 é dominante, de forma que genótipos $A_1/-$ têm aptidão diferente de A_2A_2 . Se $A_1/-$ tem aptidão maior do que A_2A_2 quando A_1 é raro, mas a aptidão diminui na medida em que A_1 aumenta de frequência, tem-se um cenário de seleção com dependência de frequência (também cha-

¹⁷Ver também Página 68.

mada de vantagem do alelo raro). Ou seja, a aptidão marginal de um alelo é negativamente correlacionada à sua frequência na população, o que leva a um balanço estável das frequências de cada variante, onde nenhuma das duas é eliminada¹⁸ (Clarke, 1962; Charlesworth e Charlesworth, 2010). Aqui, a manutenção do polimorfismo não requer vantagem do heterozigoto.

Como nesse caso as aptidões dependem das frequências dos genótipos, o argumento acima sobre a frequência de equilíbrio ser aquela que maximiza a aptidão média da população não procede neste caso: aqui, o equilíbrio estável não precisa, por definição, coincidir com o a aptidão média máxima (Charlesworth e Charlesworth, 2010).

Exemplos de dependência de frequência É provável que cenários como aqueles descritos acima sejam bastante comuns em sistemas naturais. Um exemplo é a *seleção apostática*, em que predadores exercem pressão seletiva sobre suas presas ao “aprenderem” enquanto caçam: esse cenário favorece presas com fenótipos raros (Clarke, 1962; Charlesworth e Charlesworth, 2010). Outro exemplo é o mimetismo batesiano, onde uma espécie modelo perigosa, ou impalatável, é mimetizada por outra espécie (não perigosa), que é sujeita a predação. Variantes raras da espécie mimética, que porventura se pareçam com a espécie modelo, se beneficiam por serem evitadas por predadores, e aumentam de frequência; esse aumento de frequência aumenta a chance de um predador comer a espécie mimética e associar seu padrão a um gosto agradável (Charlesworth e Charlesworth, 2010), assim diminuindo sua vantagem.

É importante salientar que esses exemplos não levam inevitavelmente a um polimorfismo balanceado: isso depende da relação entre aptidão e frequências

¹⁸Em geral, ao longo do texto, assume-se que um loco permite duas variantes, ou seja, é bi-alelético.

dos genótipos. Mais ainda, ao contrário do que ocorre na vantagem do heterozigoto, o equilíbrio é instável: análises teóricas apontam que nesse tipo de regime seletivo a população oscilará permanentemente em torno da frequência de equilíbrio (Charlesworth e Charlesworth, 2010).

Outro exemplo importante é o sistema de auto-incompatibilidade: plantas com capacidade de se auto-polinizar costumam ter os chamados locos “S”, que compõem, junto com os genes do MHC em vertebrados, um dos sistemas mais polimórficos em termos de número de alelos (Charlesworth e Charlesworth, 2010). Nos exemplos mais simples, uma planta com grãos de pólen *S1* não pode fertilizar óvulos que também carregam alelos *S1*; assim, todas as plantas são heterozigotas por definição, e um novo alelo que surge tem uma enorme vantagem seletiva, pois o pólen que o carrega poderá fertilizar qualquer planta da população. Nesse cenário, um equilíbrio poderia ser atingido se todos os heterozigotos tivessem a mesma frequência (Charlesworth e Charlesworth, 2010).

Finalmente, outro exemplo é o da “corrida armamentista” entre hospedeiros e parasitas. No contexto de imunidade adquirida, um tipo comum de parasita provavelmente terá infectado uma grande proporção dos hospedeiros, que se tornam imunes a novas infecções. Assim, parasitas com novos tipos de antígenos têm uma vantagem dependente de frequência, acarretando altos níveis de polimorfismos na população de parasitas (Trachtenberg et al., 2003; Borghans et al., 2004; Slade e McCallum, 1992; Spurgin e Richardson, 2010).

É provável que parte da diversidade observada nos genes MHC seja mantida desta forma, uma vez que diferentes variantes dos genes que codificam as moléculas apresentadoras de antígenos são capazes de apresentar um repertório determinado de epítomos, e novas variantes que surgem no hospedeiro podem ter uma vantagem dependente de frequência (Trachtenberg et al., 2003;

Borghans et al., 2004). Essa hipótese é corroborada pela diversidade particularmente alta na fenda apresentadora de antígeno das moléculas MHC (Garrigan e Hedrick, 2003) e por uma correlação entre diversidade de patógenos e variabilidade no MHC humano (Prugnolle et al., 2005), mas, conforme mencionado anteriormente, o debate acerca dos mecanismos responsáveis pela diversidade dos genes MHC permanece acirrado.

Aptidões que variam no tempo Podemos imaginar cenários em que um alelo tem má performance em uma geração, e boa na geração subsequente (Figura 3B). Um bom exemplo aqui são as flutuações sazonais observadas por Dobzhansky (1937) em inversões cromossômicas de *D. pseudoobscura* ao longo das estações do ano. Mais recentemente, Bergland et al. (2014) detectaram centenas de loci (SNPs) em *D. melanogaster* cujas frequências variam dramaticamente entre estações do ano¹⁹, e argumentam que tais polimorfismos estão sujeitos a forte pressão seletiva, variável ao longo do tempo, dado que eles estão associados a fenótipos que variam entre as estações e que tais polimorfismos respondem a variações climáticas.

De fato, a teoria prevê que “polimorfismos protegidos” (Charlesworth e Charlesworth, 2010; Prout, 1968) podem permanecer em uma população com aptidões que variam ao longo do tempo, desde que a média geométrica das aptidões do heterozigoto seja mais alta que a de ambos os homozigotos ao longo do espaço de tempo considerado (Bergland et al., 2014; Charlesworth e Charlesworth, 2010; Gillespie e Langley, 1974)²⁰.

¹⁹Cerca de 10 gerações por verão, e duas por inverno (Bergland et al., 2014).

²⁰Entretanto, no estudo de Bergland et al. (2014) modelos mais realistas foram considerados, que levam em conta a possibilidade de gerações que se sobrepõem, múltiplos loci ligados e uma combinação de variações espaciais e temporais nas pressões seletivas (todos compatíveis com *D. melanogaster*), uma discussão além dos objetivos desta introdução.

Aptidões que variam no espaço Um polimorfismo protegido também pode ser mantido em ambientes que variam ao longo do espaço (Figura 3B), sob as condições previstas no modelo de Levene (1953) (ver Figura 1), se: (a) a aptidão relativa dos genótipos varia entre diferentes nichos ocupados pela espécie; (b) toda a seleção ocorre dentro de cada nicho, e (c) a reprodução ocorre aleatoriamente entre os indivíduos de cada nicho.

Com alguma abstração, podemos também incluir aqui casos em que existem aptidões diferentes para um dado genótipo considerando os dois sexos: se alelos A_1 e A_2 têm efeitos opostos sobre a aptidão de machos e fêmeas (antagonismo sexual), polimorfismos podem ser mantidos na ausência de vantagem do heterozigoto²¹. Em *D. melanogaster*, estima-se que 8% dos genes têm padrões compatíveis com esse tipo de seleção (Innocenti e Morrow, 2010).

No caso de pressões seletivas que variam ao longo do espaço ocupado por uma espécie, a existência ou não de uma frequência de equilíbrio depende não apenas da relação entre as aptidões dos diversos alelos entre os ambientes, mas da proporção de indivíduos alocados a cada ambiente e da quantidade de fluxo gênico entre os ambientes (Gloss e Whiteman, 2016; Charlesworth e Charlesworth, 2010). Sendo assim, nesse mecanismo a manutenção de polimorfismos a longo-prazo pode ou não ocorrer.

Em muitos dos casos documentados de seleção balanceadora de curto e longo prazo (i.e, que não são detectados com base em estudos de geração atual), é difícil ou impossível determinar qual dos mecanismos acima definidos é responsável pelo padrão observado. Ou seja, a detecção de assinaturas genômicas compatíveis com regimes seletivos de seleção balanceadora não permite dife-

²¹Ver também Página 18, onde o cenário de pleiotropia antagonista foi definido junto com o de vantagem de heterozigoto, pois ambos os mecanismos resultam em assinaturas indistinguíveis.

reenciar qual mecanismo é responsável pelo padrão observado (ver discussão no próximo item). Além disso, múltiplos mecanismos podem ter atuado simultaneamente ou em diferentes momentos ao longo da história evolutiva (Hedrick, 2012), como é o caso do MHC.

Evolução adaptativa no genoma humano

“It is easy to invent a selectionist explanation for almost any specific observation; proving it is another story. Such facile explanatory excesses can be avoided by being more quantitative.” (Kimura, 1983)

Durante décadas, o esforço de detectar casos de seleção positiva através do exame individual de genes candidatos resultou em alguns casos bem documentados, como o dos genes associados à persistência da enzima lactase em adultos (e.g. Bersaglieri et al., 2004) e à pigmentação da pele em humanos (Jablonski e Chaplin, 2010). Até recentemente, essa abordagem era a única forma prática de encontrar alvos de seleção positiva em humanos (Sabeti et al., 2006).

Nos últimos 15 anos, abordagens genômicas tornaram-se populares, e houve uma explosão de *scans* genômicos²² em busca de assinaturas de seleção positiva no genoma humano (Bamshad e Wooding, 2003; Bustamante et al., 2005; Enard et al., 2010; Fay et al., 2001; Nielsen, 2005; Sabeti et al., 2006; Sabeti et al., 2007). Até 2009, 21 *scans* para seleção positiva em humanos já haviam sido feitos (revisado em Akey, 2009).

Abordagens genômicas têm a vantagem de possibilitar a compreensão do

²²Em um *scan* genômico, uma estatística de interesse é calculada em janelas do genoma, e destaca-se aquelas que estão nos extremos empíricos da estatística e/ou além de algum limiar definido com base em simulações neutras.

impacto da seleção natural ao longo de todo o genoma e de inferir categorias funcionais mais sujeitas à ação da seleção positiva em humanos (Sabeti et al., 2006). Graças à grande quantidade de *scans* genômicos em busca de assinaturas de seleção positiva, sabe-se hoje que muitos dos genes com tais assinaturas estão relacionados com certas categorias: “imunidade e defesa”, “percepção sensorial” e “imunidade mediada por células T”, “gametogênese”, “espermatogênese e mobilidade” (Harris e Meyer, 2006; Nielsen et al., 2005). Para as regiões não-codificadoras, destacam-se processos biológicos como “neurogênese”, “outras atividades neuronais” e “desenvolvimento muscular” (Haygood et al., 2010).

Assinaturas de seleção balanceadora

Desde a proposição da teoria neutra, muitos testes para seleção balanceadora no passado recente²³ foram desenvolvidos. Por usarem o modelo neutro como modelo nulo, testes usados para detectar assinaturas de seleção natural são também chamados de “testes de neutralidade”²⁴.

Seleção atual

Classicamente, a detecção de instâncias de vantagem do heterozigoto foi restrita a estudos da geração atual, que focam em desvios das proporções genotípicas esperadas sob certas premissas (Hardy-Weinberg, panmixia) ou em relações entre genótipo/alelo e aptidão (Figura 5). Tais estudos são de extrema importância no sentido de avaliar aptidões de fenótipos e comprovar a relação entre o fenótipo selecionado e o genótipo subjacente, mas não focaremos neles aqui.

²³Na literatura, frequentemente usa-se a expressão “seleção balanceadora de curto prazo” (últimos milhares de anos).

²⁴Ver página 8.

Escala de Tempo	Teste	Fatores evolutivos
Geração atual	Proporções HW	Seleção
	Proporções mendelianas	Seleção
	Proporções reprodução aleatória	Seleção
	Associações genótipo-aptidão	Seleção
Passado recente (até ~ 1 milhão de anos)	Ewens-Watterson	Deriva, seleção, tempo
	Desequilíbrio de ligação	Recombinação, deriva, seleção, tempo
	Congruência geográfica	Estruturação populacional, deriva, seleção, tempo
Passado distante (> 1 milhão de anos)	Razão dN/dS	Mutação, seleção, tempo
	Mcdonald-Kreitman	Mutação, tempo de divergência entre espécies, seleção, tempo
	Teste HKA	Mutação, seleção, tempo
	D de Tajima	Mutação, deriva, estruturação populacional, tempo
	Polimorfismos trans-específicos	Mutação, tempo de divergência entre espécies, seleção, tempo

Figura 5: HW, Hardy-Weinberg, refere-se às proporções $p^2 + 2pq + q^2 = 1$ para os genótipos. Figura adaptada a partir de Hedrick (2012). As assinaturas e os testes de neutralidade correspondentes são abordados no item “Assinaturas de seleção balanceadora”.

Seleção de curto prazo

Em relação a eventos de seleção ocorridos no passado “recente” (até cerca de 1 milhão de anos atrás; Fu e Akey, 2013), um dos primeiros testes do modelo neutro foi o de Ewens-Watterson (e.g. Watterson, 1978), ou teste de homozigose da amostra (Figura 5). Esse teste assume o modelo IAM (modelo de infinitos alelos²⁵), e que a população encontra-se em equilíbrio entre mutação e deriva. O teste relaciona a expectativa de número de loci homozigotos esperados em uma população em equilíbrio com aquela que de fato é observada. Uma deficiência de homozigose pode ser interpretada como indício de seleção balanceadora (i.e.

²⁵Um modelo que postula que cada novo alelo que surge em uma população é “novo” ou “único”, i.e, diferente de todos os que surgiram antes. Esse modelo foi proposto por Kimura e Crow (1964) em uma tentativa de estimar a proporção de loci homozigotos em uma população diploide finita.

os alelos segregam em frequências intermediárias, diminuindo a homozigose observada). Esse teste foi muito influente em estudos prévios à era genômica, mas hoje em dia é preterido por outros, que têm maior poder.

