UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

INSTITUTO DE BIOCIÊNCIAS

PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

Explorando Epistasia Adaptativa em Humanos e Outros Primatas

Bruna Oliveira Missaggia

Tese submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS como requisito parcial para a obtenção do grau de Doutora em Genética e Biologia Molecular

Orientadora: Profa. Dra. Maria Cátira Bortolini

Coorientador: Prof. Dr. Márcio Dorn

Porto Alegre, 2024

Esta tese foi realizada no Laboratório de Evolução Humana e Molecular (LEHM) do Departamento de Genética, Instituto de Biociências da Universidade Federal do Rio Grande do Sul, no Laboratório de Laboratório de Bioinformática Estrutural e Biologia Computacional (SBCB) do departamento de informática da Universidade Federal do Rio Grande do Sul e no Laboratory of Computational and Quantitative Biology (LCQB) da Sorbonne Université. Este trabalho foi subsidiado por recursos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). A bolsa de doutorado foi concedida por meio do Programa de Pós-Graduação em Genética e Biologia Molecular (PPGBM – UFRGS) e através do programa CAPES STIC/AMSUD (Cooperação em Ciência e Tecnologia da Informação e da Comunicação França-América do Sul - CAPES/CDEFI).

Agradecimentos

Agradeço à minha orientadora, Profa. Dra. Maria Cátira Bortolini, pelo empenho com que me ajudou em todas as etapas do desenvolvimento desta pesquisa. A ela também sou grata pelos inúmeros ensinamentos e por toda a generosidade científica e humana que demonstra ao conduzir o LEHM.

Ao meu coorientador, Prof. Dr. Márcio Dorn, por compartilhar seu conhecimento de maneira tão dedicada e inspiradora. Agradeço a ele também por disponibilizar a infraestrutura computacional do SBCB, e pela valiosa oportunidade que me deu ao convidar para fazer meu estágio doutoral na Sorbonne.

À Profa. Dra. Juliana Bernardes, por ter me recebido de forma acolhedora na França e pela orientação nos trabalhos que desenvolvemos por lá.

Agradeço aos colegas do LEHM, do SBCB e do LCQB pela convivência agradável. Sou grata à Dra. Bibiana Fam, ao Dr. Pedro Vargas-Pinilla e ao Lucca Fanti por colaborarem nesta pesquisa.

Aos funcionários e professores do PPGBM, pela competência e ensinamentos. Especialmente ao Elmo Cardoso, por ser sempre tão gentil e solícito, e ao professor Dr. Aldo Mellender Araújo pela orientação durante a graduação.

Aos meus familiares e amigos agradeço o suporte e incentivo sempre. Em especial à minha mãe, pelo apoio, e aos meus irmãos pela cumplicidade. Ao meu pai, *in memoriam*, por tudo dele que vive em mim e naqueles que tiveram o privilégio de conhecê-lo.

À CAPES, pela concessão da bolsa de pesquisa.

RESUMO

A presente tese investigou a relação entre genes e fenótipos adaptativos em humanos e outros primatas, por meio de dois estudos independentes, um de natureza macroevolutiva e o outro em escala microevolutiva. Ambos os trabalhos visam identificar a epistasia adaptativa, utilizando a metáfora da paisagem de aptidão para compreender os padrões observados nos "mapas genótipo-fenótipo adaptativos". O estudo microevolutivo partiu da "hipótese da vitamina D-folato", a qual propõe a radiação ultravioleta como a principal pressão seletiva para a coloração da pele humana. Segundo esta hipótese, não seria apenas por meio da variação na cor de pele que as populações se adaptam aos seus ambientes. Assim, a cor de pele diferente do esperado em nativos americanos não significa necessariamente que essas populações não estejam adaptadas aos seus ambientes. Tais diferenças poderiam ser atribuídas à complexidade da paisagem de aptidão em relação à radiação, o que envolve diferentes vias e a possível coadaptação de alelos entre elas. Ao testar a hipótese, identificamos redes de alelos em duas rotas genéticas relacionadas à adaptação ao ambiente de radiação — coloração da pele e metabolismo da vitamina D — em populações nativas americanas. As possíveis relações epistáticas que diferem entre grupos que vivem em condições ambientais diferentes podem ajudar a explicar a exceção do padrão de variação de cor nessas populações. O estudo macroevolutivo empregou técnicas de Inteligência Artificial para desenvolver o método ProteinPhenotypeInsights (ProPhIn). Este método enfoca a interpretabilidade e visa auxiliar biólogos evolutivos na compreensão das relações entre proteínas candidatas e fenótipos categóricos de interesse em espécies relacionadas. O ProPhIn tem dois objetivos principais: selecionar variações genéticas associadas ao fenótipo e interpretar os padrões gerais dessas variações em relação ao fenótipo e à relação entre as variações selecionadas e as espécies. Ao aplicar o ProPhIn aos dados de 64 espécies de primatas para genes candidatos do sistema oxitocinérgico, identificamos variações genéticas estatisticamente associadas aos fenótipos monogamia social, cuidado biparental e tamanho de ninhada. Além disso, pudemos visualizar os padrões de variação genética e formular hipóteses sobre a relação entre as espécies, as redes de atributos genéticos e os fenótipos na paisagem adaptativa.

