

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE BIOCÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

Estudos Macro e Micro Evolutivos de Genes da Audição e Fala
em Mamíferos e Populações Humanas

Ágnis Iohana de Souza Grefenhagen

Porto Alegre

Março de 2023

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

INSTITUTO DE BIOCÊNCIAS

PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

Estudos Macro E Micro Evolutivos De Genes Da Audição E Fala Em Mamíferos E
Populações Humanas

Ágnis Iohana de Souza Grefenhagen

Dissertação submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS como requisito parcial para obtenção do título de mestre em Genética e Biologia Molecular.

Linha de Pesquisa: Genética de Populações, Evolução, Bioinformática.

Orientadora: Ursula da Silveira Matte

Co-orientadora: Maria Cátira Bortolini

Porto Alegre

Março de 2023

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

INSTITUTO DE BIOCÊNCIAS

PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

ESTUDOS MACRO E MICRO EVOLUTIVOS DE GENES DA AUDIÇÃO E FALA EM
MAMÍFEROS E POPULAÇÕES HUMANAS

Ágnis Iohana de Souza Grefenhagen

INSTITUIÇÕES E FONTES FINANCIADORAS

Este trabalho foi realizado no Núcleo de Bioinformática do Laboratório Células, Tecidos e Genes no Centro de Pesquisa Experimental, no Núcleo de Bioinformática do Hospital de Clínicas de Porto Alegre e no Laboratório de Evolução Humana Molecular do Departamento de Genética da UFRGS.

A fonte financiadora foi a Coordenação de Aperfeiçoamento de Nível Superior (CAPES).

Ao meu avô Henrique (*in memoriam*), que sempre acreditou no meu potencial e, mesmo sem saber ler e escrever, me ensinou muito sobre o poder da educação.

Obrigada vô.

AGRADECIMENTOS

Inicialmente, agradeço ao PPGBM e à CAPES pela oportunidade de realizar essa pesquisa, ao nosso secretário Elmo por sempre ser solícito e aos meus professores que me auxiliaram nessa caminhada.

Agradeço à USP por me aceitar como aluna especial em dois semestres. Agradeço à Coursera por me permitir realizar um curso sem custo algum e que foi primordial para as análises feitas e descritas nesta dissertação. Foram essas oportunidades que recebi ao longo do meu mestrado que me fizeram chegar até aqui com um empurrãozinho amigo.

Agradeço à minha orientadora Ursula Matte por todo companheirismo. Descobri contigo que uma orientação pode ser leve e engraçada. Agradeço por todas as risadas, por todo apoio e toda ajuda que tive. Levarei pra sempre comigo essa relação saudável de orientanda e orientadora. Obrigada, aliás, por me ajudar a manter meu trabalho e a minha pós-graduação ao mesmo tempo.

Agradeço à minha co-orientadora Maria Cátira Bortolini pela confiança e estímulo que recebi. Agradeço também por toda orientação excelente que tive nesse período. Será sempre uma pessoa que irei admirar pelo brilhantismo científico.

Agradeço a todos os meus colegas e amigos do Laboratório de Células, Tecidos e Genes e do Núcleo de Bioinformática do Hospital de Clínicas de Porto Alegre. Agradecimentos especiais à Cristal, Martiela, Pâmella, Laura, Luís, Paola, Ana, Marina e Lari pela amizade, pelos joguinhos e por toda a ajuda. Sem vocês, o mestrado não seria tão leve e suportável.

Agradeço ao Laboratório de Evolução Humana e Molecular, em especial à Luane Landau e à Bibiana Fam que auxiliaram esse trabalho e me ensinaram tudo o que sei. Obrigada por sempre serem solícitas.

Agradeço ao André Zani por toda ajuda nas horas em que precisei e por toda paciência para me ensinar evolução e genética de populações. Aqui vai meu muito obrigada por tudo mesmo.

Agradeço a todos os meus amigos que herdei da época da iniciação científica: Zi, Julinha, Bruno, Betinha, Lou, Piethro e Fernando. Obrigada por vibrarem comigo em todos os meus momentos importantes e por me apoiarem quando passei pelas fases mais difíceis.

Agradeço aos meus “cupinxas”, amigos que a graduação me deu. Obrigada também por vibrarem comigo por todas as minhas conquistas.

Agradeço a minha comadre e amiga Gracy e ao meu afilhado Kainã por todo carinho comigo. Um agradecimento especial a minha comadre por acreditar sempre em mim. Eu te admiro como mulher e como amiga.

Agradeço aos meus pais Jonatas e Teresinha e aos meus irmãos Ágatha, Ariadni, Rafael e Adllen por sempre acreditarem em mim. Obrigada por serem meu suporte e meu porto seguro. Agradeço também a minha avó Ambrosina, e minha avó do coração Iolanda que torceram e torcem pelo meu sucesso e felicidade. Eu amo vocês.

Agradeço as minhas cadelinhas Chica e Serena, que vieram no finalzinho do mestrado, mas foram super importantes pra minha saúde mental nesse período.

Agradeço ao meu amor Alexander por todo o companheirismo em todos esses anos. Tu me acompanhaste desde que entrei na graduação até chegar na pós-graduação e espero que esteja comigo em mais momentos importantes como este. Obrigada por ser meu maior fã e maior incentivador pra tudo na vida. Eu te amo imensamente. Obrigada por escolher formar uma mini familinha comigo.

Por fim, agradeço aos membros da banca que avaliaram este trabalho.

“Eis um problema! E cada sábio nele aplica
As suas lentes abismais.
Mas quem com isso ganha é o problema, que fica
Sempre com um X a mais...”
Mário Quintana

RESUMO

A comunicação pode ser definida como um sinal que fornece informação de um sinalizador para um receptor. Enquanto a ecolocalização é comum em cetáceos e morcegos, a comunicação acústica através da língua falada é uma característica de humanos. A genética da comunicação em mamíferos tem sido amplamente estudada e o uso de língua tonal e não tonal em humanos já foi correlacionada com a distribuição geográfica das populações. Portanto, o objetivo deste trabalho foi realizar estudos macroevolutivos de genes de audição em mamíferos com diferentes estratégias acústicas e estudos microevolutivos de genes relacionados à fala em populações humanas que falam línguas tonais e não-tonais. Para os estudos macroevolutivos, nós comparamos mamíferos com estratégias acústicas marinhas e terrestres (cetáceos e morcegos respectivamente), bem como roedores, artiodáctilos e primatas. As análises evolutivas foram realizadas nos pacotes PAML e MEME. A busca de frequências alélicas em humanos nos sítios sob seleção encontrados foram feitas através do dbSNP. Para as análises microevolutivas, usamos os dados disponíveis no 1000 Genomes e no *World Atlas of Language Structure*. Os genes candidatos para “fala” no *Human Phenotype Ontology* (HPO) foram usados para criar uma rede de genes no STRING e analisada no CYTOSCAPE. Duas populações, Vietnamitas e Cambojanos, foram analisadas através do *pairwise* Fst. A análise no PAML mostrou um gene, *DIAPH1*, com um X^2 estatisticamente significativo. Os resultados do *DIAPH1* demonstraram que não há posições comuns no MEME e no PAML. Esses sítios estão localizados em um domínio, FH1, que dimeriza o filamento de actina. Os sítios sob seleção encontrados no gene *DIAPH1*, detectados por cada um dos métodos, parecem ser táxon-específicos uma vez que *Homo sapiens* não apresentou polimorfismos importantes nessas posições, embora alelos raros são descritos em bancos de dados públicos. Em relação aos achados microevolutivos, nós encontramos duas variantes no gene *HSPG2*, na análise *pairwise* F_{ST} . Esse gene é relacionado ao aumento anormal do *pitch* (frequência) da voz. Outras investigações relacionadas a impactos funcionais desses achados precisam ser realizadas.

ABSTRACT

Communication can be defined as a signal that provides information from a signaler to a receiver. Although echolocation is common in cetaceans and bats, acoustic communication through spoken language is a feature of humans. The genetics of communication in mammals has been extensively studied and the use of tonal and non-tonal language in humans has already been correlated with the geographic distribution of populations. Therefore, the aim of this work was to carry out macroevolutionary studies of hearing genes in mammals with different acoustic strategies and microevolutionary studies of genes related to speech in human populations that speak tonal and non-tonal languages. For macroevolutionary studies, we compare mammals with marine and terrestrial acoustic strategies (cetaceans and bats, respectively), as well as rodents, artiodactyls and primates. The evolutionary analysis was performed in the PAML and MEME packages. The search for allele frequencies in humans at the selection sites found was performed using dbSNP. For microevolutionary analyses, we used available data from 1000 Genomes and the World Atlas of Language Structure. Speech candidate genes in the Human Phenotype Ontology (HPO) were used to create a gene network in STRING and analyzed in CYTOSCAPE. Two populations, Vietnamese and Cambodian, were analyzed using the pairwise F_{ST} . PAML analysis showed one gene, DIAPH1, with a statistically significant X^2 . DIAPH1 results demonstrated that there are no common positions in MEME and PAML. These sites are located in a domain, FH1, that dimerizes the actin filament. The sites under selection found in the *DIAPH1* gene, detected by each of the methods, seem to be taxon-specific, since *Homo sapiens* did not show significant polymorphisms in these positions, although rare alleles are described in public databases. Regarding the microevolutionary findings, we found two variants in the F_{ST} pairwise analysis, located in the *HSPG2* gene. This gene is related to the abnormal increase in the pitch (frequency) of the voice. Further investigations related to the functional impacts of these findings need to be carried out.