Outros testes com poder para detectar assinaturas de eventos de seleção balanceadora compatíveis com essa escala de tempo olham: (a) a distribuição de frequências alélicas observada, comparando com aquela esperada sob o modelo neutro, (b) a variação genética e desequilíbrio de ligação²⁶ em certas regiões genômicas com as observadas em regiões evoluindo de forma neutra, (c) a diferenciação geográfica observada em certos loci com aquelas encontradas para marcadores neutros (Hedrick, 2006; Hedrick, 2012; Mitchell-Olds et al., 2007) (Figura 5).

A desvantagem de testes focados em desequilíbrio de ligação (Figura 5) é que essa assinatura é essencialmente indistinguível daquela deixada por varreduras seletivas incompletas (seleção positiva). Uma outra assinatura que pode ser observada ao nível populacional é que diferentes subpopulações terão pouca diferenciação entre elas para loci-alvo de seleção balanceadora, dado que ela mantém variantes em frequências intermediárias em ambas as populações. Essa assinatura depende de uma congruência entre regimes seletivos entre os ambientes ocupados pelas duas populações e de ausência de adaptação local de uma ou mais subpopulações (Figura 5).

Por outro lado, com a disponibilidade de dados de sequência, tornou-se possível investigar o efeito cumulativo da seleção ao longo de diversas gerações. O sinal de seleção no passado recente é gerado ou perdido ao longo de dezenas

²⁶Uma medida que reflete se dois alelos em dois diferentes *loci* coexistem de forma não-neutra em uma população. Alelos em desequilíbrio de ligação são encontrados mais frequentemente no mesmo haplótipo do que seria esperado se a recombinação ocorresse livremente entre eles (revisado em Cutter e Payseur, 2013).

a milhares de gerações, dependendo da influência de deriva genética, fluxo gênico e recombinação (Hedrick, 2012).

Seleção de longo prazo

O sinal de seleção que começou no passado distante e durou muito tempo²⁷ é determinado principalmente por mutação e seleção (Figura 5), e geralmente leva de milhares a milhões de gerações para ser gerado. Quando persiste por milhares a milhões de gerações, a seleção balanceadora resulta não apenas na manutenção de maior quantidade de alelos nas populações (Andrés et al., 2009), mas também em uma maior persistência, ao longo do tempo, da diversidade alélica em relação à variação neutra (Richman, 2000). Em outras palavras, em casos de seleção balanceadora de longo prazo, alelos segregando para um dado loco têm um TMRCA²⁸ mais longo, o que implica em determinadas assinaturas genômicas: excesso de polimorfismos na região do sítio selecionado (pois haverá mais tempo para que mutações ocorram no haplótipo) e uma manutenção de tais polimorfismos por mais tempo que o esperado para uma mutação neutra, que acaba por ser fixada ou perdida após, em média, $4N$ gerações (onde N é o tamanho da população).

Relação entre taxas de substituição não-sinônimas e sinônimas De acordo com a teoria neutra (Kimura, 1968; Kimura, 1983), mutações capazes de alterar a função de uma proteína (mutações não-sinônimas) geralmente são deletérias e, portanto, alvo da seleção purificadora. Análises comparativas apontam

²⁷À qual nos referiremos como seleção balanceadora de longo prazo, SBLP, e em humanos corresponde a eventos que ocorreram há mais de um milhão de anos, ainda que possam ter persistido até recentemente (Fu e Akey, 2013).

²⁸*Time to most recent common ancestor*, tempo de coalescência até o ancestral comum mais recente.

que ao menos 38% das mutações não-sinônimas sejam deletérias (Eyre-Walker e Keightley, 1999). Já as mutações sinônimas, que não alteram a sequência de aminoácidos da proteína, evoluiriam de forma neutra²⁹. A evolução adaptativa pode levar a um aumento da taxa de substituição de mutações não-sinônimas (dN), tornando-a mais alta do que a taxa de substituição sinônima (dS).

Nesse sentido, a razão $dN/dS > 1$ (ou $\omega > 1$) é uma assinatura genética de seleção positiva (Gillespie, 1991; Nielsen, 2005), mas também de seleção balanceadora (e.g. Bitarello et al., 2015³⁰) (Figura 5). Entretanto, o critério de $\omega > 1$ para considerar que genes estejam sob evolução adaptativa é muito conservador. Partindo da premissa de que a maior parte das mutações não-sinônimas é deletéria (Kimura e Crow, 1963; Kimura, 1968; Eyre-Walker e Keightley, 1999), o critério muitas vezes não é atendido quando genes inteiros são analisados. Isso ocorre porque geralmente apenas alguns códons estão sob seleção positiva ou balanceadora, enquanto a maior parte das mutações não-sinônimas são deletérias e, portanto, estão sob seleção purificadora³¹. Por isso, há algum tempo convencionou-se analisar subconjuntos de códons em busca de seleção (e.g. Hughes e Nei, 1988; Hughes e Nei, 1989; Bitarello et al., 2015) ou através de modelos que estimam diferentes valores de dN/dS para grupos de códons (Yang e Swanson, 2002; Bitarello et al., 2015), tornando possível inferir quais deles evoluíram adaptativamente.

Espectro de frequências alélicas Evidências de seleção do passado distante podem também ser baseadas na distribuição de frequência dos alelos de uma

²⁹Embora algumas mutações sinônimas possam ser alvo de seleção devido ao viés no uso de códons e uma parcela das mutações não-sinônimas ser neutra, a premissa é válida devido às proporções (e.g. Comeron et al., 2008).

³⁰Esta referência está disponibilizada no Apêndice A.4.

³¹Essa ideia é indiretamente explorada no Capítulo 2.

amostra, o espectro de frequências alélicas (SFS, *site frequency spectrum*). O SFS é uma contagem do número de mutações que existem em uma frequência de $x_i = i/n$ para $i = 1, 2, \dots, n - 1$ em uma amostra de tamanho n , onde x é a frequência relativa de cada contagem. Em outras palavras, o SFS sumariza as frequências alélicas das várias mutações presentes em uma amostra (Nielsen, 2005). Muitos testes estatísticos em genética de populações usam informação acerca da proporção de SNPs que são comuns ou raros, e acerca da frequência de alelos derivados e ancestrais, a fim de fazer inferências sobre a história demográfica e possíveis regimes seletivos. Um dos testes dentro dessa categoria é o D de Tajima (Tajima, 1989; Mitchell-Olds et al., 2007) (Figura 5). No Capítulo 1, nós propomos dois novos testes que exploram essa assinatura³².

Relação entre níveis de polimorfismo e divergência Outros testes para seleção balanceadora nessa escala de tempo se baseiam em comparações entre os níveis de polimorfismo (observados dentro de uma espécie) e os níveis de divergência (entre duas espécies)³³. Aqui inclui-se o teste Hudson-Kreitman-Aguadé, ou HKA (Hudson et al., 1987), usado para testar previsões do modelo neutro de evolução molecular. Um excesso de polimorfismo em relação a divergência pode ser interpretado como evidência de seleção balanceadora (Figura 5).

Uma versão modificada do teste HKA – o teste de McDonald-Kreitman (MK) – compara polimorfismos e divergência entre diferentes classes de mutação, como as substituições sinônimas e não-sinônimas (McDonald e Kreitman, 1991).

³²As figuras 1 e 5 do Capítulo 1 mostram espectros de frequência alélica esquemático e de dados reais, respectivamente.

³³Ao longo do texto, e particularmente no Capítulo 1, refiro-me a substituições entre humanos e chimpanzés embora em princípio possa se tratar de substituições entre quaisquer duas espécies.

Nas situações em que há um excesso de divergência não-sinônima, temos um padrão consistente com seleção positiva (que fixa diferenças entre espécies e remove polimorfismos, explicando o padrão descrito). Já um excesso de polimorfismos não-sinônimos pode ser interpretado como uma assinatura de seleção balanceadora (Figura 5). Entretanto, conforme discutido por Eyre-Walker (2006), essa estatística é muito sensível a mudanças de tamanho populacional, e mesmo aumentos modestos de N_e podem criar evidências espúrias de evolução adaptativa. Sella et al. (2009), por sua vez, argumentam que uma das premissas mais problemáticas do teste MK é a de que a fração de novas mutações que é neutra, que é estimada a partir dos dados de polimorfismo de uma das espécies, tenha permanecido constante durante a história evolutiva das duas espécies sendo comparadas. A razão pela qual essa premissa é problemática é que uma história demográfica que resulta em uma população atual fora de equilíbrio torna essa premissa falsa quando a seleção é fraca, resultando em estimativas erradas de taxa de evolução adaptativa (por exemplo, Eyre-Walker, 2006; Fay et al., 2001; Nielsen, 2005; Sella et al., 2009).

Partilhamento de polimorfismos entre espécies Finalmente, se um polimorfismo é mantido por seleção balanceadora por um tempo suficientemente longo, o mesmo polimorfismo pode ser encontrado em duas espécies-irmãs: um polimorfismo trans-específico (Hedrick, 2012; Klein et al., 1998) (Figura 5).

Se a possibilidade de que mutações idênticas tenham ocorrido independente nas duas espécies puder ser descartada – por exemplo requerindo-se que mais de um polimorfismo seja compartilhado dentro de um haplótipo de tamanho reduzido (e.g. Leffler et al., 2013; Teixeira et al., 2015) –, a outra possível explicação é que o polimorfismo seja neutro e mantido “ao acaso” nas duas espécies.

Se o tempo de divergência entre as espécies é muito maior do que o tempo médio de coalescência intra-específica, essa alternativa tem baixíssima probabilidade.

Teixeira et al. (2015), por exemplo, demonstraram que a probabilidade de um polimorfismo em humanos ser também um polimorfismo em chimpanzés e bonobos é da ordem de 10^{-10} . Mesmo supondo independência entre todos os SNPs, considerando amostras de 20 indivíduos de cada espécie, a probabilidade de haver um polimorfismo partilhado é de ~ 0.00005 , ou seja, praticamente zero. Portanto, embora essa assinatura seja extremamente convincente como evidência de seleção balanceadora de longo prazo, ela é bastante rara.

Seleção balanceadora no genoma humano

“Balancing selection is not unique to the human lineage, nor is it the dominant force in shaping the human genome. It is, however, there and its effects are without dispute.” (Vallender e Johnson, 2008)

Embora diversos *scans* genômicos tenham sido feitos com o intuito de localizar alvos de seleção positiva (revisado em Akey, 2009), poucos trabalhos, comparativamente, buscaram localizar alvos de seleção balanceadora. Em parte isso é devido às dificuldades de detecção desse tipo de seleção em escala genômica (Andrés et al., 2009). A Figura 6 resume os estudos que buscaram por assinaturas de seleção balanceadora em humanos.

Até muito recentemente, não se havia estabelecido mais que alguns poucos casos de seleção balanceadora em humanos. Mesmo com o advento de dados de sequência pra diversos genes, poucos alvos foram propostos além dos genes

HLA, do gene ABO e do gene da hemoglobina S (Allison, 1954; Asthana et al., 2005; Bubb et al., 2006; Hughes e Nei, 1988; Ségurel et al., 2013).

Artigo	Dados	Assinatura de seleção	Pop	Candidatos
Asthana et al. (2004)	Exons	Polimorfismos trans-específicos	dbSNP	8 trSNPs, dos quais 3 não-sinônimos
Bubb et al. (2006)	Exons	Densidade de SNPs elevada	-	HLA-DQA1, -DRB1, ABO, + 16 regiões**
Andres et al. (2009)	Exons	SFS, excesso de SNPs ^a	2	60 genes
Leffler et al. (2013)	Genoma	Polimorfismos trans-específicos (H-C-B)	1	PROKR2, FREM3, MTRR, IGFBP7, HUS1, ST3GAL1; MHC
DeGiorgio et al. (2014)	Genoma	SFS, excesso de SNPs	2	200 genes
Teixeira et al. (2015)	Exons	Polimorfismos trans-específicos (H-C)	1	LAD1, HLA-DQA1, HLA-C, HLA-DPB1
Bitarello et al. (in prep)	Genoma	SFS, excesso de SNPs	4	212 genes

Figura 6: Resumo dos estudos que buscaram assinaturas de seleção balanceadora de longo prazo em humanos. SFS, espectro de frequências alélicas; dbSNP, banco de dados de SNP, indica que o estudo não usou informações de frequências de SNPs estimadas para populações individuais, mas de um banco de dados comum; trSNPs, SNPs trans-específicos entre (H-C) humano-chimpanzé ou (H-C-B) humano, chimpanzé, bonobo. ^a, excesso de SNPs em relação a sítios divergentes entre humanos e chimpanzé, uma assinatura de seleção balanceadora. ** apesar de que eles encontram 16 regiões com alta diversidade genética além de genes HLA e o gene ABO, os autores as descartam como reais candidatas pois não possuem evidência de polimorfismo trans-específico, o que é um critério muito rigoroso (ver “Assinaturas de seleção balanceadora”).

O primeiro estudo em escala genômica que identificou novas assinaturas de seleção balanceadora em humanos foi o de Andrés et al. (2009). Nele, foi feito um *scan* para genes sob seleção balanceadora de longo prazo (SBLP) no genoma humano. O *scan* de Andrés et al. utilizou uma base de dados de exoma consistindo de 13.400 genes de duas populações humanas (19 norte-americanos com ancestralidade europeia e 20 norte-americanos com ancestralidade africana). O método consistiu em contrastar padrões de polimorfismo em cada gene com o resto do genoma e com as expectativas neutras, obtidas através de simulações

parametrizadas pelos próprios dados de sequenciamento. O método buscava genes com duas assinaturas de SBLP: excesso de variantes em frequências intermediárias e excesso de sítios polimórficos em humanos (em relação a substituições entre humanos e chimpanzés)³⁴.