ABSTRACT

The present thesis investigated the relationship between genes and adaptive phenotypes in humans and other primates through two independent studies, one of macroevolutionary nature and the other on a microevolutionary scale. Both endeavors aimed to identify adaptive epistasis, utilizing the fitness landscape metaphor to comprehend the patterns observed in adaptive genotype-phenotype maps. The microevolutionary study was grounded on the "vitamin D-folate hypothesis," proposing ultraviolet radiation as the primary selective pressure for human skin coloration. According to this hypothesis, population adaptation to environments cannot be solely explained by skin color variation. Thus, unexpected skin color in Native American populations does not necessarily indicate lack of adaptation to their environments. These differences could be attributed to the complexity of the fitness landscape concerning radiation, which entails considering different pathways of phenotypic adaptation and the potential co-adaptation of alleles among them. Testing this hypothesis, we identified allele networks in two genetic pathways related to radiation environment adaptation - skin coloration and vitamin D metabolism - in Native American populations. Epistatic relationships differing among groups living in different environmental conditions may help explain deviations from the color variation pattern in these populations. The macroevolutionary study employed Artificial Intelligence techniques to develop the ProteinPhenotypeInsights (ProPhIn) method. This method focuses on interpretability and aims to assist evolutionary biologists in understanding the relationships between candidate proteins and categorical phenotypes of interest in related species. ProPhIn has two primary objectives: selecting genetic variations associated with the phenotype and interpreting the general patterns of these variations concerning the phenotype and the relationship between the selected variations and the species. Applying ProPhIn to data from 64 primate species for candidate genes related to the oxytocinergic system, we identified genetic variations statistically associated with social monogamy, biparental care, and litter size phenotypes. Additionally, we visualized patterns of genetic variation and formulated hypotheses about the relationship between species, genetic attribute networks, and phenotypes in the adaptive landscape.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE SIGLAS E ABREVIATURAS

Ala^{8}OXT - Oxitocina Alanina 8

AVP - Vasopressina

CYTB- Citocromo b

FDR- *False-discovery rate*

GPCRs - Receptores acoplados às proteínas G

IRAP- Aminopeptidase regulada por insulina

Leu^{8}OXT- Oxitocina Leucina 8

LNPEP- Leucil-cistinil aminopeptidase

MA- Milhões de anos atrás

mRNA- RNA mensageiro

OXT - Oxitocina

OXTR - Receptor da Oxitocina

PCA- Análise de componentes principais

Phe^{2}OXT - Oxitocina Fenilalanina 2

P-LAP- Leucina aminopeptidase placentária

Pro^{8}OXT - Oxitocina Prolina 8

SNPs- *Single-nucleotide polymorphism*

Thr^{8}OXT - Oxitocina Treonina 8

t-SNE- *t-distributed Stochastic Neighbor Embedding*

UV- Radiação ultravioleta

UVMED- dose mínima de eritema

Val^{3}Pro^{8}OXT - Oxitocina Valina 3 Prolina 8

VDR - Receptor de vitamina D

SUMÁRIO

CAPÍTULO I: INTRODUÇÃO

Um dos maiores desafios que as Ciências Biomédicas têm na atualidade é conseguir conectar variações em nível de genoma com aquelas observadas em nível de fenótipos. A complexidade aumenta quando o chamado mapa genótipo-fenótipo (Alberch, 1991) está ligado à aptidão, ou ainda à uma "paisagem adaptativa", conceito que busca demonstrar como a seleção natural direciona a evolução biológica (De Visser & Krug, 2014; Bank et al., 2022).