LISTA DE ABREVIATURAS E SIGLAS

HPO - Human Phenotype Ontology

MHC - Complexo de Histocompatibilidade

EHH - Homozigosidade Haplotípica Estendida

GARD - Genetic Algorithm for Recombination Detection

MEGA - Molecular Evolutionary Genetics Analysis

PAML - Phylogenetic Analysis by Maximum Likelihood

BEB - Bayes Empirical Bayes

MEME - Mixed Effect Model of Evolution

GDV -Genome Data Viewer

ALFA - Allele Frequency Aggregator

PolyPhen - Polymorphism Phenotyping

HGDP - Human Genome Diversity Project

WALS - World Atlas of Language Structures

CRAM - Compressed Reference-oriented Alignment Map

VEP - Variant Effect Predictor

VCF - Variant Call Format

GWAS -Genome Wide Association Study

FDR - False Discovery Rate

MCC - Maximal Clique Centrality

PC2 - Componente Principal 2

GO - Gene Ontology

CdLS - Síndrome de Cornelia de Lange

CRISPR - Conjunto de Repetições Palindrômicas Regularmente Espaçadas

LISTA DE FIGURAS

Figura 1 – Elementos envolvidos na comunicação.

Figura 2 - Tipos de comunicação encontrados na natureza.

Figura 3 - Distribuição de populações falantes de línguas tonais e não-tonais.

Figura 4 - Diferenças entre convergência e paralelismo na história evolutiva de uma proteína hipotética.

Figura 5 - Filogenia das espécies analisadas e sítios do DIAPH1 sob seleção segundo análise no PAML.

Figura 6 - Mapa da intersecção entre os dados do 1000 Genomas e do WALS.

Figura 7- Manhattan plot com os dados de pairwise F_{ST} entre as populações de Cambojanos e Vietnamitas.

Figura 8 - Ontologia dos genes presentes entre os 0,01% selecionados no F_{ST} .

Figura 9 - Resultado do “GWAS Catalog” para os genes presentes entre os 0,01% selecionados no F_{ST} .

Figura 10 - Rede topológica criada pelo CytoHubba.

Figura 11 - Testes funcionais para estudos de variantes.

LISTA DE TABELAS

Tabela 1 - Amplitude de frequência audível em diferentes grupos de animais.

Tabela 2 - Genes relacionados aos termos “audição” e “frequência”.

Tabela 3 - Sítios sob seleção em genes relacionados aos termos “audição” e “frequência”

Tabela 4. Análise de cada grupo taxonômico que apresenta ausência ou presença de ecolocalização e do resíduo mais comum ao longo da filogenia.

Tabela 5 - Impacto das variantes preditas pelo PolyPhen.

Tabela 6 - Frequência dos alelos nas posições com indícios de seleção positiva no DIAPH1.

Tabela 7 - Valores de frequência e de F_{ST} para os SNPs do gene *HSPG2*.

Tabela 8 - *Rank* dos 20 *top* genes da rede em ordem do maior para o menor escore.

SUMÁRIO

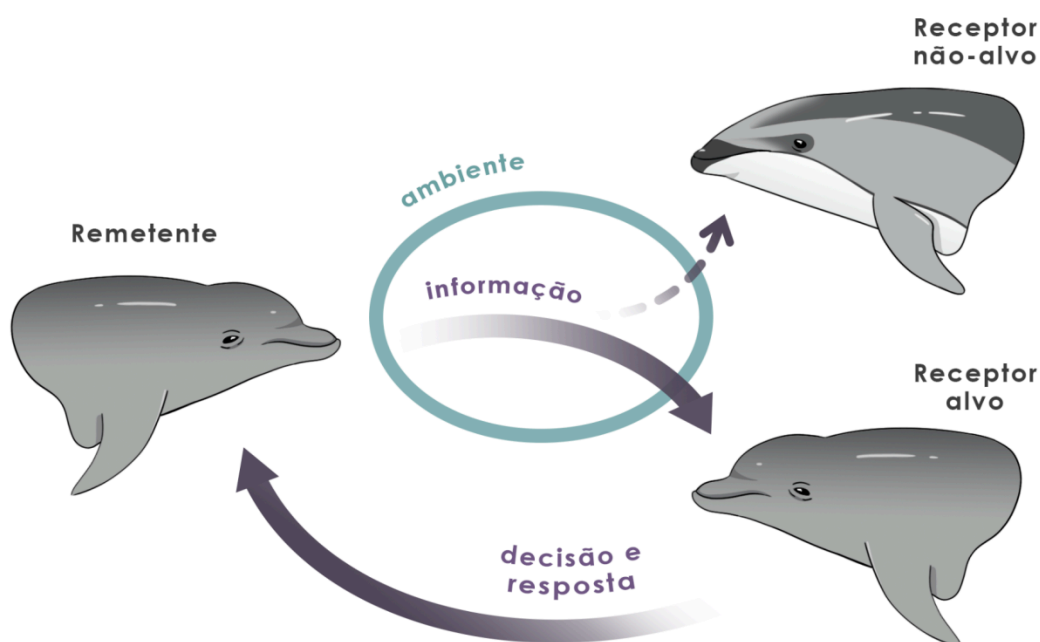
Resumo	8
Abstract	9
Lista de abreviaturas e siglas	10
Lista de Figuras	11
Lista de tabelas	12
1 Introdução	14
1.1 Comunicação no reino animal	14
1.2 Audição	18
1.3 Fala	22
1.4 Variabilidade dos genes ASPM e MCPH1 em populações humanas e seu papel nas linguagens tonais e não-tonais	25
1.5 Papel do ambiente	27
1.6 Estudos macroevolutivos	28
1.6.1 Estudo filogenético das espécies e das moléculas	28
1.6.2 Paralelismo e co-evolução	29
1.7. Estudos microevolutivos	30
1.7.1. Seleção e deriva	30
2. Objetivos	34
3 Métodos	35
3.1 Métodos do estudo macroevolutivo	35
3.2 Métodos do estudo microevolutivo	37
4. RESULTADOS	40
4.1. Resultados dos estudos macroevolutivos	40
4.2. Resultados dos estudos microevolutivos	50
5. DISCUSSÃO	58
6. REFERÊNCIAS BIBLIOGRÁFICAS	66

1. INTRODUÇÃO

1.1. Comunicação no reino animal

A comunicação no reino animal pode ser definida como qualquer compartilhamento de informações de qualquer fonte (Smith, 1977). Wilson, por sua vez, define tal tipo de comunicação como “a ação ou sugestão dada por um organismo [o remetente] que é percebida por ele e altera o padrão de probabilidade de comportamento em outro organismo [o receptor] de uma forma adaptável a qualquer um dos dois participantes" (Wilson, 1975). Nesse sentido, há um sinal que fornece informações vindas de um sinalizador para um receptor (Figura 1) e que ajuda os animais fazerem suas escolhas. Segundo Smith, quando um indivíduo adquire informações, torna-se mais capaz de antecipar e responder adequadamente aos eventos e usar o *feedback* para selecionar ações e promover mudanças em seu estado (Smith, 1977).

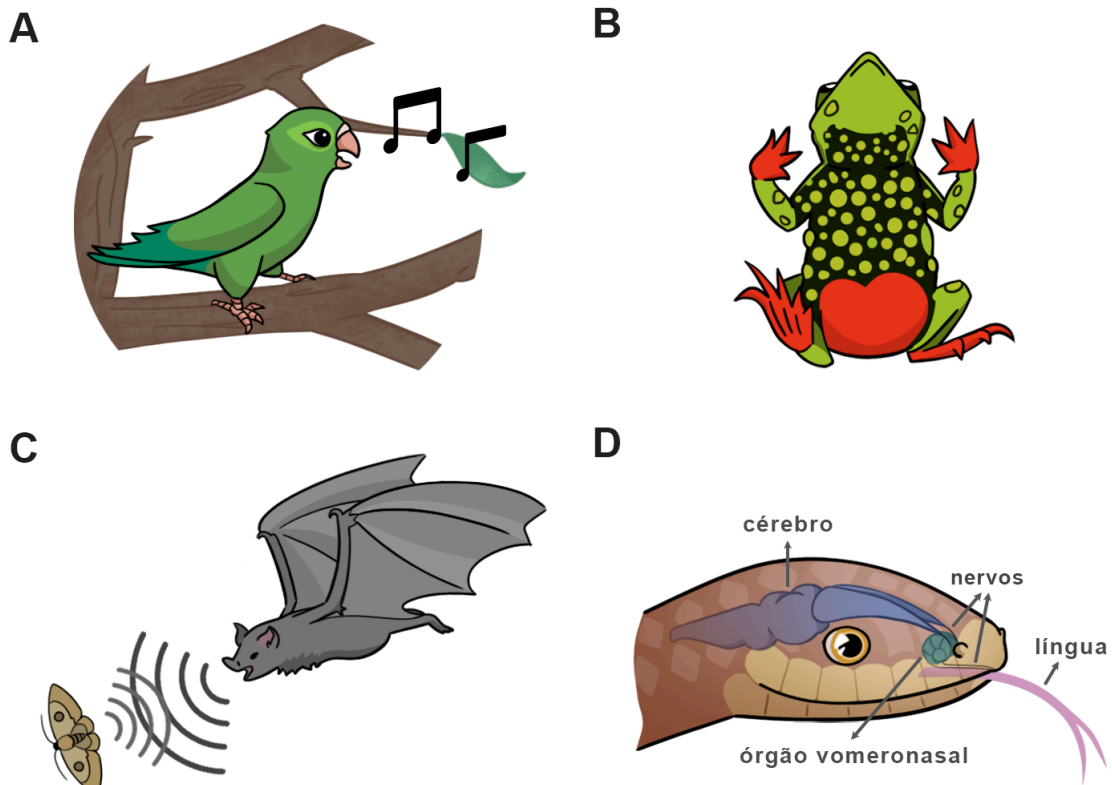
Figura 1. Elementos envolvidos na comunicação.