Apenas genes extremos em relação a ambas as assinaturas, i.e, altamente significativos para dois testes independentes, foram considerados como candidatos. Apesar do baixo poder estatístico deste tipo de abordagem (Fijarczyk e Babik, 2015), ela tem poucos falsos positivos. Foram encontrados 60 genes com assinaturas de SBLP, muitos deles envolvidos com a imunidade, mas não restritos aos exemplos clássicos até então descritos (Key et al., 2014).

Foi observado que a maior parte dos loci sob seleção balanceadora eram compartilhados entre as duas populações, com poucas exceções : quatro genes com evidência de seleção balanceadora apenas nos americanos com ascendência africana e nove apenas nos americanos com ascendência europeia.

A ausência de sequências flanqueadoras não-codificadoras, intergênicas e intrônicas nas bases de dados desse estudo tem duas consequências: primeiro, os efeitos da seleção balanceadora sobre regiões ligadas a loci sob seleção não puderam ser quantificados; segundo, alvos de seleção balanceadora em regiões não-codificadoras não puderam ser detectadas.

Mais recentemente, DeGiorgio et al. (2014) desenvolveram dois novos testes para a identificação de padrões locais de diversidade esperados em posições ligadas a um polimorfismo balanceado (T1 e T2) para a identificação e usaram esses métodos em dados genômicos de duas populações (YRI, Yoruba, uma população africana e CEU, uma população de norte-americanos do Utah com ascendência norte-europeia). Eles identificaram 200 genes candidatos (Figura 6)

³⁴ Ambas as assinaturas servem de base para testes discutidos na seção anterior.

– alguns deles conhecidos (e.g. Andrés et al., 2009), e outros novos (Key et al., 2014).

Três importantes limitações destes trabalhos são que: (1) apesar de o trabalho mostrar que os métodos T1 e T2 têm poder elevado sob modelos simples, os autores não exploram modelos demográficos humanos (e.g. Gravel et al., 2011); (2) os 200 genes reportados provêm de uma lista dos “100 genes mais extremos” para cada teste e população, mas não foi estabelecido um critério para que 100 genes fossem reportados; portanto, esse trabalho, apesar de inquestionavelmente contribuir muito para o conhecimento acumulado de regiões do genoma humano que possuem assinaturas de SBLP, não fornece uma estimativa aproximada do quão frequente ela pode ser no genoma humano; (3) apesar de os autores terem usado dados de genoma completo, eles reportam como alvos apenas genes codificadores de proteínas, e não exploram ou reportam regiões genômicas candidatas que estão fora dos limites gênicos.

Finalmente, dois estudos em escala genômica buscaram por polimorfismos partilhados entre humanos e outras espécies: trata-se da assinatura mais extrema de SBLP, pois requer a manutenção de polimorfismos por milhões de anos. Leffler et al. (2013) olharam o genoma completo de humanos e chimpanzés e buscaram por polimorfismos compartilhados, encontrando seis genes com forte evidência de polimorfismos balanceados compartilhados (Figura 6), e Teixeira et al. (2015) olharam os éxons de humanos, chimpanzés e bonobos, encontrando forte evidência de polimorfismo balanceado partilhado entre as três espécies no gene *LAD1*, além de alguns genes HLA (Figura 6).

Eventos de seleção balanceadora de curto e longo prazo deixam assinaturas genômicas que não permitem diferenciar entre os mecanismos de seleção balanceadora expostos anteriormente. Além disso, um sinal de seleção de longo

prazo não significa que a seleção tenha perdurado até o passado recente ou até a geração atual. Vê-se, portanto, que definir escalas de tempo dos regimes seletivos é importante tanto no sentido de determinar quais as ferramentas adequadas, quanto no grau de resolução que se pode alcançar.

Por si só as análises de genética de populações acima descritas são capazes de identificar genes que evoluíram de forma não-neutra durante milhares ou até milhões de gerações. Essa é o tipo de investigação proposta no Capítulo 1, a fim de melhor compreender o impacto da seleção balanceadora de longo prazo no genoma humano. Tal tipo de investigação não fornece, todavia, informação conclusiva acerca do caráter fenotípico que foi alvo de seleção (Mitchell-Olds et al., 2007). Trata-se de um desafio em andamento na genômica de populações, e que será discutido no item Discussão e Considerações Finais (Página 214).

Carga genética induzida por seleção balanceadora

“Hence there must be a far greater number of different kinds of ailments whose characteristics are traceable to genetic changes of natural origin than there are different kinds of infectious diseases. This general reasoning does not in itself give us much idea, however, of the actual frequencies with which these mutational disorders occur in populations. For this it is necessary to turn to quantitative studies.” (Muller, 1950)

Como vimos, os polimorfismos de sequência podem estar evoluindo de forma neutra (provavelmente a grande maioria deles), mas podem também representar estados transientes de variantes genéticas rumo à eliminação por serem de-

letérias, ou rumo à fixação por serem vantajosas (Hudson et al., 1987; Mitchell-Olds et al., 2007; Sellis et al., 2011). Finalmente, uma certa quantidade desses polimorfismos é mantida em populações individuais por seleção balanceadora ou ao longo de toda a distribuição da espécie, por adaptação local³⁵ (revisado em Mitchell-Olds et al., 2007).

Estudos genômicos têm documentado grandes quantidades de polimorfismos, e tem-se feito muito progresso no sentido de compreender quais processos evolutivos moldaram tal variação (revisado em Mitchell-Olds et al., 2007). Para um geneticista evolutivo, detectar possíveis alvos de seleção natural no genoma é fundamental. Primeiramente, padrões compatíveis com um cenário de evolução adaptativa (seleção positiva ou balanceadora) devem ser buscados. Um caminho são as assinaturas de seleção e os testes de neutralidade, vistos anteriormente. Uma vez detectadas as regiões de interesse, é necessário investigar se há base biológica para sustentar que tais padrões de fato resultem de seleção. Finalmente, pode-se investigar se a seleção sobre um dado loco no genoma tem consequências deletérias sobre regiões vizinhas.

A existência de assinaturas de seleção positiva ou balanceadora indica que pode ter havido adaptação para algum traço (aquele relacionado à pressão seletiva), mas a consequência da pressão seletiva pode incluir também mudanças que causam má-adaptação, em função de mudanças que ocorrem em outras regiões do mesmo gene que está sendo selecionado, ou mesmo em regiões não codificadoras – porém funcionais – adjacentes.

³⁵Situação na qual genótipos de diferentes populações têm aptidão maior em seus ambientes de origem, devido a seleção natural histórica na região.

Carga genética

A compreensão da base genética da adaptação requer a definição de “carga genética”. O conceito foi discutido pela primeira vez por Haldane (1937), sendo posteriormente elaborado por Muller (1950). A ideia central é que uma população fica aquém de sua aptidão máxima por dois motivos: (1) a ocorrência de mutações deletérias recorrentes (carga mutacional) e; (2) a produção de homozigotos menos aptos, nas situações em que os heterozigotos são o genótipo mais apto (carga segregacional). Esse decréscimo na aptidão máxima de uma população é a carga genética.

Embora as mutações deletérias incorram em um custo de aptidão, elas nem sempre são removidas da população de forma eficiente. É ainda mais difícil remover mutações deletérias de populações pequenas (baixos valores de N_e ³⁶), e seu acúmulo pode levar a reduções no tamanho populacional e, em última instância, à extinção (Chun e Fay, 2011).

No âmbito da evolução molecular, a maior parte das mutações *não* são adaptativas, piorando (mutações mal-adaptativas) ou não interferindo (neutras) no grau de adaptação dos caracteres ao ambiente (Orr, 1998). A eficácia de remoção de variantes deletérias de uma população depende de vários fatores: mutação (que cria novas variantes deletérias constantemente), dominância (que influencia o quanto a mutação é “visível” para a seleção) (e.g. Sellis et al., 2011), demografia e ligação (Gravel, 2016).

Além disso, uma situação de má-adaptação de uma população pode ser cau-

³⁶Tamanho populacional efetivo: reflete o tamanho de uma população idealizada que estaria sujeita à deriva da mesma forma que a população de fato. O N_e pode ser menor que o tamanho real da população devido a vários fatores, incluindo variância no sucesso reprodutivo, uma história demográfica com gargalos genéticos (reduções extremas de tamanho populacional, seguida de uma expansão a partir de uma amostra da população original) e endogamia (revisado em Cutter e Payseur, 2013).

sada por falta de variação genotípica segregante para responder à seleção. Deriva genética e endogamia, por exemplo, removem as populações de seus picos adaptativos e podem levar à má-adaptação fenotípica (Crespi, 2000). A pleiotropia³⁷ pode resultar em populações mal-adaptadas pois a otimização conjunta de muitos caracteres é inviável (Charlesworth e Charlesworth, 2010; Crespi, 2000). Finalmente, a migração entre populações de indivíduos que se adaptaram em diferentes subpopulações pode também levar à má-adaptação (Charlesworth e Charlesworth, 2010; Crespi, 2000).

A má-adaptação pode, ainda resultar de pressões seletivas para a adaptação em sítios ligados, que discutirei em maior detalhes abaixo. Nesse sentido, seja em termos fenotípicos, seja em termos genéticos (acúmulo de mutações deletérias), o termo má-adaptação alude a um “custo adaptativo” com o qual as populações têm que arcar em função de estarem evoluindo sob determinadas pressões seletivas.

Efeitos de ligação sobre polimorfismos neutros

A última década documentou uma explosão de estudos sobre a prevalência e efeito da seleção natural, em particular no genoma da nossa espécie. Um dos achados mais marcantes foi o fato de haver uma proporção relativamente grande de mudanças entre humanos e chimpanzés que resultam da seleção natural. Por exemplo, alguns estudos mostraram que até 10% das substituições que carregamos resultam de seleção positiva (Bustamante et al., 2005; Fay et al., 2001), uma fração muito maior do que seria esperado por um neutralista (revisado em Eyre-Walker, 2006). Esses resultados foram obtidos usando uma

³⁷Fenômeno em que um gene afeta múltiplos caracteres, considerado o modo quase universal de atuação gênica.

abordagem que contrasta o grau de polimorfismo e divergência entre humanos e primatas³⁸, e revelou que há mais diferenças fixas não-sinônimas entre as duas espécies do que seria esperado sob neutralidade – diferenças essas que são explicáveis se supusermos que a seleção fixou mutações vantajosas diferentes entre as linhagens de humanos e chimpanzés (revisado em Eyre-Walker, 2006).

É esperado que a seleção natural deixe assinaturas específicas sobre os padrões de variação neutra intimamente ligados ao sítio com a mutação vantajosa. Essa ideia é a base dos métodos de genética molecular de populações que buscam por adaptações no genoma humano (Kreitman e Di Rienzo, 2004). Essa propriedade é crucial para que testes de neutralidade tenham poder para detectar regiões com assinaturas de seleção, ainda que o sítio selecionado seja apenas um (revisado em Charlesworth, 2006). A informação genética em escala genômica permite também fazer inferências sobre consequências da seleção natural sobre regiões do genoma fisicamente próximas aos genes ou regiões genômicas selecionados(as).

Existe também uma bem-documentada correlação positiva entre taxas de recombinação e níveis de polimorfismo em *Drosophila* (Zhang e Parsch, 2005) e em humanos (Hellmann et al., 2003). Essa correlação é consistente com a ideia de que a seleção positiva ocorre em diversos locais do genoma, e quando afeta um gene numa região de baixa recombinação, “arrasta” com ele uma parte do cromossomo (isto é, promove um evento de carona genética), que tem como consequência a perda da variação naquela região do genoma. Nesse processo, a variação neutra em sítios ligados é reduzida. E quanto menor for a taxa de recombinação na região, mais pronunciada será a perda de diversidade.

³⁸Conforme discutido na página 31 e na Figura 5.

Estas varreduras seletivas³⁹ levam a uma queda de diversidade em torno do sítio selecionado, que, com o passar do tempo, vai progressivamente aumentado na medida em que aumenta a distância entre o sítio neutro e o sítio-alvo. O tamanho da área vizinha que perderá polimorfismos neutros depende da intensidade da seleção, da taxa de mutação e da taxa de recombinação (revisado em Bamshad e Wooding, 2003). Sabe-se hoje que a seleção positiva em humanos afeta bastante os sítios neutros próximos ao sítio selecionado, reduzindo em 6% seu nível de polimorfismo ao longo de todo o genoma, e 11% na porção codificadora de proteínas do genoma (Cai et al., 2009).

Analogamente, a seleção balanceadora de longo prazo também afeta sítios neutros vizinhos (Charlesworth, 2006). Por aumentar o tempo de coalescência, a seleção balanceadora leva a um aumento de diversidade em torno do sítio selecionado, além de mudar o formato do espectro de frequências alélicas local (SFS), que passa a ter um excesso de alelos segregando em frequências próximas à do polimorfismo balanceado (Andrés et al., 2009; Andrés, 2011; Bamshad e Wooding, 2003; Charlesworth, 2006).

Entretanto, ao contrário das varreduras seletivas, que envolvem altos coeficientes seletivos e diminuem a diversidade neutra em relativamente poucas gerações, gerando assinaturas que se estendem por longos trechos do cromossomo (Bamshad e Wooding, 2003), a seleção balanceadora de longo prazo, por envolver escalas de tempo de milhares a milhões de anos, gera assinaturas em curtos segmentos em torno do polimorfismo balanceado⁴⁰. Isso ocorre porque, ao longo de muitas gerações, a recombinação tem a oportunidade de ir "que-

³⁹Fenômeno em que uma mutação recém-surgida e altamente adaptativa sobe rapidamente de frequência na população.