A ideia de paisagem de aptidão foi visualizada de forma icônica por Sewall G. Wright ainda na primeira metade do século passado (Wright, 1932). O notável geneticista norte-americano vislumbrou uma paisagem adaptativa como um mapa topográfico com picos e vales, na qual populações evoluem em direção a caminhos que aumentam sua aptidão. Em outras palavras, o modelo tenta explicar como indivíduos de uma população seriam capazes de sobreviver e de se reproduzir ao longo do tempo evolutivo por meio de mutações benéficas, até atingirem um pico de aptidão em determinado contexto. Mudando o contexto e as pressões seletivas, novos picos adaptativos se impõem. Além disso, uma paisagem de aptidão irregular deve possuir múltiplos picos adaptativos que representam soluções genéticas estáveis, mas diferentes, às vezes sub-ótimas, para o desafio de sobreviver (e prosperar ao longo das gerações) em um ambiente específico.

A relação entre espaço genotípico e aptidão veio da ideia de que o valor adaptativo de uma mutação depende do contexto genético em que ela ocorre (Wright, 1931; 1932), já que o autor acreditava na existência de uma epistasia generalizada. O termo epistasia é polissêmico (conforme detalhadamente discutido em Phillips, 1998; 2008), sendo usado para descrever diferentes fenômenos, incluindo a interação funcional entre produtos dos genes, o resultado genético de mutações que atuam na mesma via e o desvio estatístico da ação aditiva (Phillips, 2008). Em um conceito abrangente, a epistasia pode ser definida como qualquer tipo de interação genética em que os efeitos mutacionais dependam do *background* genético (De Visser & Krug, 2014). Existem diversos subtipos ou classes de epistasia dependendo do seu efeito específico, diferentes definições matemáticas e

distintas metodologias para detectar o fenômeno. Domingo (et al., 2019) e Mackay e Anholt (2024) apresentam ótimas revisões sobre estes assuntos.

A epistasia é altamente dependente do equilíbrio entre a eficiência da seleção, a intensidade da deriva e a taxa de mutação (Gros et al., 2009; Sydykova et al., 2020). A variância epistática para a aptidão é a base da teoria da mudança do equilíbrio de Wright, que sugere como ocorrem os movimentos de populações ao longo das paisagens adaptativas considerando os efeitos desses fatores evolutivos (Mackay & Anholt, 2024). Em termos gerais, a epistasia adaptativa é definida como a interação entre alelos de diferentes *loci* em seus efeitos na aptidão (Da Silva et al, 2010). De modo simplificado, ela pode ser entendida como os efeitos não aditivos da interação entre produtos de diferentes genes (intergênico), ou ainda da interação entre aminoácidos dentro de proteínas (intragênico) que resultam em desfechos evolutivos que terão impacto nas populações e, consequentemente, na evolução a longo prazo.

A ideia de paisagem adaptativa e relação entre epistasia e fenômenos evolutivos vêm sendo expandida desde os trabalhos pioneiros de Wright. O paleontólogo estadunidense George Simpson e o geneticista russo Theodosius Dobzhansky extrapolaram o conceito de paisagem de aptidão, criada para descrever dinâmicas populacionais (microevolutivas), expandindo a metáfora para o nível macroevolutivo, que representa relações entre espécies (Huneman, 2017). Nesse contexto, cada espécie ocuparia um pico diferente no campo das combinações genéticas. Em paisagens macroevolutivas os vales adaptativos estariam vazios, já que os fenótipos são muito pouco adaptativos. Neste relevo, espécies de nichos ecológicos semelhantes, como o leão e o tigre, ocupariam picos adjacentes, os quais se separariam, por profundos vales, dos picos que ocupam roedores ou primatas, por exemplo (Dobzhansky, 1982). A epistasia também está associada com a especiação, já que geraria as chamadas incompatibilidades de Dobzhansky-Muller. Nesse caso, defende-se que alelos geneticamente neutros ou benéficos na origem de duas populações que divergem passam a ser deletérios em híbridos dessas populações (Orr & Turelli, 2001).