O remetente envia uma informação que é recebida pelos demais indivíduos presentes no meio, mas faz sentido apenas para o receptor, que então pode tomar uma decisão ou enviar uma resposta. Adaptado de Erin Gillam, 2011. Autor: Laura Haleva. Imagem adaptada no Adobe Photoshop.

A comunicação e a linguagem não são exclusivas dos seres humanos, ainda que a associação da linguagem à imaginação, segundo alguns autores, pareça ser única da nossa espécie (Harari, 2012). A comunicação nos animais pode ocorrer por diversos motivos: encontrar e capturar presas, evitar predadores, orientação no espaço geográfico, encontrar e selecionar um parceiro e cuidar da prole (Hoelzel, 2002). Os animais usam uma grande quantidade de canais sensoriais ou modalidades de sinais para se comunicar. Podemos dividir o tipo de comunicação em pelo menos 3 formas: a comunicação visual, acústica e química (Figura 2). A comunicação visual é extremamente efetiva para animais que são ativos durante o dia. A comunicação química desempenha um papel importante na vida de alguns animais e depende da detecção de sinais químicos por parte do receptor. Alguns animais possuem órgãos especializados para isso, como o órgão vomeronasal. Por exemplo, elefantes asiáticos machos (*Elephas maximus*) usam esse órgão para processar sinais químicos na urina da fêmea de modo a detectar se ela está sexualmente receptiva (Rasmussen *et al*, 1982). Comparada à comunicação visual e acústica, os sinais químicos viajam mais lentamente pelo espaço, podendo ser transmitidos por longas distâncias e demoram a desaparecer depois de sua produção.

Figura 2. Tipos de comunicação encontrados na natureza, considerando o reino animal.



A) Comunicação acústica. B) Comunicação visual. C) Comunicação por ecolocalização. D) Comunicação química. Autor: Laura Haleva. Imagem criada no Adobe Photoshop.

Já a comunicação acústica é muito abundante na natureza, provavelmente porque os sons podem se adaptar a uma ampla variedade de condições ambientais e situações comportamentais (Gillam, 2011). Em aves canoras, a siringe é o órgão responsável pela produção de som e os sinais vocais são divididos em chamados (“calls”) e em canto (“song”). O canto é utilizado para controle territorial e para encontrar um parceiro, enquanto o chamado é utilizado para contato, acasalamento, ameaça, “implorar” (usado por filhotes), vôo e alarme (Kumar, 2003).

Mamíferos, anfíbios e répteis também se utilizam da comunicação acústica para transmitir informações. As vocalizações e comunicação acústica é mais bem conhecida em anuros, nos quais tem papel importante na reprodução (Ovaska & Rand, 2001; Ryan, 2001; Gerhardt & Huber, 2002; Duellman & Trueb, 1986; Feder & Burggren, 1992; Walkowiak,

2006; Yang & Kelley, 2008). Os machos de muitas espécies vocalizam e competem uns com os outros de modo a ganhar a atenção da fêmea (Gerhardt & Huber, 2002; Wells 2010; Bee, 2015; Clulow et al., 2017). Entre os répteis, os chamados são usados para interações entre a mãe e a prole, acasalamento e interações territoriais (Vergne *et al.*, 2009; Frankenberg, 1982; Galeotti *et al.*, 2005).

A comunicação acústica também é observada nos primatas não-humanos, entre outros mamíferos. Esses animais utilizam um tipo de vocalização que é denominada “calls” ou “chamados”. Em um estudo com macacos-de-nariz-branco (*Cercopithecus nictitans*), os pesquisadores perceberam que os machos combinavam dois chamados básicos para sinal de perigo em uma sequência única. Os membros do bando que escutam esse chamado combinado tendiam a percorrer uma distância maior do que quando o chamado básico era utilizado de maneira individual, logo indicando que o chamado combinado teria um significado próprio denominado “afastar” (Arnold & Zuberbühler, 2006a; Arnold & Zuberbühler, 2006b). Evidências similares também foram encontradas em macacos-de-Campbell (*Cercopithecus campbelli*), gibões (família Hylobatidae) e bonobos (*Pan paniscus*) (Ouattara et al., 2009a; Ouattara *et al.*, 2009b; Clay & Zuberbühler, 2011), sugerindo que as sequências organizadas sintaticamente (chamados em sequência) são comuns na comunicação de primatas. Além disso, chimpanzés (*Pan troglodytes*) parecem entender o significado dos chamados. Fêmeas de chimpanzés, por exemplo, tendem a evitar os chamados de copulação dependendo do número de indivíduos em sua volta, isto é, as fêmeas tendem a não utilizar o chamado de copulação caso haja um número maior de indivíduos ao redor como forma de prevenir competição social (Townsend *et al.*, 2008).

Um tipo particularmente especializado de comunicação encontrada em morcegos microquirópteros e cetáceos é a ecolocalização. Resumidamente, é um processo fisiológico em que sons de alta frequência são emitidos para detectar e localizar a presa ou obstáculos. Depois da emissão do som, o indivíduo detecta e processa o eco gerado. Isso permite que o animal faça avaliações muito precisas da localização do objeto ou presa. Além de ser usado para forrageamento e predação, também é utilizado para navegação e proteção (Whitlow 2009). Entretanto, há evidências que demonstram que a ecolocalização também pode facilitar a comunicação social. Em um artigo de 2012, Knörnschild e colegas relatam que morcegos machos da espécie *Saccopteryx bilineata* discriminam a aproximação de machos e fêmeas da mesma espécie apenas com base em seus chamados de ecolocalização. Ainda, os machos produzem vocalizações de cunho agressivo quando escutam os chamados de ecolocalização

de outros machos e, quando escutam chamados de ecolocalização produzidos por fêmeas, produzem chamados de copulação (Knörnschild et al., 2012)

As descobertas da ecolocalização, ou de suas evidências, em golfinhos e morcegos remontam ao século passado. Em 1938, G. Pierce e D. Griffin forneceram a evidência de que morcegos emitem pulsos ultrassônicos, apesar de suspeitarem disso desde os anos 1910. Em relação aos golfinhos, só em 1960 se demonstrou esse tipo de comunicação nesses animais, a partir dos experimentos de Norris e colegas (1960). A ecolocalização também pode ser denominada como autocomunicação, uma vez que o animal que ecolocaliza é o sinalizador e receptor (Bradbury & Vehrencamp, 1998).

1.2. Audição

Para que haja a comunicação entre dois indivíduos, é preciso que uma série de órgãos trabalhem em conjunto. O ar que atravessa a laringe é transformado em som pela vibração das cordas vocais e transforma-se em um som (Lieberman *et al.* 1992). Agora, o ouvinte precisa escutar e processar o som, as palavras pronunciadas, utilizando seu aparelho auditivo e regiões do cérebro como o córtex auditivo e o córtex motor (Friederici, 2012). Sendo assim, essa habilidade tão complexa necessita de ao menos 3 regiões: o aparelho fonador, o aparelho auditivo e as regiões cerebrais associadas.

Especificamente em relação à audição, ela ocorre através da recepção de ondas sonoras que chegam ao pavilhão auricular e fazem a membrana timpânica vibrar. O tímpano, por sua vez, transfere essa energia mecânica para os ossículos do ouvido até chegar no fluido presente na cóclea. Tudo isso então é detectado pelas células ciliadas que são estimuladas e enviam esse estímulo para o nervo auditivo, que leva essa informação para o cérebro (Seikel, Konstantopoulos & Drumright, 2020).

Os sons chegam ao ouvido em diferentes frequências e a capacidade de detectar os sons varia amplamente entre as espécies (Tabela 1). Por exemplo, a maioria dos morcegos emitem chamados ultrassônicos (> 20 kHz), os quais são de frequência alta e inaudíveis para os seres humanos (20Hz - 20kHz; Jones, 2005; Rosen, 2011), mas audíveis para os outros morcegos. Golfinhos também podem ouvir em frequências muito altas, acima de 100 kHz, enquanto baleias, que fazem parte do mesmo grupo taxonômico dos golfinhos, escutam em frequências um pouco menores. Há outros animais que escutem em frequências muito baixas, inaudíveis para humanos. Os elefantes, por exemplo, escutam em frequências que variam de

15 Hz a 20 Hz (Garstang, 2004). Os sons audíveis em diferentes frequências possuem papéis muito importantes para a sobrevivência e o comportamento, e são essenciais para a comunicação entre os indivíduos de uma mesma espécie.