⁴⁰Uma propriedade que é explorada quando avaliamos o poder de nossas estatísticas no Capítulo 1.

brando" a ligação entre o polimorfismo balanceado e os sítios neutros vizinhos (Andrés, 2011; Charlesworth, 2006), assim reduzindo o efeito de ligação a segmentos curtos do cromossomo.

Efeitos de ligação sobre polimorfismos não-neutros

Existem dois modos através dos quais a seleção sobre um traço interfere sobre a seleção sobre outros traços. O primeiro se dá em condições em que um gene tem funções pleiotrópicas. Seleção positiva ou balanceadora pressupõe adaptação para algum traço (aquele relacionado à pressão seletiva), mas a consequência da pressão seletiva em termos de fixação de mutações pode não ser uma adaptação para todas as possíveis funções que exerce. Nesse caso, o “subproduto” seria uma má-adaptação. Como exemplo, temos os genes HLA, que estão relacionados à resposta imune em humanos e têm fortes evidências de seleção positiva e balanceadora. Por outro lado, muitas doenças inflamatórias e autoimunes também estão relacionadas aos genes de HLA (Becker et al., 1998).

O segundo modo ocorre quando o sítio selecionado interfere sobre o destino de mutações não-neutras em sítios ligados geneticamente⁴¹. A seleção natural não atua independentemente sobre locos ligados e, sendo assim, sua eficiência está diretamente relacionada à taxa de recombinação (Hill e Robertson, 1966; Comeron et al., 2008). A ligação entre sítios reduz a eficácia da seleção natural em populações finitas (Hill e Robertson, 1966; Comeron et al., 2008). Mantendo-se as outras variáveis constantes, espera-se que o grau de interferência entre alelos selecionados varie entre regiões com diferentes taxas de recombinação, potencialmente levando a mudanças nas taxas de evolução⁴². Regiões de baixa

⁴¹Ver Figura 1 do Capítulo 2 na página 170.

⁴²*Hill-Robertson Effect* é o nome dado a essa interferência (Hill e Robertson, 1966; Comeron et al., 2008).

recombinação estariam sujeitas a efeitos mais pronunciados de Hill-Robertson, dado que a baixa recombinação reduz a independência entre os sítios, aumentando o efeito relativo da deriva sobre a região – o que equivale a uma redução do N_e local (Maynard-Smith e Haigh, 1974; Comeron et al., 2008).

Diante da informação de que substituições adaptativas (fixadas por seleção positiva) são relativamente comuns, tornou-se importante investigar a forma como esses eventos influenciam a atuação da seleção natural em regiões adjacentes do genoma. As trajetórias até a fixação das variantes neutras ligadas a variantes vantajosas permanecem inalteradas – dado que a probabilidade de uma mutação vantajosa carregar consigo uma mutação neutra é diretamente proporcional à frequência do alelo na população (Birky e Walsh, 1988; Chun e Fay, 2011; Comeron et al., 2008; Harris, 2010; Charlesworth, 2012). Entretanto, mutações levemente deletérias, quando ligadas geneticamente a variantes vantajosas, terão probabilidade de fixação maior do que a esperada sob o modelo neutro (Lynch, 2007).

O efeito da seleção positiva sobre regiões ligadas em *Drosophila* foi investigado por Betancourt e Presgraves (2002), que verificaram que os genes de evolução adaptativa lenta (sujeitos principalmente à seleção purificadora) se concentram principalmente nas regiões de baixa recombinação. Além disso, foi encontrada uma correlação negativa significativa entre a taxa de evolução (medida em termos de dN) e a proporção de uso de códons ótimos⁴³. Os autores descartaram a possibilidade de restrição seletiva relaxada; eles concluem que a causa para essa correlação é a interferência do sítio codificador sobre o fraco coeficiente seletivo para uso ótimo de códons em sítios vizinhos fortemente li-

⁴³Diferentes sítios sinônimos são seletivamente diferentes, uma vez que certos códons (ditos “ótimos”) são utilizados mais frequentemente que outros, possivelmente em função de eficiência e precisão da tradução (Betancourt e Presgraves, 2002).

gados.

A conclusão geral é que, ao menos em *Drosophila*, parece haver uma “hierarquia de pressões seletivas”: a seleção contra mutações deletérias é mais forte do que a seleção sobre mutações não-sinônimas vantajosas, que por sua vez é maior do que a seleção para códons ótimos. O que não significa, conforme os próprios autores salientam, que o efeito cumulativo de muitos códons subótimos seja negligenciável.

Além disso, foi demonstrado que a seleção direcional forte (positiva ou negativa) é capaz de gerar um aumento da proporção de variantes deletérias segregando em regiões adjacentes àquelas que foram alvo da seleção direcional em humanos (Chun e Fay, 2011). Assim como os *scans* genômicos para seleção balanceadora são muito menos abundantes que aqueles para seleção positiva, o mesmo ocorre em relação ao acúmulo de deletérias: até o momento, nenhum estudo investigou especificamente o impacto que a seleção balanceadora tem sobre o acúmulo de variantes deletérias em sítios na vizinhança do polimorfismo balanceado.

Existe evidência de que genes na vizinhança dos genes HLA têm um excesso de variantes potencialmente deletérias (Mendes, 2013; Lenz et al., 2016). Entretanto, dadas as várias particularidades dessa região genômica (Meyer et al., 2006), permanece em aberto qual o efeito que a seleção balanceadora tem, em geral, sobre variantes não-neutras ligadas. No Capítulo 2 eu abordo essa questão.

Relevância, Questões & Hipóteses

Relevância

Na última década, com a implementação cada vez mais frequente de *scans* genômicos, foram geradas várias listas de regiões e genes do genoma humano e de outras espécies que têm assinaturas de seleção positiva. O acesso a sequências de genomas de diversas espécies com ferramentas bioinformáticas estimulou a quantificação da evolução adaptativa, que resulta nos padrões de polimorfismo observados em humanos. O fato de grupos de genes definidos com base em sua função em categorias tais como “espermatogênese”, “olfação”, “percepção sensorial” e “resposta imune” serem recorrentes nas listas de genes candidatos à ação da seleção positiva (e.g. Nielsen, 2005; Sabeti et al., 2006) é algo que, intrinsecamente, é provido de sentido biológico: muitos genes nessas categorias estão diretamente envolvidos em interações com o ambiente. Tais observações aumentam nossa confiança em relação aos *scans* genômicos, ao mesmo tempo em que nos ajudam a compreender as adaptações específicas de nossa espécie. Por ser considerada amplamente como o principal mecanismo responsável pela evolução adaptativa, a seleção positiva foi e é intensamente estudada.

A variabilidade genética é o objeto de estudo do geneticista de populações. O regime seletivo conhecido como seleção balanceadora engloba uma série de mecanismos capazes de manter polimorfismos adaptativos segregando nas populações por curtos ou longos períodos de tempo. Se por um lado hoje temos um mapa abrangente de genes que sofreram seleção positiva em humanos, existe uma deficiência de informação acerca dos alvos e da abrangência da seleção balanceadora na história evolutiva humana.

Um dos motivos é histórico, conforme exposto na introdução: se outrora foi um regime seletivo que entusiasmou gigantes da biologia evolutiva como Dobzhansky e Fisher, a seleção balanceadora deixou de ser o foco de pesquisa dos biólogos evolutivos por mais de duas décadas após o advento da teoria neutra de Kimura (1968) e Kimura (1983). De certa forma, a demonstração de Hughes e Nei (1988) de que a seleção balanceadora mantém níveis de diversidade especificamente na fenda apresentadora de antígenos de genes do MHC – compatível com uma ação da seleção balanceadora – marca um renascimento do interesse por esse tópico.

O MHC segue sendo o melhor e menos controverso exemplo de seleção balanceadora em humanos, bem como um exemplo de que múltiplos mecanismos podem atuar em um mesmo sistema: existem evidências independentes de vantagem do heterozigoto, seleção dependente de frequência e pressão seletiva variável ao longo do tempo e/ou do espaço atuando sobre esses genes, e um acirrado debate acerca da proporção com que cada mecanismo contribui para os padrões de alta diversidade do MHC.

Ainda assim, a defasagem de conhecimento que temos acerca dos alvos de seleção balanceadora no genoma humano em relação aos de seleção positiva, e as propriedades entre esses regimes seletivos, é enorme. Entre os poucos estudos que buscaram assinaturas de seleção balanceadora, existem diversas limitações incluindo: uma falta de poder para detectar as assinaturas, dados insuficientes (poucas populações ou restritos a exons, por exemplo). Dos que buscaram por assinaturas em escala genômica, um estudo discute apenas os alvos codificadores de proteínas – ainda que todo o genoma tenha sido analisado (DeGiorgio et al., 2014) – e o outro (Leffler et al., 2013) documenta que a maior parte dos alvos é não-gênico, i.e, possivelmente regulatórios ou relacionados à

expressão gênica. Entretanto, este último usou um critério muito estrigente na determinação de alvos de seleção balanceadora (a presença de polimorfismos compartilhados com chimpanzé), de forma que faltam estudos que apoiem ou contestem essa observação.

A detecção de assinaturas de seleção natural é possível, em grande parte, devido ao sinal produzido sobre sítios neutros adjacentes aos sítios alvo de seleção. O método desenvolvido no Capítulo 1 (NCD, *Non-Central Deviation*) é um exemplo de utilização dessa propriedade: mesmo que apenas um sítio seja efetivamente mantido polimórfico por seleção balanceadora⁴⁴, ele irá afetar níveis de polimorfismo neutro adjacentes, por efeito de ligação, e a extensão desse efeito sobre a vizinhança depende da escala de tempo, duração e intensidade de seleção.

Por outro lado, pouco se sabe sobre o efeito que polimorfismos balanceados têm sobre variantes não-neutras adjacentes. Esse conhecimento é escasso mesmo para alvos de seleção positiva, e praticamente inexistente para alvos de seleção balanceadora – com exceção de alguns estudos em genes HLA (e.g. Oosterhout, 2009; Lenz et al., 2016). Entender melhor o impacto da seleção balanceadora na evolução humana é importante não apenas para entender como nos tornamos o que somos, mas também para melhor entender doenças complexas que ocorrem com frequência relativamente alta em humanos (Vallender e Johnson, 2008).

⁴⁴Uma possibilidade, mas não uma certeza. Para os genes HLA sabe-se que muitos sítios são ativamente mantidos polimórficos.

Questões & Hipóteses

Nesse contexto, as principais questões abordadas nesta tese foram: **(1)** é possível desenvolver métodos mais poderosos para encontrar regiões do genoma que evoluem sob seleção balanceadora? **(2)** quais são os alvos de seleção balanceadora de longo prazo em humanos? **(3)** quais são as propriedades biológicas desses alvos: eles são majoritariamente genes (codificadores de proteínas), regiões regulatórias, ou regiões que afetam a expressão gênica? **(4)** quais são as categorias funcionais mais abundantes entre genes-alvo de seleção balanceadora: fora os genes HLA, que estão envolvidos da resposta imune, o que podemos dizer sobre os alvos em termos de função? **(5)** o que sabemos sobre a importância biológica de alguns desses genes candidatos, com base em estudos independentes? **(6)** os alvos de seleção balanceadora são partilhados entre populações ou continentes? **(7)** qual é a prevalência de assinaturas de seleção balanceadora de longo prazo no genoma humano? Podemos quantificar a proporção do genoma humano que foi moldado por mecanismos de manutenção de diversidade? **(8)** a seleção balanceadora sobre um ou mais sítios interfere na eficácia da seleção purificadora sobre sítios não-neutros adjacentes?

As hipóteses exploradas nesse contexto foram:

- (a) a seleção balanceadora não é muito frequente no genoma, mas provavelmente mais frequente do que o que se estimou até agora, dado que os métodos e/ou os dados utilizados não permitiram obter uma estimativa menos conservadora de sua frequência no genoma. A fim de testar essa hipótese, propusemos uma nova estatística (Capítulo 1), com poder aumentado em relação a testes de neutralidade comumente usados e otimizada para vasculhar o genoma humano.

- (b) a seleção balanceadora afeta tanto regiões gênicas quanto regiões regulatórias/controladoras de expressão.
- (c) a seleção balanceadora afeta majoritariamente genes relacionados com a defesa do organismo (e.g. genes que integram vias do sistema imunológico, proteínas de membrana, de matriz extracelular, etc), ou relacionados a interações com o ambiente extracelular ou com outras células, e com a reprodução⁴⁵.
- (d) haverá maior compartilhamento de alvos de seleção balanceadora entre populações de um mesmo continente do que entre populações de continentes distintos, e o nível geral de compartilhamento entre quaisquer populações será alto; poucos genes/regiões apresentariam sinais opostos de seleção em diferentes populações/continentes.
- (e) existe um excesso de SNPs não-sinônimos (ou de SNPs deletérios) em regiões próximas a alvos de seleção balanceadora. A possível explicação seria que por efeitos de ligação, mutações fracamente deletérias poderiam aumentar de frequência juntamente com os sítios alvos de seleção balanceadora. Por outro lado, uma deficiência de mutações deletérias poderia estar relacionada à frequência das variantes (um artefato) ou ao fato de que a seleção balanceadora aumenta o tamanho efetivo local.

As hipóteses (a-d) foram testadas com os alvos obtidos com um método que desenvolvi com colaboradores, que vasculhou o genoma todo em busca de assinaturas de seleção balanceadora, para quatro populações de dois continentes.