O papel da epistasia adaptativa em contexto macroevolutivo é bem conhecido. Dungan e Chang (2017), por exemplo, levantaram a hipótese de que a

epistasia intramolecular ajudou a conservar as propriedades cinéticas do pigmento visual de pouca luz, a rodopsina, ao mesmo tempo em que permitiu substituições de sintonia espectral com deslocamento para o azul, à medida que os cetáceos se adaptavam ao ambiente aquático. A sinalização é feita através da interação entre pigmento e receptor acoplado à proteína G (GPCR), de modo que os autores descobriram que substituições específicas de aminoácidos nesta molécula promoviam compensações entre os diferentes aspectos funcionais do sistema.

Por outro lado, a epistasia costuma ser negligenciada em nível microevolutivo. Isso acontece porque alguns autores consideram o efeito epistático transitório e pouco importante nesses contextos (Whitlock et al, 1995), além de haver dificuldades técnicas inerentes aos métodos estatísticos que costumam ser usados em dados populacionais de alta dimensionalidade (Pośpiech et al., 2014). Os estudos de GWAS (*Genome-Wide Association Studies*) medem individualmente a associação entre cada variação genética e um fenótipo. Trabalhos de seleção genética populacional costumam assumir que o efeito de uma mutação selecionada é independente do contexto genético em que ela ocorre, e na genética quantitativa clássica a epistasia tende a aparecer como um termo de ruído (Walsh e Lynch, 2018; Tam et al., 2019; Mackay & Anholt, 2024). Porém, cada vez mais os estudos têm mostrado que a epistasia adaptativa abrange todos os níveis biológicos de organização: dentro ou entre genes, dentro de uma população, entre hospedeiro e simbionte, e entre espécies, de modo que o mapa genótipo-fenótipo adaptativo não pode ser completamente entendido sem levar esse fenômeno em consideração (Bank, 2022).

Grande parte do que se conhece sobre o papel da epistasia em nível populacional veio de estudos com organismos modelo, como *Drosophila melanogaster*, *Saccharomyces cerevisiae* e *Escherichia coli*, devido a possibilidade de manipulação experimental (Domingo et al., 2019; Mackay e Anholt, 2024). Diversos estudos mostraram evidências de efeitos fenotípicos diferentes da mesma mutação em diferentes *backgrounds* genéticos, tanto em *S. cerevisiae* quanto em *D. melanogaster* (Gibson & Dworkin, 2004; Dworkin et al., 2009; Yamamoto et al., 2009; Chandler et al., 2017; Galardini et al., 2019). Ainda, uma investigação com mutações introduzidas na enzima isopropilmalato desidrogenase (IMDH) da bactéria

Escherichia coli mostrou sinal de epistasia adaptativa. A IMDH está envolvida na biossíntese de leucina e pode utilizar duas coenzimas alternativas, e o uso de uma delas confere mais aptidão do que o uso da outra. As mutações foram feitas na região que controla o uso das coenzimas por IMDH, alterando a ligação. Os resultados indicam que apesar de cada mutação contribuir de forma aditiva em relação ao uso da coenzima, os efeitos são epistáticos em termos de aptidão (Lunzer et al., 2015). Em estudos com humanos e outras espécies não modelo, as quais não se pode controlar experimentalmente, é muito difícil detectar o fenômeno.

Atualmente é bem conhecido que a epistasia adaptativa determina quais sequências de mutações funcionais consecutivas são viáveis, uma vez que a existência dessas interações limita as rotas disponíveis para que seja atingido um pico adaptativo (Ferretti et al., 2016; Fragata et al., 2019; Østman et al., 2012). Ou seja, uma mudança benéfica para uma característica ou função pode resultar em uma mudança prejudicial para outro traço. Portanto, a seleção natural frequentemente prioriza certas características vantajosas a despeito de outras (Guillaume e Otto, 2012). Esses *trade-offs* (Domingo et al., 2019) são comuns na evolução porque os processos biológicos não ocorrem isoladamente, mas fazem parte dos sistemas altamente integrados e hierárquicos que constituem os organismos vivos. Além disso, os recursos de um organismo são limitados e, muitas vezes, não é possível atingir a adaptação perfeita em todos os fenótipos ao mesmo tempo. Desse modo, a construção dessas paisagens adaptativas requer a seleção das variantes genéticas envolvidas no sinal epistático (Yoo et al., 2020).