Tabela 1. Amplitude de frequência audível em diferentes grupos de animais.

Grupo	Amplitude de Frequência	Referência
Rã e sapo	50Hz - 4kHz	Fay, R.R. (1988). Hearing in Vertebrates: A Psychophysics Databook. Winnetka, IL: Hill-Fay Associates.
Crocodilo	300Hz - 8KHz	Wever EG. (1971). Hearing in the Crocodilia. Proc. Natl Acad. Sci. USA 68, 1498-1500.
Ave	1kHz - 4kHz	Beason, R.C. (2004). What can birds hear? Proceedings of the Vertebrate Pest Conference, 21
Baleia Azul	14kHz - 36kHz	Cummings, W.C., Thompson, P.O. (1971). Underwater sounds from the blue whale, <i>Balaenoptera musculus</i> . J Acoust Soc Am. 50:1193-1198.
Golfinho	150Hz - 150kHz	Fay, R.R. (1988). Hearing in Vertebrates: A Psychophysics Databook. Winnetka, IL: Hill-Fay Associates.
Morcego	15kHz - 90kHz	Adams, R.A, Pedersen, S.C. (2000). Ontogeny, Functional Ecology, and Evolution of bats. Cambridge University Press. pp. 139-140.

Rato	250Hz - 80kHz	Escabi, C.D., Frye, M.D., Trevino, M, Lobarinas, E. (2019). The rat animal model for noise-induced hearing loss. The Journal of the Acoustical Society of America, 146(5), 3692-3709.
Cachorro	67Hz - 45kHz	Condon, Timothy (2003). Elert, Glenn (ed.). Frequency Range of Dog Hearing. The Physics Factbook.
Humano	20Hz - 20kHz	Rosen, S. (2011). Signals and Systems for Speech and Hearing (2nd ed.). BRILL. p. 163.

Uma série de genes controlam o desenvolvimento de órgãos, células e regiões cerebrais associadas à audição. É no ectoderma que o placode óptico surge, ventralmente no rombencéfalo, e se invagina para dar origem ao otocisto, que mais tarde resultará no ouvido. O placode óptico é induzido pela sinalização do fator de crescimento de fibroblasto (FGF) e a expressão de *SOX2* marca o domínio prossensorial da cóclea. A diferenciação de células neurais da cóclea ocorre com a expressão de vários fatores de transcrição. Muitos desses fatores de transcrição surgiram há muito tempo na evolução, presentes no último ancestral comum de eucariotos (de Mendoza et al. 2013). Assim, esses genes são pleiotrópicos, e alterações nas suas sequências impactam não somente regiões ligadas à audição. Variantes nesses genes podem produzir *trade-offs*, com uma desvantagem associada e levando a restrições evolutivas (Hughes & Leips, 2017).

Para avaliar genes que podem estar associados, de uma forma mais específica, a um fenótipo, é possível analisar aqueles cujas mutações causam alguma anomalia relativa ao fenótipo de interesse. Por exemplo, no caso da audição, podemos analisar genes que apresentem mutações que resultam em um dano direto na recepção e processamento do som. Danos auditivos podem ter caráter síndrômico ou não síndrômico (Gettelfinger & Dahl, 2018; Venkatesh, Moorchung & Puri, 2015). A perda auditiva genética não síndrômica corresponde a cerca de 70% dos casos e pode ser herdada de forma autossômica recessiva ou dominante

ou por herança ligada ao X (Petit, Levilliers & Hardelin, 2001). Casos mais raros envolvem até mesmo a mitocôndria (Korver et al., 2017).

Além disso, o modo como a doença se expressa varia bastante, ocorrendo de forma mais atenuada até mais severa. Ainda, podemos ter a perda auditiva condutiva, que se refere a defeitos do ouvido médio/externo, ou a perda auditiva sensorineural, que envolve defeitos no ouvido interno ou até mesmo nas regiões do cérebro associados à audição. A perda auditiva não sindrômica é quase que exclusivamente sensorineural. Além do grau de severidade da doença, também há formas relacionadas à perda da audição para determinadas frequências sonoras. Há casos de indivíduos que não escutam faixas de frequência altas (entre 3000 e 6000 Hz), faixas médias (entre 1000 Hz e 3000 Hz) e frequências baixas (125 Hz a 1000 Hz) (Human Phenotype Ontology).

Ainda que a surdez hereditária seja uma condição com alto grau de heterogeneidade genética, alguns genes se destacam como os principais responsáveis pela surdez não sindrômica. Entre eles, o *GJB2*, que codifica a conexina 26. A mutação por mudança de matriz de leitura (c.35delG) nesse gene é uma das causas mais comuns de perda auditiva em humanos. A conexina 26 é expressa na cóclea e na epiderme e forma canais entre as células os quais, quando abertos, permitem a difusão de pequenas moléculas entre as células. Problemas na proteína conexina 26 levam a falha da difusão de potássio intracelular e na transmissão dos sinais auditivos até as regiões que processam o som (Chen et al. 2022). Além dessa mutação, uma série de outras mutações foram descritas no *GJB2* como causadoras de perda auditiva, desde mutações do tipo deleção, inserção, mutações de sentido trocado e mutações sem sentido (Kenneson et al., 2002; Shen et al., 2019).

Além do *GJB2*, também no gene *GJB6* que codifica a conexina 30 já foram descritas mutações associadas à perda auditiva (Batissoco et al. 2009; Wang et al. 2011; del Castillo et al. 2003). Por exemplo, em 2002, Pallares-Ruiz e colaboradores analisaram um caso de surdez congênita sem ligação com o gene *GJB2* e descobriram uma deleção em homozigose no gene *GJB6* em torno 150kb de extensão, incluindo a terminação 5'. Esse achado demonstra que a ausência dessa região do gene em homozigose causa danos severos à audição. Em outro estudo, uma deleção que compreende 309kb e uma segunda deleção na região 5' não traduzida, identificada na população da Espanha e Reino Unido foram descritas (el Castillo et al., 2005).

Além dos genes *GJB2* e *GJB6*, mutações no gene *SLC26A4* são a segunda causa mais proeminente de perda auditiva (Albert et al., 2006). Um exemplo é a síndrome de Pendred, a forma mais comum de perda auditiva sindrômica associada a mutações nesse gene (Pera et al., 2008). O *SLC26A4* codifica uma proteína (pendrina) que realiza o transporte transmembrana de ânions na cóclea (Everett et al., 2006). A mutação mais comum nesse gene é uma mutação que afeta o *splicing*, ou seja, o processamento do RNA mensageiro (c.919-2A>G). Pacientes afetados por esse tipo de mutação apresentam variados graus de severidade da doença. Por exemplo, existem mutações na parte C-terminal do gene que podem levar à geração de um sítio de terminação precoce, sintetizando uma proteína truncada. Apesar disso, há manutenção da função residual da pendrina, o que pode explicar a presença de um fenótipo menos severo da doença em alguns pacientes (Van Hauwe et al., 1998). Outras mutações patogênicas foram descritas e promovem o surgimento de novos sítios de *splicing* (López-Bigas et al., 1999; Park et al., 2003). Porém, ainda mantém resíduos importantes da proteína, levando a fenótipos menos severos (Lopez-Bigas et al., 1999). Algumas dessas mutações, como p.Thr416Pro, p.Leu236Pro e p.Glu384Gly, foram associadas à retenção da pendrina no retículo endoplasmático (Rotman-Pikielny et al., 2002).

Além de mutações em genes associados à perda auditiva sindrômica, mutações em outros genes são responsáveis por perda auditiva não sindrômica. Um desses genes é o *MYO15A*, que codifica a miosina XVa. A maioria das mutações nesse gene são associadas à perda auditiva congênita de grau severo a profundo. Pacientes com este tipo de fenótipo não ouvem nenhum som de fala e poucos sons do ambiente. Estudos com roedores demonstraram que a miosina XVa é necessária para a organização da actina nas células auditivas da cóclea (Belyantseva, Boger & Friedman, 2003).

1.3. Fala

Segundo Chomsky (2017), a linguagem complexa é uma habilidade de comunicação encontrada exclusivamente em humanos. E não somente ela é complexa como também diversa. Em levantamento atual do “Ethnologue” (Eberhard et al., 2020), um banco de dados especializado em linguagens, são contabilizados em torno de 7117 idiomas e dialetos diferentes no mundo. Cerca de 40% deles estão sob risco de desaparecer e apenas 23 línguas são faladas por mais da metade da população mundial (Eberhard et al., 2020).

Os idiomas podem ser classificados em famílias conforme sua origem comum e suas características como vocabulário básico, morfologia e fonologia (Campbell & Poser, 2008).

Como exemplos de famílias linguísticas temos a indo-europeia, com aproximadamente 850 idiomas, e a mais ampla de todas, que é a atlântico-congolesa com pelo menos 1430 línguas, abrangendo quase todo o continente Africano. Já a família com maior distribuição espacial é a austronésia, presente desde o Hawaii até Madagascar, sendo composta por cerca de 1274 idiomas (Hammarström, 2016).