⁴⁵Embora haja relativamente poucos casos descritos até o momento e os mecanismos não sejam totalmente compreendidos, eles parecem promissores. Por exemplo, genes envolvidos em espermatogênese, reconhecimento entre espermatozoide e óvulo, e hormônio folículo estimulante (revisado em Vallender e Johnson, 2008).

Tais hipóteses são exploradas no **Capítulo 1**.

A premissa para a hipótese (e) é que a seleção balanceadora é suficientemente mais forte do que a seleção purificadora contra as variantes deletérias, de forma que a segunda não consiga contrabalancear a primeira. Esta última hipótese foi explorada no **Capítulo 2**.

Bibliografia

- Akey, J. M. (2009). "Constructing genomic maps of positive selection in humans: where do we go from here?" Em: *Genome Research* 19 (5), pp. 711–722.
- Allison, A. C. (1954). "Protection Afforded by Sickle-cell Trait Against Subtertian Malarial Infection." Em: *British Medical Journal* 1 (4857), pp. 290–294.
- Andrés, A. M. (2011). "Balancing Selection in the Human Genome". Em: *eLS*, pp. 1–8.
- Andrés, A. M. et al. (2009). "Targets of balancing selection in the human genome." Em: *Molecular Biology and Evolution* 26 (12), pp. 2755–64.
- Asthana, S., S. Schmidt e S. R. Sunyaev (2005). "A limited role for balancing selection". Em: *Trends in genetics : TIG* 21 (1), pp. 30–32.
- Bamshad, M. e S. P. Wooding (2003). "Signatures of natural selection in the human genome". Em: *Genetics* 4 (February), pp. 99–111.
- Becker, K. G., R. M. Simon, J. E. Bailey-Wilson, B. Freidlin, W. E. Biddison, H. F. McFarland e J. M. Trent (1998). "Clustering of non-major histocompatibility complex susceptibility candidate loci in human autoimmune diseases". Em: *Proceedings of the National Academy of Sciences of the United States of America* 95 (17), pp. 9979–9984.
- Bergland, A. O., E. L. Behrman, K. R. O'Brien, P. S. Schmidt e D. A. Petrov (2014). "Genomic Evidence of Rapid and Stable Adaptive Oscillations over Seasonal Time Scales in *Drosophila*". Em: *PLoS Genetics* 10 (11), e1004775.
- Bernardi, G. (2007). "The neoselectionist theory of genome evolution." Em: *Proceedings of the National Academy of Sciences of the United States of America* 104 (20), pp. 8385–90.
- Bersaglieri, T., P. C. Sabeti, N. Patterson, T. Vanderploeg, S. F. Schaffner, J. A. Drake, M. Rhodes, D. E. Reich e J. N. Hirschhorn (2004). "Genetic signatures of strong recent positive selection at the lactase gene." Em: *American Journal of Human Genetics* 74 (6), pp. 1111–20.

- Betancourt, A. J. e D. C. Presgraves (2002). "Linkage limits the power of natural selection in *Drosophila*." Em: *Proceedings of the National Academy of Sciences of the United States of America* 99 (21), pp. 13616–20.
- Birky, W. e J. B. Walsh (1988). "Effects of linkage on rates of molecular evolution." Em: *Proceedings of the National Academy of Sciences of the United States of America* 85, pp. 6414–6418.
- Bitarello, B. D., R. D. S. Francisco e D. Meyer (2015). "Heterogeneity of dN/dS Ratios at the Classical HLA Class I Genes over Divergence Time and Across the Allelic Phylogeny". Em: *Journal of Molecular Evolution* 82 (1), pp. 38–50.
- Borghans, J., J. Beltman e R. Boer (2004). "MHC polymorphism under host-pathogen coevolution". Em: *Immunogenetics* 55 (11), pp. 732–739.
- Bromham, L. e D. Penny (2003). "The modern molecular clock." Em: *Nature reviews. Genetics* 4 (3), pp. 216–224.
- Bubb, K. L. et al. (2006). "Scan of human genome reveals no new Loci under ancient balancing selection." Em: *Genetics* 173 (4), pp. 2165–77.
- Bustamante, C. D. et al. (2005). "Natural selection on protein-coding genes in the human genome". Em: *Nature* 437 (7062), pp. 1153–1157.
- Cai, J. J., J. M. Macpherson, G. Sella e D. A. Petrov (2009). "Pervasive Hitchhiking at Coding and Regulatory Sites in Humans". Em: *PLoS Genetics* 5 (1), pp. 1–13.
- Chakraborty, M. e J. D. Fry (2015). "Evidence that Environmental Heterogeneity Maintains a Detoxifying Enzyme Polymorphism in *Drosophila melanogaster*". Em: *Current Biology* 26 (2), pp. 1–5.
- Charlesworth, B. (2000). "Fisher, Medawar, Hamilton and the Evolution of Aging". Em: *Genetics* 156 (3), pp. 927–931.
- (2012). "The effects of deleterious mutations on evolution at linked sites". Em: *Genetics* 190 (1), pp. 5–22.
- Charlesworth, B. e D. Charlesworth (2010). *Elements of Evolutionary Genetics*. 1ª ed. Roberts e Company Publishers, p. 768. ISBN: 0981519423.
- Charlesworth, D. (2006). "Balancing selection and its effects on sequences in nearby genome regions." Em: *PLoS Genetics* 2 (4), pp. 379–384.
- Chun, S. e J. C. Fay (2011). "Evidence for hitchhiking of deleterious mutations within the human genome." Em: *PLoS genetics* 7 (8), e1002240.

- Clarke, B. (1962). "Balanced polymorphism and the diversity of sympatric species". Em: *Taxonomy and Geography*. Ed. por D. Nichols. Oxford: Systematics Association.
- Comeron, J. M., a. Williford e R. M. Kliman (2008). "The Hill-Robertson effect: evolutionary consequences of weak selection and linkage in finite populations." Em: *Heredity* 100 (1), pp. 19–31.
- Connallon, T. e A. G. Clark (2013). "Antagonistic versus nonantagonistic models of balancing selection: characterizing the relative timescales and hitchhiking effects of partial selective sweeps." Em: *Evolution; international journal of organic evolution* 67 (3), pp. 908–17.
- Crespi, B. J. (2000). "Short Review The evolution of maladaptation". Em: *Heredity* 84 (March 1999), pp. 623–629.
- Crow, J. F. (1987). "Muller, Dobzhansky, and overdominance". Em: *Journal of the History of Biology* 20 (3), pp. 351–380.
- Cutter, A. D. e B. A. Payseur (2013). "Genomic signatures of selection at linked sites: unifying the disparity among species." Em: *Nature reviews. Genetics* 14 (4), pp. 262–74.
- Darwin, C. (1859). *The origin of species: complete and fully illustrated*. 1979^a ed. New York: Gramercy Books. ISBN: 9780517123201.
- (1876). *The effects of cross and self fertilisation in the vegetable kingdom*.
- De Boer, R. J., J. a. M. Borghans, M. van Boven, C. Kesmir e F. J. Weissing (2004). "Heterozygote advantage fails to explain the high degree of polymorphism of the MHC." Em: *Immunogenetics* 55 (11), pp. 725–731.
- DeGiorgio, M., K. E. Lohmueller e R. Nielsen (2014). "A model-based approach for identifying signatures of ancient balancing selection in genetic data." Em: *PLoS genetics* 10 (8), e1004561.
- Dempster, E. R. (1955). "Maintenance of genetic heterogeneity." Em: *Cold Spring Harbor Symposium on Quantitative Biology*. Cold Spring Harbor Laboratory Press, pp. 25–32.
- Dobzhansky, T. (1937). *Genetics and the Origin of Species*. 2nd. New York: Columbia University Press.
- Doherty, P. C. e R. M. Zinkernagel (1975). "Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex". Em: *Nature* 256 (5512), pp. 50–52.
- Enard, D., F. Depaulis e H. Roest Crolius (2010). "Human and Non-Human Primate Genomes Share Hotspots of Positive Selection". Em: *PLoS Genetics* 6 (2), pp. 1–13.

- Eyre-Walker, A. (2006). "The genomic rate of adaptive evolution." Em: *Trends in ecology & evolution* 21 (10), pp. 569–75.
- Eyre-Walker, A. e P. D. Keightley (1999). "High genomic deleterious mutation rates in hominids". Em: *Nature* 397 (6717), pp. 344–347.
- Fay, J. C., G. J. Wyckoff e C.-I. I. Wu (2001). "Positive and negative selection on the human genome." Em: *Genetics* 158 (3), pp. 1227–34.
- Fijarczyk, A. e W. Babik (2015). "Detecting balancing selection in genomes: Limits and prospects". Em: *Molecular Ecology*, n/a–n/a.
- Fisher, R. A. (1922). "On the Dominance Ratio." Em: *Proc. R. Soc.* 42, pp. 321–341.
- Fu, W. e J. M. Akey (2013). "Selection and Adaptation in the Human Genome". Em: *Annual Review of Genomics and Human Genetics* 14 (1), pp. 467–489.
- Garrigan, D. e P. W. Hedrick (2003). "Detecting adaptive molecular polymorphism : Lessons from the MHC". Em: *Evolution* 57 (8), pp. 1707–1722.
- Gillespie, J. H. (1991). *The causes of molecular evolution*. Oxford: Oxford University Press. ISBN: 0-19-509271-6.
- Gillespie, J. H. e C. Langley (1974). "A general model to account for enzyme variation in natural populations". Em: *Genetics* 76 (4), pp. 837–48.
- Gloss, A. D. e N. K. Whiteman (2016). "Balancing Selection: Walking a Tightrope". Em: *Current Biology* 26 (2), R73–R76.
- Gravel, S. (2016). "When Is Selection Effective?" Em: *Genetics* 203 (1), pp. 451–462.
- Gravel, S., B. M. Henn, R. N. Gutenkunst, A. R. Indap, G. T. Marth, A. G. Clark, F. Yu, R. A. Gibbs e C. D. Bustamante (2011). "Demographic history and rare allele sharing among human populations." Em: *Proceedings of the National Academy of Sciences of the United States of America* 108 (29), pp. 11983–8.
- Haldane, J. (1937). "The Effect of Variation on Fitness". Em: *The American Naturalist* 71 (735), pp. 337–349.
- Harris, E. E. (2010). "Nonadaptive processes in primate and human evolution." Em: *American journal of physical anthropology* 143 Suppl, pp. 13–45.
- Harris, E. E. e D. Meyer (2006). "The Molecular Signature of Selection Underlying Human Adaptations". Em: *Yearbook of Physical Anthropology* 130, pp. 89–130.

- Haygood, R., C. C. Babbitt, O. Fedrigo e G. A. Wray (2010). "Contrasts between adaptive coding and noncoding changes during human evolution". Em: *Proceedings of the National Academy of Sciences of the United States of America* 107 (17), pp. 7853–7857.
- Hedrick, P. W. (2006). "Genetic Polymorphism in Heterogeneous Environments: The Age of Genomics". Em: *Annual Review of Ecology, Evolution, and Systematics* 37, pp. 67–93.
- (2012). "What is the evidence for heterozygote advantage selection?" Em: *Trends in Ecology & Evolution* 27 (12), pp. 698–704.
- Hellmann, I., I. Ebersberger, S. E. Ptak, S. Pääbo e M. Przeworski (2003). "A neutral explanation for the correlation of diversity with recombination rates in humans." Em: *American journal of human genetics* 72 (6), pp. 1527–35.
- Hill, W. G. e A. Robertson (1966). "The effect of linkage on limits to artificial selection". Em: *Genetical Research* 8 (03), p. 269.
- Hudson, R. R., M. Kreitman e M. Aguade (1987). "A Test of Neutral Molecular Evolution Based on Nucleotide Data". Em: *Genetics* 116 (1), pp. 153–159.
- Hughes, A. L. e M. Nei (1989). "Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection". Em: *Proceedings of the National Academy of Sciences of the United States of America* 86 (3), pp. 958–962.
- Hughes, A. L. e M. Nei (1988). "Pattern of nucleotide substitution at major histocompatibility class I loci reveals overdominant selection". Em: *Letters to Nature* 335 (8), pp. 167–170.
- Innocenti, P. e E. H. Morrow (2010). "The sexually antagonistic genes of drosophila melanogaster". Em: *PLoS Biology* 8 (3), e1000335.
- Jablonski, N. G. e G. Chaplin (2010). "Human skin pigmentation as an adaptation to UV radiation". Em: *Proceedings of the National Academy of Sciences* 107 (Supplement_2), pp. 8962–8968.
- Key, F. M., J. C. Teixeira, C. de Filippo e A. M. Andrés (2014). "Advantageous diversity maintained by balancing selection in humans". Em: *Current Opinion in Genetics & Development* 29, pp. 45–51.
- Kimura, M. (1991). "The neutral theory of molecular evolution: a review of recent evidence". Em: *Japanese Journal of Genetics* 66 (4), pp. 367–386.
- Kimura, M. (1968). "Evolutionary rate at the molecular level". Em: *Nature* 217, pp. 624–626.