Outro fator que pode complexificar o entendimento da relação genótipo-fenótipo em uma paisagem adaptativa é a pleiotropia. Este conceito descreve a capacidade de um único gene ou alelo de influenciar múltiplos fenótipos em um organismo. Em uma definição mais rigorosa, a pleiotropia ocorre quando os efeitos aditivos e/ou de dominância de uma variante polimórfica são diferentes de zero para duas ou mais características (Mackay & Anholt, 2024). Genes do sistema oxitocinérgico seriam casos notáveis. Por exemplo, *OXT* codifica o nonapeptídeo oxitocina que é considerado um neuro-hormônio, uma vez que atua tanto como neurotransmissor no sistema nervoso central, onde é produzido, quanto como hormônio, pois também é liberado na corrente sanguínea e tem um efeito periférico

relacionado a processos fisiológicos, como a contração uterina e ejeção do leite em mamíferos placentários (Zingg e Laporte, 2003; Vargas-Pinilla et al., 2015; Fam et al., 2024). Sua presença no cérebro, por sua vez, modula comportamentos sociais complexos (*e.g.* formação de casais, cuidados parentais, ver Lucion e Bortolini, 2014; French et al., 2016; Mustoe et al., 2018 e referências lá contidas). A atuação de OXT depende de sua adequada interação com seu receptor preferencial OXTR, caracterizado como uma GPCR, já que atua na cascata de sinalização acoplado a proteínas G (Bockaert e Pin, 1999; Bockaert et al., 2004).

A pleiotropia tem sido considerada um elemento importante para a epistasia (de Visser et al., 2011), na medida que a última necessariamente se torna mais complexa, pois a variação das características também depende frequentemente de interações dos produtos gênicos. Por outro lado, a epistasia também forneceria a variação genética necessária para a evolução da pleiotropia, visto que duas mutações em um gene podem ter menor grau pleiotrópico quando isoladas do que quando juntas (Polster et al., 2016). A epistasia pode reestruturar as correlações genéticas entre características ao combinar padrões de covariação de características favorecidas pela seleção (Jones et al., 2014) e também aliviar as restrições impostas pela pleiotropia (Polster et al., 2016). Desse modo, os efeitos da epistasia podem ser ainda mais importantes do que quando incidem sobre um traço único, alterando os caminhos evolutivos para picos de aptidão mais elevados (Weinreich et al. 2006; Poelwijk et al. 2007; Franke et al. 2011; Polster et al., 2016). Assim, os dois fenômenos estariam correlacionados em um ciclo de influência mútua, moldando paisagens de aptidão e as trajetórias evolutivas das espécies.

Em resumo, a pleiotropia e a epistasia, bem como fenômenos associados (*e.g trade-offs*), não seriam diferentes daqueles processos encontrados em qualquer outro sistema complexo, nos quais múltiplas partes interagindo devem funcionar juntas para desempenhar as funções adequadamente (Mauro e Ghalambor, 2020). No caso dos organismos vivos, o resultado é sua extraordinária capacidade de se adaptar e evoluir (evolvabilidade).

Incorporar tais fenômenos em estudos evolutivos é difícil devido à complexidade das interações multiloci, resultando em desafios teóricos e estatísticos notáveis (Mackay 2014; Bank, 2022). Desse modo, buscamos, através de dois

trabalhos, responder como um gene (ou um grupo deles) e suas variantes se conectam com fenótipos adaptativos em humanos e outros primatas (“mapa genótipo-fenótipo adaptativo”).

Em um dos estudos que fazem parte de presente tese, e já publicado, utilizamos programa de análise conhecido, mas desenvolvemos uma abordagem estatística própria, para identificar a conexão (redes) de alelos de duas rotas genéticas potencialmente correlacionadas (coloração da pele e metabolismo da vitamina D), considerando populações nativas americanas que vivem em diferentes latitudes e altitudes e com diferentes hábitos de dieta (Missaggia et al., 2020).

O outro estudo envolve uma abordagem macroevolutiva, o qual foi desenvolvido utilizando técnicas de Inteligência Artificial, a partir de um método inovador denominado de *ProteinPhenotypeInsights* (ou *ProPhIn*), capaz de identificar os aminoácidos táxon-específicos em genes pleiotrópicos do sistema oxitocinérgico que potencialmente fariam parte do repertório genético epistático por trás de alguns fenótipos adaptativos, incluindo alguns relativamente raros em mamíferos, tais como monogamia social, intenso cuidado paterno (machos cuidadores) e parto gemelar (Fernandez-Duque et al., 2009; Lukas e Clutton-Brock, 2012; Stockley e Hobson, 2016).