As línguas também podem ser divididas em tonais e não-tonais de acordo com o uso de tons lexicais em consoantes e vogais, independente da família a que pertencem. Línguas tonais utilizam o “tonema”, que é uma terceira classe fonética, na qual o tom é o elemento diferenciador de palavras. Ou seja, o tom é alterado para dar significado diferente à palavra (Jones, 1944). Alguns idiomas desse tipo utilizam tons em todas as sílabas, enquanto outros restringem seu uso a sílabas acentuadas de palavras específicas (Best, 2019). Alguns exemplos de linguagens tonais são mandarim, cantonês, vietnamita, coreano, suruí (utilizada pelo povo indígena amazônico identificado com o mesmo nome) e alguns dialetos africanos (Dryer & Haspelmath, 2013; Moore, 2005). As linguagens não-tonais são aquelas muito presentes no continente europeu como o inglês, alemão, italiano e em algumas nas regiões da Oceania, África e Ásia (Dryer & Haspelmath, 2013). Neste grupo não há distinção tonal entre as palavras, ou seja, entonações de sílabas ou vogais não refletem um significado diferente, mas uma carga emocional atribuída às palavras (Zhang, 2018).

Embora não haja uma resposta concreta para a existência de tantas línguas diferentes, diversos fatores parecem influenciar na estrutura e evolução da linguagem. O tamanho populacional, por exemplo, influencia no surgimento ou perda de palavras do vocabulário (Bromham *et al.*, 2015; Greenhill *et al.*, 2018). Além disso, utilizando dados genéticos (*loci* de microssatélites), Creanza e colaboradores (2015) observaram que a migração moldou a evolução dos fonemas. Linguagens que são relativamente isoladas têm mais variância no número de fonemas do que linguagens com muitos “vizinhos”, ou seja, o isolamento populacional contribuiria para o nível de mudanças fonéticas observadas.

A anatomia do trato vocal também é outro fator importante. A forma da crista alveolar do palato tem um papel relevante em algumas populações sul e leste-africanas que utilizam *clicks* (produzido pelo estalo da língua e que funciona como uma consoante). A forma do palato duro reflete em estratégias articulatórias e acústicas na produção de vogais (Moisik & Dediu, 2017; Dediu, Janssen & Moisik, 2019). Além disso, de forma mais ampla, a alteração na configuração da mordida no período pós-neolítico possibilitou o surgimento de consoantes

como o “f” e o “v” (Blasi *et al.*, 2019). Portanto, estes achados demonstram a influência da anatomia do trato vocal na linguagem, e por consequência do repertório genético por trás desses traços, mas que ainda não foram identificados.

Por outro lado, um dos genes mais bem conhecidos com importância na linguagem humana é o *Forkhead Box Protein 2 (FOXP2)* que codifica um fator de transcrição expresso no cérebro em altos níveis durante o desenvolvimento fetal (Lai *et al.*, 2001). O papel deste gene foi descoberto através do caso da família KE, que apresenta um distúrbio de linguagem e fala com padrão de herdabilidade autossômica dominante (Hurst *et al.*, 1990). Um pouco mais de uma década depois, Lai e colaboradores (2001) demonstraram que havia uma mutação de ponto no exon 14 deste gene nos indivíduos acometidos pelo distúrbio. No mesmo estudo, um indivíduo não aparentado apresentava um distúrbio semelhante, sendo identificado que neste caso o gene *FOXP2* havia sido interrompido por uma translocação (Lai *et al.*, 2001). Logo depois, Enard *et al.* (2002), utilizando uma abordagem evolutiva, detectaram que duas substituições não sinônimas no exon 7 (p.Asn325Ser e p.Thr303Asn) do *FOXP2* de *Homo sapiens* diferiam do ortólogo de chimpanzé. Posteriormente, se viu que os alelos derivados 325Ser e 303Asn estavam também presentes no *Homo neanderthalensis* (Krause *et al.*, 2007). Mais recentemente, o grupo do Laboratório de Evolução Humana e Molecular do Departamento de Genética da UFRGS, mostrou que a modificação p.Asn325Ser promove o ganho de dois motivos lineares curtos de ligação, denominados MOD_CK1_1 e MOD_GSK3_1, devido a presença da serina nesta posição. Ambos os motivos são promotores da fosforilação por quinases (Viscardi *et al.*, 2017). Cooper (2006), por sua vez, sugeriu que a fosforilação pela quinase C na proteína FOXP2 pode estar relacionada a características comportamentais humanas, como a linguagem. Interessantemente, Viscardi *et al.* (2017) mostraram que a outra substituição específica do gênero *Homo*, a presença de uma asparagina na posição 303 da proteína FOXP2 levou à perda de um sítio de fosforilação. Mudanças nos padrões de fosforilação podem modificar a forma como um fator de transcrição, caso de FOXP2, atua, de modo a alterar sua afinidade de ligação com outros elementos genéticos, sejam proteínas ou o próprio DNA alvo. Sendo assim, essas mudanças podem alterar a expressão gênica, o crescimento e a diferenciação celular (Iakoucheva *et al.*, 2004). Diante destes achados, o *FOXP2* foi o primeiro gene a ser fortemente associado à língua falada. Além deste gene, o *CNTNAP2* e o *SRPX2* (um dos alvos de FOXP2) têm sido estudados devido a sua associação com distúrbios da fala (Roll *et al.*, 2006; Vernes *et al.*, 2008; Roll *et al.*, 2010).

Quando se trata da existência de idiomas tonais e não-tonais, dados apontam para a importância de mecanismos neurais para o processamento linguístico (Chien *et al.*, 2020; Berthelsen *et al.*, 2020). Regiões cerebrais como áreas fronto-parietais esquerdas, conhecidas pela função de processamento de tom, são evocadas em falantes desses dois tipos de linguagem. Por outro lado, algumas regiões adicionais, como áreas subcorticais, são evocadas apenas naqueles que utilizam linguagens tonais, sugerindo que há um processamento semântico diferente do tom linguístico (Chien *et al.*, 2020). O componente biológico neste contexto faz emergir a ideia de que uma população que é geneticamente predisposta a perceber uma particular característica linguística tenderia a adotar essa característica em sua língua. Essa hipótese vem sendo sustentada por vários estudos genéticos nos últimos 20 anos, dentro os quais aqueles que investigaram os genes *ASPM* e *MCPHI* como veremos abaixo.

1.4. Variabilidade dos genes *ASPM* e *MCPHI* em populações humanas e seu papel nas linguagens tonais e não-tonais

Os genes *MCPHI* e *ASPM* codificam as proteínas “microcefalina” (MCPH1) e “proteína associada à microcefalia anormal do fuso” (ASPM), ambas relacionadas com o desenvolvimento do cérebro e do córtex cerebral, indicando seu papel crucial na neurogênese. Alelos desses genes/proteínas apresentam marcantes diferenças em suas distribuições ao longo das populações humanas (Dediu & Ladd, 2007). Dediu e Ladd (2007) observaram que o alelo derivado C de um polimorfismo presente no exon 18 de *ASPM* (rs41310927) apresentava altas frequências na Ásia Central e no oeste asiático, na Europa Ocidental, norte da África e Papua Nova Guiné, enquanto apresentava frequência menor no leste asiático, África Subsaariana e nas Américas. Já o alelo derivado C de um polimorfismo no exon 8 de *MCPHI* (rs930557) se mostrou frequente na Ásia, Europa e nas Américas, enquanto sua frequência foi menor no sudeste asiático e muito rara na África Subsaariana. Vale destacar que ambos os alelos derivados resultam em mudanças de aminoácidos nas proteínas, *ASPM* (p.Ser2562Gly) e *MCPHI* (p.Asp314His). Interessantemente, Dediu e Ladd (2007) encontraram uma correlação negativa e significativa entre esses alelos derivados e uma tendência de uso de tons lexicais em uma língua, mesmo depois que fatores geográficos e históricos foram controlados. Em outras palavras, em populações falantes de línguas tonais o alelo ancestral T/Ser seria o mais frequente. Uma explicação para esses achados seria que *ASPM* e *MCPHI* estão associados ao processamento de sinais acústicos simultâneos dentro

de uma janela temporal curta, visto que falantes de tom lexical teriam que ter habilidades para processar a sílaba e o tom como uma unidade integral (Dediu e Ladd, 2007).

Posteriormente, Wong *et al.* (2012) também investigaram a temática da co-evolução de genes e certas características das linguagens humanas, introduzindo em seu estudo exames de neuroimagem. Os autores encontraram um resultado diverso daquele de Dediu e Ladd (2007) pois sugeriram que o alelo derivado de *ASPM/ASPM* (C/Gly) apresentava correlação positiva com a percepção do tom. Porém, os autores também encontraram correlação entre o alelo C/Gly com percepção tonal numa amostra de norte-americanos que não falavam uma língua tonal (Wong *et al.*, 2012). Os exames de neuroimagem, por sua vez, corroboram que é no córtex temporal direito que o processamento dos tons ocorre preferencialmente, e que havia uma correlação positiva significativa entre a percepção do tom e a presença do alelo derivado Gly de *ASPM*. Por outro lado, nenhuma correlação significativa foi encontrada entre o alelo derivado de *MCPHI/MCPH1* (C/His) e a percepção do tom. Wong *et al.* (2012) atribuem as diferenças com relação aos achados de Dediu e Ladd (2007) ao fato bem conhecido de que o mesmo alelo pode resultar em diferenças fenotípicas devido aos efeitos da complexa rede de interações gene-gene e/ou gene-cultura encontrados nas diferentes populações humanas.