- (1983). *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge University Press. ISBN: 9780511623486. URL: <http://ebooks.cambridge.org/ref/id/CB09780511623486>.
- Kimura, M. e J. F. Crow (1963). "The Measurement of Effective Population Number". Em: *Evolution* 17 (3), pp. 279–288.
- (1964). "The Number of Alleles that Can Be Maintained in a Finite Population". Em: *Genetics* 49, pp. 725–738.
- Klein, J., A. Sato, S. Nagl e C. O'hUigin (1998). "Molecular trans-species polymorphism". Em: *Annual Review of Ecology and Systematics* 29, pp. 1–21.
- Kreitman, M. e A. Di Rienzo (2004). "Balancing claims for balancing selection". Em: *Trends in Genetics* 20 (7), pp. 300–304.
- Lande, R. (1975). "The maintenance of genetic variability by mutation in a polygenic character with linked loci". Em: *Genetical Research* 26 (3), pp. 221–35.
- Leffler, E. M. et al. (2013). "Multiple Instances of Ancient Balancing Selection Shared Between Humans and Chimpanzees". Em: *Science* 339 (6127), pp. 1578–1582.
- Lenz, T. L., V. Spirin, D. M. Jordan e S. R. Sunyaev (2016). "Excess of Deleterious Mutations around HLA Genes Reveals Evolutionary Cost of Balancing Selection". Em: *bioRxiv*, pp. 1–30.
- Levene, H. (1953). "Genetic Equilibrium When More Than One Ecological Niche is Available". Em: *The American Naturalist* 87 (836), pp. 331–333.
- Lewontin, R. C. e J. L. Hubby (1966). "A Molecular Approach to the Study of Genic Heterozygosity in Natural Populations. II. Amount of Variation and Degree of Heterozygosity in Natural Populations of *Drosophila pseudoobscura*". Em: *Genetics* 54 (2), pp. 595–609.
- Lynch, M. (2007). "The evolution of genetic networks by non-adaptive processes." Em: *Nature reviews. Genetics* 8 (10), pp. 803–13.
- Maynard-Smith, J. e J. Haigh (1974). "The hitch-hiking effect of a favorable gene." Em: *Genetical Research* (23), pp. 23–35.
- McDonald, J. H. e M. Kreitman (1991). "Adaptive protein evolution at the *Adh* locus in *Drosophila*." en. Em: *Nature* 351 (6328), pp. 652–4.
- Mendes, F. (2013). *Natural selection on HLA and its effects on adjacent regions of the genome*. Rel. téc. Universidade de São Paulo. URL: <http://www.teses.usp.br/teses/disponiveis/41/41131/tde-02082013-161104/pt-br.php>.

- Meyer, D., R. M. Single, S. J. Mack, H. A. Erlich e G. Thomson (2006). "Signatures of demographic history and natural selection in the human major histocompatibility complex Loci." Em: *Genetics* 173 (4), pp. 2121–2142.
- Mitchell-Olds, T., J. H. Willis e D. B. Goldstein (2007). "Which evolutionary processes influence natural genetic variation for phenotypic traits?" Em: *Nature reviews. Genetics* 8 (11), pp. 845–856.
- Muller, H. J. (1950). "Our Load of Mutations". Em: *The American Journal of Human Genetics* 2 (2), pp. 111–176.
- Nielsen, R. (2005). "Molecular Signatures of Natural Selection". Em: *Annual Review of Genetics* 39 (1), pp. 197–218.
- Nielsen, R. et al. (2005). "A Scan for Positively Selected Genes in the Genomes of Humans and Chimpanzees". Em: *PLoS Biology* 3 (6), e170.
- Ohta, T. (1973). "Slightly Deleterious Mutant Substitutions in Evolution". Em: *Nature* 246 (5428), pp. 96–98.
- (1995). "Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory". Em: *Journal of Molecular Evolution* 40, pp. 56–63.
- Ohta, T. e J. H. Gillespie (1996). "Development of Neutral and Nearly Neutral Theories". Em: *Theoretical Population Biology* 49 (2), pp. 128–142.
- Oosterhout, C. van (2009). "A new theory of MHC evolution: beyond selection on the immune genes." Em: *Proceedings of the Royal Society of London. Series B, Biological Sciences* 276 (1657), pp. 657–65.
- Orr, H. A. (1998). "The population genetics of adaptation: the distribution of factors fixed during adaptive evolution." Em: *Evolution* 52 (4), pp. 935–949.
- (2005). "The genetic theory of adaptation: a brief history." Em: *Nature Reviews Genetics* 6 (2), pp. 119–27.
- Prout, T. (1968). "Sufficient Conditions for Multiple Niche Polymorphism". Em: *The American Naturalist* 102 (928), pp. 493–496.
- (2000). "How well does opposing selection maintain variation?" Em: *Evolutionary genetics: from molecules to morphology*. Cambridge: Cambridge University Press, pp. 157–181.

- Prugnolle, F., A. Manica, M. Charpentier, J. F. Guégan, V. Guernier e F. Balloux (2005). "Pathogen-driven selection and worldwide HLA class I diversity." Em: *Current Biology* 15 (11), pp. 1022–7.
- Richman, A. (2000). "Evolution of balanced genetic polymorphism". Em: *Molecular Ecology* 9 (12), pp. 1953–1963.
- Sabeti, P. C., S. F. Schaffner, B. Fry, J. Lohmueller, P. Varilly, O. Shamovsky, A. Palma, T. S. Mikelsen, D. Altshuler e E. S. Lander (2006). "Positive natural selection in the human lineage." Em: *Science* 312 (5780), pp. 1614–20.
- Sabeti, P. C. et al. (2007). "Genome-wide detection and characterization of positive selection in human populations." Em: *Nature* 449 (7164), pp. 913–8.
- Ségurel, L., Z. Gao e M. Przeworski (2013). "Ancestry runs deeper than blood: The evolutionary history of ABO points to cryptic variation of functional importance". Em: *BioEssays* 35 (10), pp. 862–867.
- Sella, G., D. A. Petrov, M. Przeworski e P. Andolfatto (2009). "Pervasive Natural Selection in the *Drosophila* Genome?" Em: *PLoS Genetics* 5 (6).
- Sellis, D., B. J. Callahan, D. a. Petrov e P. W. Messer (2011). "Heterozygote advantage as a natural consequence of adaptation in diploids". Em: *Proceedings of the National Academy of Sciences* 108 (51), pp. 20666–20671.
- Slade, R. e H. McCallum (1992). "Overdominant vs. frequency-dependent selection at MHC loci." Em: *Genetics* 132, pp. 861–864.
- Spurgin, L. G. e D. S. Richardson (2010). "How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings." Em: *Proceedings. Biological sciences / The Royal Society* 277 (1684), pp. 979–88.
- Tajima, F. (1989). "Statistical method for testing the neutral mutation hypothesis by DNA polymorphism." Em: *Genetics* 123 (3), pp. 585–595.
- Teixeira, J. C. et al. (2015). "Long-Term Balancing Selection in LAD1 Maintains a Missense Trans-Species Polymorphism in Humans, Chimpanzees, and Bonobos". Em: *Molecular Biology and Evolution* 32 (5), pp. 1186–1196.
- Tishkoff, S. A. e S. M. Williams (2002). "Genetic analysis of African populations: human evolution and complex disease." Em: *Nature Reviews Genetics* 3 (8), pp. 611–621.

- Trachtenberg, E. et al. (2003). "Advantage of rare HLA supertype in HIV disease progression". Em: *Nature Medicine* 9, pp. 928–935.
- Vallender, E. J. e W. E. Johnson (2008). "Balancing Selection in Human Evolution". Em: *eLS*.
- Watterson, G. A. (1978). "The homozygosity test of neutrality." Em: *Genetics* 88 (2), pp. 405–17.
- Williams, G. C. (1957). "Pleiotropy, Natural Selection, and the Evolution of Senescence". Em: *Evolution* 11 (4), p. 398.
- Wright, S. (1937). "The Distribution of Gene Frequencies in Populations." Em: *Proceedings of the National Academy of Sciences* 23 (6), pp. 307–320.
- Yang, Z. e W. J. Swanson (2002). "Codon-Substitution Models to Detect Adaptive Evolution that Account for Heterogeneous Selective Pressures Among Site Classes". Em: *Molecular Biology and Evolution* 19 (1), pp. 49–57.
- Zhang, Z. e J. Parsch (2005). "Positive correlation between evolutionary rate and recombination rate in *Drosophila* genes with male-biased expression." Em: *Molecular Biology and Evolution* 22 (10), pp. 1945–7.

Considerações Finais e Perspectivas

Aqui, eu recapitulo as questões que propus abordar na Introdução (página 49), resumo as conclusões a que chegamos com as investigações dos Capítulos 1 e 2, e discuto perspectivas decorrentes destes trabalhos.

Seleção balanceadora no genoma humano

Desenvolvimento e avaliação de um novo método para a detecção de assinatura de seleção balanceadora

No Capítulo 1, descrevemos um novo método para detecção de assinaturas de seleção balanceadora de longo prazo (SBLP) em humanos: *Non-Central Deviation* (NCD). Esse método apresenta duas estatísticas: *NCD1* utiliza apenas o espectro de frequências alélicas, ao passo que *NCD2* usa também informação contida na divergência entre humanos e chimpanzés. A combinação de duas assinaturas de SBLP em *NCD2* confere maior poder em relação a *NCD1*. Apesar disso, a performance de *NCD1* é comparável à de outros métodos comumente usados para detectar genes ou regiões sob seleção natural e recomendamos que seja utilizada em espécies para as quais dados de divergência com espécies pró-

ximas não estejam disponíveis.

Através de simulações neutras e com seleção, baseadas num modelo detalhado de demografia humana, demonstramos que o poder das duas estatísticas é alto para detectar assinaturas de SBLP em populações africanas e europeias, para sequências não muito longas (≤ 6.000 pares de base). Avaliamos qual a combinação de possíveis implementações que maximiza o poder das estatísticas e vimos que, para seleção que surgiu há pelo menos 3 milhões de anos, o método *NCD2* tem maior poder para sequências de 3.000 pares de base. Para seleção mais recente (isto é, que teve início há menos de 1 milhão de anos), *NCD1* tem poder maior que *NCD2*, mas como os valores em geral são baixos, na ordem de 30-40% (para taxa de falso positivo de 5%), enfatizamos que ambas as estatísticas são indicadas para a detecção de eventos de seleção balanceadora que perduram há pelo menos 3 milhões de anos (em humanos). Além disso, mostramos que a performance de *NCD2* supera a de outros métodos já existentes: *D* de Tajima (Tajima, 1989), teste HKA (Hudson et al., 1987), testes T1 e T2 (DeGiorgio et al., 2014), e uma combinação de *NCD1*+HKA.

Um diferencial das nossas estatísticas em relação às já existentes é que nela pode-se definir uma “frequência-alvo” (*target frequency*) a partir da qual o desvio de frequências alélicas é calculado. Isto é, a estatística pode ser calculada assumindo-se que os polimorfismos balanceados estejam segregando em frequências diferentes de 0.5. Na avaliação de poder, consideramos as frequências 0.3, 0.4 e 0.5. Com as simulações, vimos que o ganho em poder de *NCD1* e *NCD2* em relação às outras estatísticas é maior quando frequências de equilíbrio diferentes de 0.5 são simuladas. Assim, mostramos que *NCD2* tem poder maior do que as outras estatísticas usadas para detectar assinaturas de seleção balanceadora.

É importante ressaltar que o poder foi avaliado no contexto de um modelo

demográfico para populações humanas que é bastante complexo e realista (Gravel et al., 2011). Isso sugere que a observação de que nosso método supera os outros não é restrita a um cenário não-realista, mas baseada nos padrões de polimorfismo previstos em humanos.

Prevalência de SBLP no genoma humano

Há ainda considerável controvérsia sobre a importância da seleção balanceadora como processo microevolutivo que molda a diversidade genética humana (Andrés et al., 2009; Bubb et al., 2006; Leffler et al., 2013). Ele é raro, envolvendo poucas regiões genômicas? É mais comum atuar em regiões codificadoras de proteína ou regulatórias? Quais funções exercem os genes sob seleção balanceadora? O regime é partilhado entre populações distintas? O método que desenvolvemos tem como motivação contribuir para a resolução dessas questões.

As análises do Capítulo 1 mostram que, para humanos, a performance de *NCD2* é melhor do que a de *NCD1*. Assim, calculamos *NCD2* para janelas de 3.000 pares de bases ao longo de todo o genoma. Usamos dados genômicos de 4 populações (duas africanas, duas europeias) do Projeto 1000 Genomas. Como na prática não se sabe em qual frequência os polimorfismos balanceados estão segregando, calculamos *NCD2* considerando três frequências-alvo (0.3, 0.4, 0.5) e combinamos os resultados. Tomamos cuidado especial em filtrar os dados *a priori*, removendo regiões que poderiam ter assinaturas semelhantes às esperadas sob seleção balanceadora, mas por outras causas. Essas incluem regiões com motivos em *tandem*, grandes duplicações cromossômicas, regiões que não têm ortologia com chimpanzés e que não são únicas.

A fim de determinar os prováveis alvos de seleção balanceadora, combinamos duas estratégias: (1) um critério de significância baseado em simulação,

em que uma janela é considerada significativa se seu valor de *NCD2* para uma dada frequência-alvo é menor do que aquele de 10.000 simulações neutras com número igual de sítios informativos (resultando em cerca de 0,50% das janelas por população, considerando a união de todas as frequências-alvo) e; (2) um critério de *ranking* na distribuição genômica, após a aplicação de uma correção que leva em conta o número de sítios informativos da janela. Com o segundo critério, definimos como *outliers* as janelas na cauda da distribuição empírica (0,05%), que é basicamente um subconjunto das janelas obtidas com o primeiro critério.

Finalmente, reportamos como genes *outlier* aqueles que têm pelo menos uma janela *outlier* (independente da frequência-alvo) em pelo duas populações do mesmo continente. Com isso, esperamos reduzir os falsos positivos que poderiam ter surgido devido a alguma propriedade dos dados de uma certa população, dado que, na escala de tempo que investigamos, esperamos que populações de um mesmo continente tenham compartilhado pressões seletivas, bem como história demográfica. Nossos resultados mostraram que pelo menos 1% dos genes do genoma têm assinaturas extremas de seleção balanceadora (*outlier*), mas talvez mais, podendo chegar até 8% (Tabela S8, Capítulo 1). Mesmo a estimativa mais conservadora de 1% é bem mais alta do que o que já tinha sido observado até hoje. Por exemplo, apenas 0.4% dos 13.500 genes analisados por Andrés et al. (2009) apresentaram fortes assinaturas de seleção balanceadora. O fato de nossa estimativa ser mais alta é provavelmente decorrente de múltiplos fatores: o alto poder de *NCD2*, os dados genômicos utilizados, o fato de mesmo com todos os nossos filtros termos retido mais de 18.000 genes autossômicos nas análises e o fato de as janelas analisadas serem pequenas, o que aumenta a probabilidade de detectar uma assinatura de SBLP (Andrés, 2011; Charlesworth,

2009).

Nosso estudo pôde identificar genes com assinaturas extremas e reportou o quão prevalente a seleção balanceadora pode ter sido na história evolutiva humana. Dentre os 213 genes com assinaturas extremas de SBLP, 30% já foram detectados em algum *scan* prévio e outros (pelo menos quatro) estudos de genes candidatos. Ou seja, cerca de 70% dos genes que apresentamos são novos na literatura de seleção balanceadora. Adicionalmente, a nossa lista mais inclusiva (i.e, menos conservadora) de 1.470 genes com assinaturas menos extremas indica que talvez a SBLP tenha sido ainda mais comum.

Partilhamento entre continentes

Boa parte dos das janelas candidatas é compartilhada entre ao menos duas das populações analisadas (87%), particularmente entre populações do mesmo continente (78%). Mesmo nos casos em que um gene não passa o critério de pertencer aos dois continentes, a grande maioria tem assinaturas em ambos os continentes (ou seja, em pelo menos 3 das quatro populações analisadas), com raras exceções. Finalmente, cerca de 32% dos genes *outlier* (69 genes, Tabela 3, Capítulo 1) são partilhados entre as quatro populações.

Nossos achados confirmam que o grau de compartilhamento entre populações de um mesmo continente é maior do que entre populações de continentes distintos. Tal observação pode ser interpretada como um compartilhamento de pressões seletivas históricas, bem como de fatores demográficos em comum, que influenciam a variabilidade genética que fica disponível para a atuação da seleção balanceadora.

O fato de muitos dos alvos – genes e janelas – serem compartilhados entre continentes é compatível com a escala de tempo do regime seletivo que investi-

gamos (≥ 3 milhões de anos). Mesmo que, na história humana recente, África e Europa tenham divergido em diversos aspectos – em termos de história demográfica e de pressões seletivas – é plausível que muitos alvos de seleção balanceadora de longo prazo tenham sido mantidos em ambas, e/ou que tenham cessado de ser selecionados em um dos continentes apenas recentemente, preservando assim as assinaturas de SBLP até o presente.

Características das regiões candidatas

Resposta imune

Observamos um enriquecimento para certas categorias funcionais entre os genes significativos e *outliers*. Cerca de metade das categorias enriquecidas são relacionados à resposta imune, de forma ampla, e dessas, cerca de metade é diretamente ligada à apresentação de antígenos por moléculas HLA.

Evidências de seleção balanceadora em diversos genes HLA clássicos de classe I e II são abundantes na literatura. De fato, eles estão contidos nas janelas significativas e também nas *outlier*, que têm as assinaturas mais extremas. Portanto, investigamos se os genes HLA estariam causando os enriquecimentos de categorias relacionadas ao sistema imune. A remoção de tais genes levou à observação de que nenhuma categoria permaneceu enriquecida para os genes *outlier*, o que demonstra, em primeiro lugar, a grande influência dos genes HLA no conjunto mais restrito de genes candidatos e, em segundo lugar, que o conjunto de dados restante é pequeno (em média 177 genes por população), o que pode acarretar perda de poder para testes que visam detectar enriquecimento de alguma classe funcional entre os genes selecionados (mesmo categorias que não são compostas exclusivamente por genes HLA deixam de ser significativas com

a remoção dos mesmos).

Por outro lado, é interessante observar que mesmo após a remoção dos genes HLA clássicos, algumas categorias funcionais permaneceram enriquecidas para os genes significativos, algumas delas relacionadas ao sistema imune, mas envolvendo outros genes, incluindo genes HLA não-clássicos. De fato, 1/3 dos genes significativos são relacionados a funções imunes, mesmo que não compo-nham categorias enriquecidas. Entre as outras categorias, temos por exemplo “região extra-celular”, que confirma a observação de que tende a haver um ex-cesso de genes relacionados à matriz extracelular entre os alvos de SBLP em humanos (revisado em Key et al., 2014).

Corroboramos, assim, que a resposta imune é uma importante pressão sele-tiva responsável por instâncias de seleção balanceadora, e detectamos fortes as-sinaturas em alguns genes candidatos relacionados à reprodução. Cinco genes significativos são relacionados à espermatogênese, embora não haja enriqueci-mento para a categoria, e um dos 10 genes mais extremos (*C1orf101*) é altamente expresso em testículo e, embora tenha função ainda desconhecida, há indícios de que poderia estar relacionado ao complexo *CATSPER* de canais de Ca^{+2} , que são cruciais para a sinalização na superfície celular que leva à fertilização. Em suma, embora estas duas pressões seletivas de inegável importância (defesa do organismo e reprodução) não aparentam estar por trás da *maioria* dos alvos de SBLP, elas estão envolvidas em mais de 1/3 dos genes com assinaturas mais fortes de SBLP.

Confiabilidade acerca dos alvos de SBLP

Outra categoria de genes enriquecida entre os candidatos são os receptores ol-fatórios. Trata-se de uma família gênica complicada de se analisar pois são o

resultado de diversas duplicações. Nossas análises não permitem excluir as hipóteses de que: a) as assinaturas de SBLP nesses genes sejam causadas por conversão gênica entre parálogos situados próximos uns aos outros (trata-se de um fenômeno biológico, porém diferente de seleção balanceadora, capaz de gerar assinatura semelhante); b) que o excesso de SNPs com frequências intermediárias nesses genes seja decorrente de *reads* de genes distintos porém com alta identidade terem sido mapeados a uma só gene no genoma referência, assim inflando artificialmente a frequência de alelos em frequência intermediárias.

Seria plausível supor que ambos os artefatos – um deles causado por um fenômeno biológico, e o outro por problemas de bioinformática de dados de sequenciamento – poderiam estar ocorrendo de forma mais generalizada nos nossos genes candidatos. A fim de verificar a credibilidade das regiões candidatas quanto à questão de conversão gênica entre parálogos situados próximos um ao outro, comparamos a distribuição de número de casos em que parálogos por gene candidato estão situados no mesmo cromossomo (possibilitando, assim, a conversão não-homóloga), e comparamos com a distribuição para todos os outros genes. Vimos que as distribuições são essencialmente idênticas, e que portanto nossos genes candidatos não tendem a ter mais parálogos situados no mesmo cromossomo, de forma geral. Como a conversão gênica não-homóloga ocorre entre genes homólogos, a proximidade física é necessária.

A respeito de duplicações não detectadas, tomamos quatro dos 10 genes com assinaturas mais extremas (dentre os que nunca apareceram em outros estudos de seleção balanceadora) e verificamos que poucos SNPs contidos nesses genes podem ser artefatos gerados por duplicações não detectadas, e que mesmo excluindo tais SNPs, os genes em questão continuam tendo assinaturas extremas de SBLP. Finalmente, dos 213 genes com assinaturas mais extremas, apenas dois

são receptores olfatórios (Tabela 3, Capítulo 1), o que implica que: (1) é plausível que não sejam falsos positivos, dados todos os cuidados que tomamos, mas não podemos descartar essa possibilidade; (2) nossas verificações nos deixam confiantes de que vieses desse tipo não são uma característica dos genes candidatos de forma geral.

Regiões regulatórias versus regiões codificadoras de proteínas

Em um *scan* para polimorfismos balanceados compartilhados entre humanos e chimpanzés, Leffler et al. (2013) reportaram que, de 125 haplótipos compartilhados entre humanos e chimpanzés – interpretado como uma assinatura de SBLP – 123 ocorrem em regiões genômicas não-gênicas. Combinando-se essa observação com o fato de há poucos casos descritos de genes-alvo de SBLP, seria plausível supor que a maior parte dos sítios-alvo de seleção balanceadora fossem regulatórios. No nosso estudo, vimos que embora as janelas significativas representem apenas cerca de 0,5% das janelas analisadas, elas correspondem a cerca de 8% dos genes codificadores de proteínas. Por outro lado, não detectamos proporcionalmente mais janelas que incluem genes entre as significativas quando comparadas às não-significativas.

A fim de explorar se a SBLP tende a ocorrer sobre sítios regulatórios, investigamos se havia um excesso de SNPs com função regulatória nas janelas significativas. A princípio vimos que esse excesso – altamente significativo (Figura 7, Capítulo 1) – existe para SNPs que possuem diversas funções regulatórias, inclusive a de eQTL. Entretanto, SNPs sem anotação de eQTL mas com outras funções regulatórias não apresentam enriquecimento. Por fim, pudemos determinar que, considerando apenas SNPs com frequência intermediária, não existe enriquecimento para eQTLs, mostrando que uma anotação positiva para eQ-

TLs é correlacionada positivamente à frequências dos mesmos. Nosso achado mostra que, como há um excesso de variantes segregando em frequência intermediária em regiões sob seleção balanceadora, o enriquecimento de traços genômicos para os quais a detecção é sensível à frequência alélica (como é o caso de eQTLs) será enviesado. Finalmente, detectamos um excesso de SNPs sem qualquer anotação de função regulatória nas janelas candidatas.

Apesar dessa ausência de evidência de excesso de enriquecimento para SNPs com funções regulatórias entre as janelas mais extremas, detectamos um sutil, porém significativo, enriquecimento para expressão mono-alélica (MAE) entre os 213 genes com assinaturas mais extremas de SBLP (Savova et al., 2016). Um estudo recente (Savova et al., 2016) reportou que uma proporção considerável dos genes humanos (25%) apresentam expressão mono-alélica (MAE)². Eles reportam, ainda, que dentre os genes que têm assinaturas de SBLP, existe um enriquecimento de genes com assinatura MAE. Nós confirmamos essa relação com o nosso achado de excesso de genes MAE entre os genes mais extremos.

Trata-se de um achado que, conforme argumentado por Savova et al. (2016), pode indicar uma possível ligação evolutiva entre MAE e vantagem do heterozigoto: muitos dos genes MAE codificam proteínas expressas na superfície celular, e modulam interações entre a célula e o ambiente ao redor, incluindo outras células. Heterozigose em um sítio MAE poderia levar a diferentes alelos inativados em células de um mesmo tecido, diminuindo a possibilidade de uma “monocultura” e assim reduzindo a susceptibilidade do tecido como um todo a agentes infecciosos (Savova et al., 2016). Por outro lado, Savova et al. (2016)

²Para a maioria dos genes, em organismos diploides, acredita-se que a expressão gênica ocorre simultaneamente para os dois alelos. Para outros, apenas um dos alelos, o materno ou o paterno, é expresso, ao passo que o outro é inativado. Esse padrão é alcançado através de modificações epigenéticas, assim levando a uma expressão mono-alélica que é mantida ao longo das divisões mitóticas.

discutem que é inteiramente possível que expressão mono-alélica e manutenção de diversidade através de seleção sejam fenômenos independentes que têm como alvo os mesmos componentes moleculares.

Finalmente, para alguns alvos de SBLP já foram reportados casos em que uma variante causa uma mudança de tecido em que o gene é expresso. Como exemplo temos o gene *B4galnt2*: em camundongos, uma variante causa a mudança de expressão do local habitual (epitélio intestinal) para outro (endotélio vascular). O ortólogo desse gene em humanos (*B4GALNT2*) é um dos nossos genes candidatos, discutidos no Capítulo 1. Outro exemplo é o *HLA-G*, também entre os nossos candidatos e com uma ampla literatura descrevendo padrões complexos de expressão (p.ex. Tan et al., 2005). Assim, testamos se tais padrões são recorrentes entre nossos genes e detectamos um excesso significativo de genes com expressão em apenas um tecido humano: 12 com expressão na glândula adrenal e 25 com expressão no pulmão.

Em suma, muitos dos alvos de SBLP são genes codificadores de proteínas – a maioria nunca foi reportada antes em estudo de seleção balanceadora – e não encontramos evidência de excesso de funções regulatórias entre as janelas que não incluem genes. Por outro lado, encontramos enriquecimento para genes com MAE e com expressão tecido-específica, apontando que talvez haja, sim, um excesso de alvos de SBLP com funções regulatórias.

Variação deletéria em regiões e genes com assinaturas de SBLP

Além da seleção balanceadora e da seleção positiva, seleção contra mutações deletérias constitui um processo evolutivo fundamental, capaz de influenciar a variação quantitativa para caracteres de importância ecológica e médica. Com o influxo constante de novas mutações deletérias que surgem nas populações, algumas irão segregar transitoriamente dentro das populações, resultando num balanço entre mutação e seleção que é influenciado pela taxa de mutação, pelo tamanho populacional efetivo e pela intensidade de seleção sobre a mutação. Entretanto, a contribuição de tais variantes deletérias sobre caracteres moldados por variação genética quantitativa permanece pouco compreendido (Mitchell-Olds et al., 2007). Diversos estudos de associação em humanos têm identificado polimorfismos segregando em frequências intermediárias que influenciam variação de traços complexos (Mitchell-Olds et al., 2007)³.