Ciente das críticas quanto às conexões sugeridas entre alelos específicos de determinados genes e características de linguagem, bem como aos achados contraditórios dos diferentes estudos, Wong *et al.* (2020) investigaram geneticamente 400 indivíduos da etnia Han, falantes de cantonês, uma linguagem tonal que possui um dos maiores inventários tonais de todas as línguas tonais conhecidas, com seis tons lexicais. Os voluntários foram convidados a proferir tons lexicais e musicais, além de desempenharem algumas tarefas cognitivas pré determinadas. Informações sobre formação musical também foram obtidas dos participantes. A abordagem genética envolveu um conjunto de genes conhecidos por modular habilidades cognitivas e processamento da linguagem (incluindo *ASPM*). Os autores encontraram que o alelo ancestral (T/Ser) da variante rs41310927 de *ASPM/ASPM* possuía uma frequência de 84% na amostra e estava associado significativamente com uma maior habilidade de percepção de tons lexicais dos Han, sugerindo uma evidência mais direta da associação deste alelo com falantes de línguas tonais (Wong et al. 2020). Esse achado vem ao encontro daquele de Dediu e Ladd (2007). Vale destacar que os dados da fase 3 do Projeto 1000 Genomas indicam que a frequência do alelo T/Ser (rs41310927) nos Han é de ~80%,

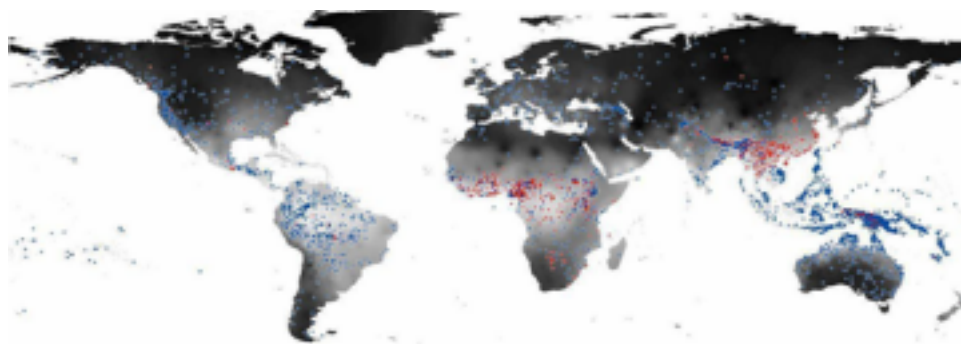
um valor intermediário ao encontrado para os britânicos (T/Ser = 52%) e os Yorubá da Nigéria (T/Ser = 97%).

Além disso, vem sendo demonstrado que o sinal da ação da seleção positiva poderia explicar pelo menos parte das distribuições alélicas dos alelos derivados de *ASPM* e *MCPHI* (Evans *et al.*, 2005; Mekel-Bobrov *et al.*, 2005; Dediu & Ladd, 2007). Mais recentemente, Haque *et al.* (2016) analisou dados do Projeto 1000 genomas e também encontrou sinal de pressão seletiva no gene *ASPM*. Eles destacaram ainda o compartilhamento de outros 11 SNPs entre todas as populações do estudo (total de 16 populações das regiões africana, europeia, sul e leste-asiática) e 135 variantes exclusivas apenas encontradas em africanos. De acordo com o padrão de variação encontrado, os autores sustentam que seus resultados condizem com a teoria da migração dos humanos modernos a partir da África (Stringer, 1988).

1.5. Papel do ambiente

Estudos também apontam o ambiente como um fator de influência na diversidade da linguagem (Everett, 2017; Bentz *et al.*, 2018; Hua *et al.*, 2019). Sabe-se que climas temperados com vegetação aberta permitem uma melhor transmissão de consoantes e/ou sons de alta frequência do que climas mais quentes com vegetação densa, com melhor propagação de vogais (Maddieson, 2011). Outros aspectos ambientais que variam espacialmente, tal como a altitude, também afetam a fisiologia da produção do som distintamente em cada população, refletindo nos sons da fala das diferentes linguagens (Everett, 2013). Everett e colaboradores (2015) sugeriram que locais quentes e úmidos são regiões onde populações falantes de idiomas tonais se encontram, enquanto populações que utilizam línguas não-tonais são mais predominantes em regiões secas (Figura 3). Os autores argumentam que a umidade do ar poderia influenciar na vibração das cordas vocais e isso explicaria, em parte, o padrão de distribuição global dessas populações. Por exemplo, linguagens em climas mais frios são menos prováveis de utilizar tons lexicais, uma vez que o ar seco pode impedir um controle vocal necessário para produção de distinções tonais (Everett, Blasi & Roberts, 2015; Everett, Blasi & Roberts, 2016).

Figura 3. Distribuição de linguagens tonais (em vermelho) e não-tonais (em azul).



Adaptado de Everett, Blasi & Roberts (2015).

1.6. Estudos macroevolutivos

Uma oportuna ferramenta para entender como os diferentes tipos de vocalizações evoluíram é analisar as assinaturas deixadas no genoma que permitiram esse processo, uma vez que a comunicação não deixa evidência fóssil.

1.6.1. Estudo filogenético das espécies e das moléculas

A filogenética é um campo da ciência que estuda a história evolutiva das espécies a partir de dados como a morfologia e dados moleculares. O uso de filogenias é bastante utilizado e representa graficamente essas relações. É com base na construção dessas filogenias que, utilizando os dados mencionados anteriormente, conseguimos visualizar espécies ou grupos que compartilham ancestrais em comum.

A filogenia molecular envolve o estudo de análises comparativas de sequências gênicas e de aminoácidos, bem como de características estruturais de proteínas a fim de entender as histórias evolutivas e suas relações, e, em alguns casos até inferir sua função. Por exemplo, conseguimos classificar os organismos em *archaea*, *bacteria* e *eucaria* com base na comparação de sequência das subunidades do RNA ribossomal (rRNA). Para poder construir as análises de filogenia molecular, precisamos escolher um alvo a ser analisado (toda a sequência de uma proteína ou gene, ou partes importantes como motivos ou domínios proteicos) e obter o material genético das espécies que queremos investigar. Com base nisso, é possível identificar regiões conservadas ou sob seleção em uma ou mais espécies ou clados. Por exemplo, foi descrito que não há variabilidade na sequência de oito aminoácidos do neuro-hormônio ocitocina (OXT) entre mamíferos placentários. Porém, Vargas-Pinilla e colaboradores (2015) analisaram as sequências gênicas do gene *OXT*, codificador da

oxitocina, e encontraram formas taxon-específicas em macacos do Novo Mundo (clado Platyrrhini), incluindo sinal da ação de seleção positiva em sítios variantes. Vargas-Pinilla et al. (2015), e posteriormente vários outros autores, correlacionaram algumas dessas formas de ocitocina com parto gemelar, cuidado paterno e monogamia social, traços adaptativos especialmente presentes em micos e saguis. Posteriormente, foi demonstrado que duas dessas variantes diminuía ou mesmo impedia a internalização do receptor de OXT (Parreiras-e-Silva et al., 2017). A capacidade das variantes estimular o cuidado paterno em ratos também foi mostrada no mesmo estudo (Parreiras-e-Silva, Vargas-Pinilla et al., 2017).

1.6.2. Paralelismo e co-evolução

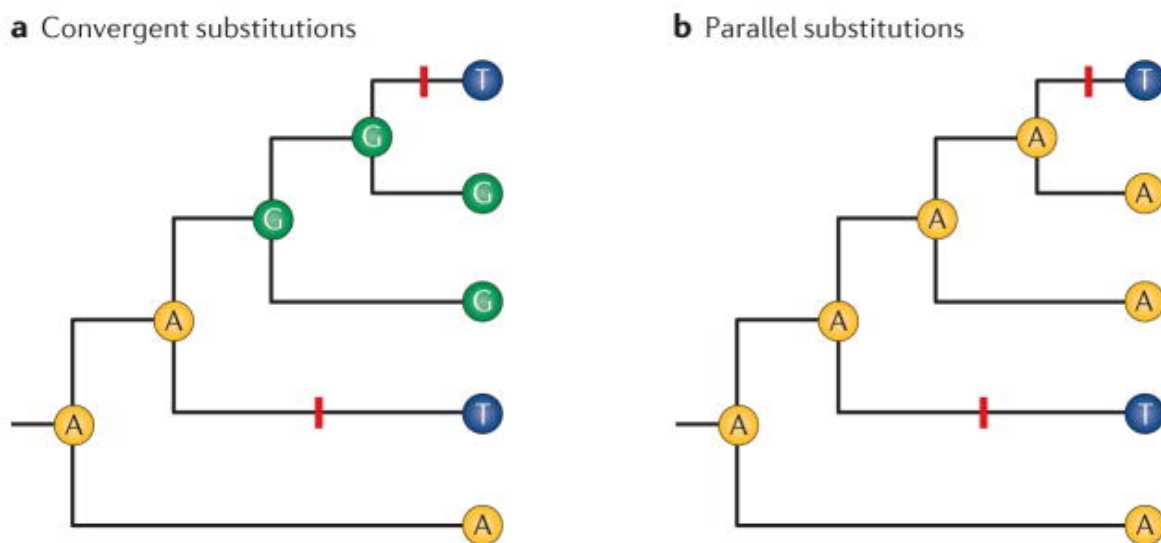
A partir dos estudos filogenéticos, é possível identificar eventos como os de convergência evolutiva, em que características comuns surgem em grupos filogeneticamente distantes. Exemplos desse mecanismo são a asa em morcegos e das aves, e a ecolocalização em golfinhos e morcegos

Uma questão ainda em aberto é a diferença entre convergência evolutiva e paralelismo. Para alguns autores, essa diferença na verdade não existe. No entanto, outros apontam quatro diferenças: os estados de caráter homoplásico (Patterson 1982, 1988); o compartilhamento de características independentes entre táxons intimamente relacionados (Scott 1896; Arendt & Reznick 2008a, b; Leander 2008); traços independentes que compartilham os mesmos estados de caráter ancestral (Hennig 1966; Reidl 1979; Leander 2008); e, por fim, traços independentes causados pelo mesmo repertório genético (Haas & Simpson 1946; Brundin 1976; Gould 2002).