No Capítulo 2, mostramos que genes com assinaturas extremas de seleção balanceadora têm maior carga genética do que regiões evoluindo presumivelmente de forma neutra. Os controles levaram em conta o fato de que o espectro de frequências alélicas dos genes balanceados tem proporcionalmente menos variantes raras do que o controle genômico. Usamos três métricas diferentes para quantificar este excesso: duas delas contam diretamente o número de variantes potencialmente deletérias dividido pelo número de variantes neutras, e a outra atribui uma medida para cada variante, que quantifica o quão dele-

³Aqui, refiro-me a traços que, acredita-se, resultam de variação genética em múltiplos genes e suas interações com fatores ambientais e comportamentais (Mitchell-Olds et al., 2007).

téria ela é. Assim, as distribuições dessas medidas para genes balanceados e controles pôde ser comparada.

As três estimativas são mais elevadas para os genes balanceados do que para os controles, com poucas exceções. Mais ainda, quando removemos os genes HLA – que têm muitos sítios mantidos de forma adaptativa e poderiam confundir a interpretação das estimativas – os resultados foram qualitativamente semelhantes. Avaliamos, por fim, o impacto que os sítios potencialmente selecionados nos genes balanceados têm sobre essas estimativas, e vimos que as observações se mantêm mesmo quando eles são removidos.

Em suma, há evidência, através de três diferentes métricas, de um excesso de carga genética na vizinhança de regiões com assinaturas de seleção balanceadora. Esse resultado pode ser interpretado de duas formas: (1) como uma evidência de *sheltered load*⁴; ou (2) como evidência de efeito carona das variantes deletérias com os polimorfismos balanceados, conforme explicado na Figura 1 do Capítulo 2.

Nossos resultados não permitem escolher entre uma ou outra explicação. Entretanto, Lenz et al., 2013 mostrou, através de simulações de genes HLA com múltiplos sítios selecionados e suas regiões adjacentes, que mesmo em um modelo aditivo (não-recessivo), espera-se um aumento da carga genética em regiões adjacentes aos genes HLA, e tal efeito diminui quanto maior é a distância em relação aos genes. Se extrapolarmos essas observações para outros genes sob seleção balanceadora, é plausível supor que o mesmo ocorre na vizinhança de outros alvos de seleção balanceadora.

A fim de discernir entre esses dois possíveis cenários, uma opção seria : (1)

⁴A ideia de que variantes deletérias recessivas raramente estarão em homozigose quando estão nos genes HLA, pois a região tem alta heterozigose. Assim, tais variantes deletérias estariam protegidas da seleção purificadora (Oosterhout, 2009).

verificar com simulações se sob modelo de seleção balanceadora não com múltiplos, mas apenas um, sítio selecionado, os mesmo padrões são observados e; (2) se existe um excesso de associações a doenças nas regiões genômicas dos genes sob seleção balanceadora; (3) se o excesso de carga genética é menor (mas ainda significativo) para genes vizinhos aos genes balanceados e/ou fixando-se janelas genômicas em torno dos genes e verificando se a carga genética diminui com a distância em relação ao gene-alvo.

Ainda que permaneçam algumas questões em aberto, nosso trabalho é uma contribuição para dois campos estimulantes da biologia evolutiva: o estudo do acúmulo de mutações deletérias no genoma humano e o estudo da importância evolutiva da seleção balanceadora para a evolução humana.

Perspectivas

Conciliando assinaturas de seleção e fenótipos

“(...) genome-wide scans are a hatchet, whereas what we need now is a scalpel. In-depth follow-up studies of individual outlier loci can be one such scalpel, more precisely defining important population genetic parameters such as the timing and magnitude of selection, the geographic distribution of selected variation, the interaction of population demographic history, recombination, and selection in shaping patterns of variation, and the functional form of selection acting on individual outlier loci” (Akey, 2009)

A rigor, evidências de evolução adaptativa não demonstram que uma dada substituição ou polimorfismo é adaptativo ao nível fenotípico, mas indicam a região onde ele provavelmente poderá ser encontrado. Estudos baseados em genética de populações são capazes de identificar genes alvo de seleção, i.e., que evoluíram de forma não-neutra ao longo da história evolutiva humana (Capítulo 1), mas não são capazes de fornecer, por si só, informações acerca dos traços fenotípicos que representam os verdadeiros alvos de seleção (Mitchell-Olds et al., 2007).

Até o momento, em muito poucos casos conseguiu-se traçar a relação causal entre um polimorfismo e um fenótipo de interesse, pois, tanto na pesquisa quanto na prática clínica, a capacidade de detectar variantes genéticas suplanta, em muito, a habilidade de sistematicamente avaliar os potenciais efeitos de tais variantes (Kircher et al., 2014). Mesmo havendo essa enorme defasagem, com a publicação de novos catálogos de genes/regiões genômicas candidatas à ação da seleção balanceadora, ensaios funcionais têm se tornado mais comuns.

Por exemplo, em um estudo elegante, Chakraborty e Fry (2015) mostraram como um polimorfismo em um gene pleiotrópico – codificador da enzima aldeído-desidrogenase – é ativamente mantido devido a diferenças no nível de concentração alcóolica em frutas em ambientes diversos ocupados por *Drosophila*. A enzima tem duas funções: metabolismo de etanol e de outros aldeídos decorrentes da fosforilação oxidativa, sendo esta a provável função ancestral e aquela a função derivada. As duas variantes têm aptidões diferentes em diferentes habitats, dependendo do regime alimentar da mosca. Os autores conseguiram identificar uma substituição de aminoácido responsável pelas duas variantes da enzima, e verificaram a eficácia das duas variantes sobre diferentes substratos, assim revelando a aptidão de cada variante em dois tipos de ambientes.

Um outro exemplo é o do gene *ERAP2*, que codifica uma proteína envolvida na via de apresentação de antígenos pelas moléculas de MHC classe I. Esse gene apresenta assinaturas de SBLP de acordo com nosso estudo (Tabela S8, Capítulo 1) e já tinha sido revelado como candidato por Andrés et al. (2009). Em um estudo posterior (Andrés et al., 2010) foi demonstrado que a seleção balanceadora mantém dois haplótipos, A e B, segregando em frequências intermediárias, e que um deles resulta em uma proteína truncada. O estudo mostra, ainda, que homozigotos para esse haplótipo resultam em expressão reduzida de moléculas de MHC de classe I na superfície de linfócitos T. Apesar de a pressão seletiva para a manutenção dessa variante ser ainda desconhecida, o estudo mostrou evidências bioinformáticas, moleculares, celulares e imunológicas que mostram que o gene pode ter sofrido seleção balanceadora, o impacto do provável sítio selecionado sobre a proteína, e uma consequência *downstream* dessa variação para a apresentação de antígenos.

Ainda que elucidar a relação causal entre genótipo e fenótipo como nos exemplos acima esteja além do escopo do presente trabalho, demos importantes passos nessa direção ao explorar propriedades das regiões candidatas. No Capítulo 1, dentro dessas limitações, buscamos explorar a base biológica dos alvos de seleção balanceadora, ao olharmos para as categorias funcionais às quais eles pertencem, para a proporção de sítios codificadores, e dentre esses, os sítios não-sinônimos. No Capítulo 2, analisamos em maior detalhe as propriedades dos sítios contidos nas regiões-alvo de seleção balanceadora. Assim, pudemos testar hipóteses acerca do acúmulo de mutações deletérias em regiões sob seleção balanceadora e aprofundamos nossa compreensão acerca dos potenciais alvos de seleção balanceadora no genoma humano.

Acreditamos que com o ressurgimento de interesse por alvos de seleção balanceadora em humanos na literatura, muitos dos genes candidatos levantados no nosso trabalho serão alvo de investigação mais detalhada tanto acerca de padrões genômicos como acerca de possíveis efeitos fenotípicos e mutações causais em estudos funcionais.

Potencial das estatísticas NCD em futuros estudos

No Capítulo 1, mostrei que as duas novas estatísticas que propusemos – *NCD1* e *NCD2* – têm poder elevado em relação a outras estatísticas comumente usadas para a detecção de assinaturas de seleção balanceadora.

Uma limitação no que tange a extrapolação de nossas observações sobre o poder das estatísticas NCD para outras espécies é que as análises de poder requerem simulações – neutras e com seleção – cujos parâmetros podem variar muito entre espécies. Por outro lado, o trabalho do Capítulo 1 deixa em aberto a possibilidade de que *NCD1* e *NCD2* sejam utilizados em outras espécies, dada

a sua extrema facilidade de implementação e interpretação de seus resultados. As simulações para outras espécies são necessárias no sentido de determinar o poder da estatística para o cenário em questão, e também para definir filtros adequados, como os que propusemos na extensa parte de métodos do trabalho.

Como exemplo, Teixeira e colaboradores (in prep)⁵ têm trabalhado em um estudo que discute as potenciais implicações biológicas de alvos de seleção balanceadora nos “grandes símios”⁶. Tal estudo tem utilizado as estatísticas NCD, valendo-se de modelos demográficos específicos e detalhados para as espécies em questão para avaliar o poder nesses cenários, bem como os filtros apropriados. Essa aplicação pra outras espécies mostra o potencial das nossas estatísticas de serem utilizadas por geneticistas evolutivos interessados em assinaturas de seleção que afetem o espectro de frequências alélicas.

⁵Sou co-autora deste trabalho.

⁶*Great Apes*, incluindo chimpanzé, bonobo, gorila e orangotango.

Bibliografia

- Akey, J. M. (2009). "Constructing genomic maps of positive selection in humans: where do we go from here?" Em: *Genome Research* 19 (5), pp. 711–722.
- Andrés, A. M. (2011). "Balancing Selection in the Human Genome". Em: *eLS*, pp. 1–8.
- Andrés, A. M. et al. (2009). "Targets of balancing selection in the human genome." Em: *Molecular Biology and Evolution* 26 (12), pp. 2755–64.
- Andrés, A. M. et al. (2010). "Balancing Selection Maintains a Form of ERAP2 that Undergoes Nonsense-Mediated Decay and Affects Antigen Presentation". Em: *PLoS Genetics* 6 (10), e1001157.
- Bubb, K. L. et al. (2006). "Scan of human genome reveals no new Loci under ancient balancing selection." Em: *Genetics* 173 (4), pp. 2165–77.
- Chakraborty, M. e J. D. Fry (2015). "Evidence that Environmental Heterogeneity Maintains a Detoxifying Enzyme Polymorphism in *Drosophila melanogaster*". Em: *Current Biology* 26 (2), pp. 1–5.
- Charlesworth, B. (2009). "Effective population size and patterns of molecular evolution and variation." Em: *Nature reviews. Genetics* 10 (3), pp. 195–205.
- DeGiorgio, M., K. E. Lohmueller e R. Nielsen (2014). "A model-based approach for identifying signatures of ancient balancing selection in genetic data." Em: *PLoS genetics* 10 (8), e1004561.
- Gravel, S., B. M. Henn, R. N. Gutenkunst, A. R. Indap, G. T. Marth, A. G. Clark, F. Yu, R. A. Gibbs e C. D. Bustamante (2011). "Demographic history and rare allele sharing among human populations." Em: *Proceedings of the National Academy of Sciences of the United States of America* 108 (29), pp. 11983–8.
- Hudson, R. R., M. Kreitman e M. Aguade (1987). "A Test of Neutral Molecular Evolution Based on Nucleotide Data". Em: *Genetics* 116 (1), pp. 153–159.

- Key, F. M., J. C. Teixeira, C. de Filippo e A. M. Andrés (2014). "Advantageous diversity maintained by balancing selection in humans". Em: *Current Opinion in Genetics & Development* 29, pp. 45–51.
- Kircher, M., D. M. Witten, P. Jain, B. J. O’Roak, G. M. Cooper e J. Shendure (2014). "A general framework for estimating the relative pathogenicity of human genetic variants". Em: *Nature genetics* 46 (3), pp. 310–315.
- Leffler, E. M. et al. (2013). "Multiple Instances of Ancient Balancing Selection Shared Between Humans and Chimpanzees". Em: *Science* 339 (6127), pp. 1578–1582.
- Lenz, T. L., B. Mueller, F. Trillmich e J. B. W. Wolf (2013). "Divergent allele advantage at MHC-DRB through direct and maternal genotypic effects and its consequences for allele pool composition and mating". Em: *Proceedings of the Royal Society B: Biological Sciences* 280 (1762), p. 20130714.
- Mitchell-Olds, T., J. H. Willis e D. B. Goldstein (2007). "Which evolutionary processes influence natural genetic variation for phenotypic traits?" Em: *Nature reviews. Genetics* 8 (11), pp. 845–856.
- Oosterhout, C. van (2009). "A new theory of MHC evolution: beyond selection on the immune genes." Em: *Proceedings of the Royal Society of London. Series B, Biological Sciences* 276 (1657), pp. 657–65.
- Savova, V., S. Chun, M. Sohail, R. B. McCole, R. Witwicki, L. Gai, T. L. Lenz, C.-t. Wu, S. R. Sunyaev e A. A. Gimelbrant (2016). "Genes with monoallelic expression contribute disproportionately to genetic diversity in humans". Em: *Nature Genetics* 48 (3), pp. 231–237.
- Tajima, F. (1989). "Statistical method for testing the neutral mutation hypothesis by DNA polymorphism." Em: *Genetics* 123 (3), pp. 585–595.
- Tan, Z., A. M. Shon e C. Ober (2005). "Evidence of balancing selection at the HLA-G promoter region". Em: *Human Molecular Genetics* 14 (23), pp. 3619–3628.