O paralelismo compartilha aspectos com o conceito de homologia padrão, uma vez que os geradores do desenvolvimento subjacentes são homólogos. Também compartilha aspectos com a convergência, uma vez que há o desenvolvimento independente de características semelhantes. Porém, a convergência difere do paralelismo em relação ao ancestral comum. A convergência se refere a mudanças independentes para o mesmo fenótipo de táxons com ancestrais diferentes, enquanto no paralelismo, os táxons compartilham o mesmo ancestral. As mudanças paralelas em nível molecular nem sempre estarão associadas com fenótipos convergentes, porém, se esse for o caso, há homoplasia no nível genotípico e fenotípico. É nesse sentido que o paralelismo pode explicar como os fenótipos convergentes evoluíram (Gould, 2002).

A figura abaixo (Figura 4) mostra uma comparação entre paralelismo e convergência, através da análise da história evolutiva de uma proteína hipotética. As substituições convergentes para o mesmo aminoácido na mesma posição de uma proteína de duas espécies diferentes podem ter evoluído de maneira independente de um mesmo aminoácido ancestral (paralelismo) ou de aminoácidos ancestrais diferentes (convergência).

Figura 4. Diferenças entre convergência e paralelismo na história evolutiva de uma proteína hipotética.



Autoria de Storz, 2016.

1.7. Estudos microevolutivos

1.7.1. Seleção e deriva

A evolução, em nível populacional, ocorre pela alteração na frequência de alelos ao longo das gerações. Existem quatro principais mecanismos que alteram as frequências alélicas em populações naturais: mutação, migração ou fluxo gênico, deriva genética e seleção natural.

A mutação é a principal fonte de variabilidade no genoma. É a partir dela que há a criação de diferentes versões dos mesmos genes, os alelos. Aqui, torna-se oportuna a conceituação de alelo ancestral e alelo derivado. Alelo derivado é o alelo que surge devido a uma mutação, enquanto o alelo ancestral é aquele que manteve seu estado inicial, sendo o alelo presente no ancestral comum de um táxon (Fay & Wu, 2000).

A variação genética produzida pela mutação é a fonte que permite às populações se adaptarem ao longo do tempo. As variações que ajudam na sobrevivência e reprodução de um organismo são passadas para a próxima geração, enquanto as variações prejudiciais são eliminadas. No entanto, as mutações podem ser neutras, isto é, não serem vantajosas ou prejudiciais. Os alelos gerados por mutação neutra podem se manter na população. Dito isso, sem a mutação, mecanismos evolutivos como a seleção não conseguem atuar. Com o surgimento de novos alelos, outras fontes de variabilidade também agem, como a recombinação e o cruzamento aleatório que misturam esses diferentes alelos em novas combinações. Porém, a taxa de mutação geralmente é muito baixa e, por isso, esse não é o principal mecanismo que altera a frequência dos alelos, embora extremamente importante.

Mecanismos evolutivos como a deriva genética e a migração alteram a frequência alélica de forma aleatória e não porque um determinado alelo ou outro é vantajoso na população. Em relação à deriva, esta é muito mais notável quanto menor for uma população. A deriva genética funciona como dados sendo lançados ao alto, não há como prever o que vai ocorrer e a chance de que um alelo seja mais frequente é a mesma de ser menos frequente. Em cada geração, alguns indivíduos podem deixar mais descendentes do que outros indivíduos apenas por acaso, não porque são mais adaptados. A deriva genética é bastante comum e pode ser observada após eventos de gargalo populacional, em que ocorrem diminuições drásticas no tamanho da população. Nesses momentos, pode ocorrer a perda e ou a fixação de alelos e, por consequência, a diminuição da variabilidade. Sendo assim, mutação e deriva atuam juntas, embora em sentidos diferentes. Motoo Kimura já havia observado essa relação quando propôs a teoria neutra da evolução molecular. Segundo essa teoria, boa parte da variação genética é explicada pelo balanço entre esses dois mecanismos. A mutação cria lentamente novas variações alélicas e a deriva lentamente elimina essa variabilidade, alcançando um estado de estabilidade.

Em relação à migração, ou também chamado de fluxo gênico, como dito anteriormente, é um mecanismo que altera a frequência dos alelos. O movimento de indivíduos (ou gametas) de uma população em direção a outra leva a introgressão gênica. Esse mecanismo inicialmente aumenta a variabilidade, tornando possível novas combinações genéticas. Quanto maior a diferença na frequência alélica entre a população local e os indivíduos migrantes e quanto maior o número de migrantes, maior é o efeito que estes têm em alterar a constituição genética da população local.

A variação genética em populações humanas tem sido altamente moldada pelo fluxo gênico. Em um estudo recente, pesquisadores analisaram o DNA mitocondrial (mtDNA) de humanos arcaicos e observaram que caçadores-coletores pré-históricos do Nordeste da Europa são similares a outros caçadores-coletores europeus. Além disso, os pesquisadores também observaram que amostras de humanos arcaicos de uma região da Rússia apresentaram uma elevada diversidade de linhagens de mtDNA com origem na Sibéria (haplogrupos C, D e Z) (der Sarkissian *et al.*, 2013). Esses haplogrupos são comuns em regiões da Ásia central, do norte e leste asiático, provavelmente sendo originários do leste da Ásia, e se expandiram através de muitas migrações após o último máximo glacial (Velichko *et al.*, 2009).

Apesar de ser uma fonte de variabilidade, o fluxo gênico tende a levar à homogeneização das populações. Por outro lado, a perda de fluxo gênico entre as populações permite que se diferenciem entre si e, em casos extremos, pode levar à especiação.

Por fim, outro mecanismo que altera a frequência alélica é a seleção natural. Diferente dos demais mecanismos citados anteriormente, a seleção natural faz isso de maneira não aleatória, em que mudanças no genoma, que impactam positivamente a sobrevivência e a taxa de reprodução dos indivíduos, são mantidas através das gerações no caso da seleção positiva. Ou, quando alelos deletérios diminuem de frequência através da seleção purificadora.

Por exemplo, foi demonstrado que três genes (*SP100*, *DUOX2* e *CLC*) apresentam indícios de seleção positiva, podendo estar relacionados à adaptação à altitude em populações dos Andes (Jacovas *et al.*, 2018). A adaptação ao frio também foi demonstrada em um estudo recente. Os autores analisaram as ácido graxo dessaturases, enzimas que removem átomos de hidrogênio de um ácido graxo, e constataram que os genes que codificam essas enzimas apresentam sinais de seleção natural em populações da América. O achado sugere que houve um evento de adaptação que ocorreu devido à passagem no estreito de Bering, antes da expansão da população no continente americano (Amorim *et al.*, 2017).

Além disso, o caso do Complexo de Histocompatibilidade (MHC) tem sido bastante documentado na literatura como alvo de seleção balanceadora (Hedrick, 1998; Talarico *et al.*, 2021; Ekblom *et al.*, 2010; Bernatchez & Landry, 2003; Piertney and Oliver, 2006; Spurgin and Richardson, 2010). Porém, em um artigo recente, um grupo de pesquisadores analisou a diferenciação nos genes MHC em populações nativo americanas, através da comparação de microssatélites distribuídos ao longo do genoma com aqueles na região do MHC. Foi

encontrada maior variabilidade nos microssatélites próximos ao *locus* MHC do que no restante do genoma (Nunes *et al.*, 2021). Em outro estudo de 2019, os pesquisadores relataram que habitantes de regiões ricas em patógenos são mais propensos a possuírem alelos MHC classe II DRB1 de baixa especificidade. Assim, as posições moleculares que definem a especificidade dessas moléculas de MHC classe II estão sob seleção positiva em populações humanas (Manczinger *et al.*, 2019).

Existem diversas formas de investigar as diferenças genéticas entre as populações e inferir os mecanismos evolutivos envolvidos. A divergência entre populações pode ser calculada através das estatísticas F de Wright. Entre elas, o F_{ST} que mede a diferenciação entre populações e por consequência, pode-se calcular também o grau de fluxo gênico entre populações. O F_{ST} é uma medida bastante conhecida em genética de populações e pode indicar alvos de seleção positiva no genoma. Quando se comparam populações par-à-par (*pairwise* F_{ST}), busca-se identificar se duas populações diferem de modo incompatível com sua história demográfica e relações filogenéticas estimadas por marcadores genéticos neutros (Weir *et al.*, 2005; Cavalli-Sforza, 1966). Desse modo, a estrutura populacional, ou seja, o nível de diferenciação entre populações pode ser fortemente influenciada pela seleção natural (Nosil *et al.*, 2009; Orsini *et al.*, 2013).

Vale lembrar que abordagens utilizando *pairwise* F_{ST} , que incluem métodos Bayesianos, foram desenvolvidas para identificar divergências genéticas que podem indicar diferentes contextos adaptativos considerando duas populações naturais (Beaumont & Balding, 2004; Foll & Gaggiotti, 2008).

Muitos exemplos de aplicação de F_{ST} como teste para se detectar ação da seleção natural podem ser encontrados na literatura. Por exemplo, em um estudo de Alonso e colaboradores, foi demonstrado que *clusters* de SNPs ao longo da região genômica onde se encontram os genes *TP53* e *WDR79* apresentavam altos valores de F_{ST} , considerando uma amostra da população de Utah (EUA), com alta ancestralidade europeia e uma representativa dos Yorubás da Nigéria (Alonso *et al.*, 2010). No estudo, tanto os resultados da análise de F_{ST} como os do teste de homozigosidade haplotípica estendida (EHH; outro teste usado para se detectar seleção) convergiram, o que reforça a robustez da análise usando F_{ST} como indicador de seleção positiva.

2. OBJETIVOS

2.1 Objetivo geral

Avaliar, através de estudos macro e microevolutivos, se variantes genéticas em genes relacionados à audição ou à fala estariam conectados ao tipo de comunicação ou linguagem.

2.2 Objetivos específicos

2.2.a Estudo macroevolutivo

1. Buscar, em bancos de dados específicos, genes relacionados à audição.
2. Comparar as sequências codificadoras desses genes entre seres humanos, primatas não-humanos, artiodátilos, cetáceos e quirópteros em busca de variantes táxon-específicas e assinaturas de seleção natural relacionadas à comunicação.
3. Determinar a frequência das variantes sob seleção em humanos modernos (*Homo sapiens*).
4. Avaliar, através de preditores *in silico*, se algumas dessas variantes teriam algum efeito sobre a proteína em seres humanos.

2.2.b Estudo microevolutivo

1. Buscar, em bancos de dados específicos, genes relacionados à fala.
2. Verificar se há indícios de seleção positiva nos genes da fala que possam, em parte, explicar as diferenças entre os dois grandes grupos linguísticos tonal e não tonal.

3. MÉTODOS

3.1 Métodos do estudo macroevolutivo

3.1.1 Levantamento de dados para a análise macroevolutiva

O primeiro passo foi buscar genes relacionados a “hearing frequency” no “Human Phenotype Ontology” (HPO), de maneira a encontrar os genes candidatos para a análise. Nesse sentido, a abordagem deste estudo foi considerar estudos com doenças humanas relacionadas aos fenótipos de interesse como evidência na escolha de genes candidatos.

Em um segundo passo, foram recuperadas sequências codificadoras dos genes previamente selecionados no passo 1 do *Homo sapiens* no Ensembl, usando a versão GRCh38 do genoma humano. Para escolher a melhor sequência codificadora neste banco, foram selecionadas aquelas que tiveram o mesmo tamanho da sequência encontrada no Uniprot. As sequências ortólogas de primatas, cetáceos e quirópteros foram obtidas no NCBI através do BlastN, utilizando o parâmetro “discontinuous megablast”. As etapas, descritas neste parágrafo, foram automatizadas através de um script escrito em Python (versão 3.7).

3.1.2 Alinhamento das sequências codificadoras

Os alinhamentos múltiplos (utilizando mais de duas sequências) foram feitos no Software MEGA (versão X) usando o algoritmo “Muscle”. Antes do alinhamento, as sequências foram selecionadas seguindo os seguintes critérios: a) das sequências obtidas através do BlastN, foram escolhidas aquelas que começassem com o códon de iniciação ATG e que apresentassem algum dos códons de terminação no fim da sequência. Se após isso ainda restasse mais de um transcrito do mesmo organismo, a sequência que permaneceu foi escolhida aleatoriamente; b) após o alinhamento, as sequências que apresentaram muitos lacunas (gaps), as quais poderiam trazer falsos positivos na análise evolutiva, passaram por uma curadoria e recorte no *software* TrimAL (versão 1.2) usando o método “gappyout” (Capella-Gutierrez *et al.*, 2009); c) após, o alinhamento foi processado no *software* “Genetic Algorithm Recombination Detection” (GARD; Kosakovsky Pond *et al.*, 2006) de modo a

identificar possíveis sítios recombinantes. Foi usada a opção “General Discrete” para a variação “site-to site” e “Rate Classes” igual a 3 para todos os alinhamentos. Nas sequências com evidência de eventos recombinantes, foi utilizada aquela fornecida pelo GARD após a análise, para evitar falsos positivos.

3.1.3 Análises evolutivas das sequências ortólogas

Após o processamento dos alinhamentos, as análises evolutivas foram realizadas no *software* “Phylogenetic Analysis by Maximum Likelihood” (PAML; versão 4.9; Yang, 2007), usando o pacote Codeml. O Codeml compara o número de taxas de substituições sinônimas e não sinônimas ($\omega = dN/dS$). O pacote detecta seleção positiva ($\omega > 1$), evolução neutra ($\omega = 1$), ou seleção purificadora ($\omega < 1$) nos códons.

Para detectar seleção positiva, foram comparados modelos neutros, com valores de ômega (ω) fixados (M1 e M8a) com modelos de seleção positiva (M2a e M8), com valores de ω variáveis. Os modelos foram comparados utilizando o Teste da Razão de Verossimilhança, ou “Likelihood Ratio Test” (LRT), montando os seguintes pares: M1 *versus* M2a ($gl = 2$), M8a *versus* M8 ($gl = 1$) (Yang, 1998). Para detectar quais sítios na sequência estão sob seleção positiva, foi utilizada a abordagem de *Bayes Empirical Bayes* (BEB), que também é implementado no Codeml. Para determinar os sítios estatisticamente significativos no BEB, foi utilizado um corte de significância de p -valor $< 0,05$.

Além disso, para testar a hipótese de que há sítios individuais sujeitos à seleção positiva, foi utilizado o Software “Mixed Effects Model of Evolution” (MEME) (Murrell *et al.*, 2012), um software implementado no pacote “Hyphy”. Esse *software* também incorpora análises de taxas de substituições sinônimas e não sinônimas por sítio. O MEME também é uma ferramenta para detectar seleção diversificadora que pode ocorrer na filogenia como um todo ou apenas em uma ou poucas linhagens (seleção episódica). Estabelecemos um valor $p \leq 0,05$ para MEME para limiares de significância estatística.

3.1.4 Reconstrução das filogenias

As filogenias das espécies utilizadas foram obtidas usando o *website* “TimeTree” (<http://www.timetree.org>) e modificadas para incluir informações sobre a presença ou ausência de ecolocalização, vocalização na faixa do ultrassom e hábitos noturnos e diurnos

bem como quais sítios estão sob seleção na análise descrita anteriormente. Essas informações fenotípicas foram obtidas através da literatura e as referências se encontram na Tabela Suplementar 1. As figuras das filogenias e as informações adicionais foram feitas no *software* Adobe Photoshop.

3.1.5 Frequência de variantes no gênero *Homo*

Além disso, foi analisada a frequência de variantes nos sítios detectados pelo MEME e PAML para o *Homo sapiens*. Identificou-se os códons e os polimorfismos de nucleotídeo único (SNPs) nessas posições através do *Genome Data Viewer* (GDV) (<https://www.ncbi.nlm.nih.gov/genome/gdv/>). Isso foi feito com o auxílio da sequência de *Homo sapiens* obtida para o alinhamento descrito anteriormente. Os ids de cada variante foram anotados em uma tabela e as frequências alélicas foram obtidas no dbSNP (<https://www.ncbi.nlm.nih.gov/snp/>). A frequência considerada foi aquela descrita pelo projeto *Allele Frequency Aggregator* (ALFA), ou, quando ausente no ALFA, o dado disponibilizado pelo dbSNP.

Finalmente, foi testado se as variantes presentes em outros grupos nas posições sob seleção teriam algum impacto sobre a proteína humana. Para isso, foi usado o seguinte preditor *in silico*: PolyPhen (Adzhubei *et al.*, 2010). Foi considerado o alinhamento da sequência utilizado acima.

3.2 Métodos do estudo microevolutivo

3.2.1 Levantamento dos dados populacionais e linguísticos

Para este estudo foram utilizados dados disponíveis no banco 1.000 Genomas (<https://www.internationalgenome.org/>), que engloba, além das 26 populações estudadas pelo projeto, dados genéticos do *Human Genome Diversity Project* (HGDP). Os dados linguísticos de cada população com dado genético disponível foram obtidos no *The World Atlas of Language Structure* (WALS; <https://wals.info/>). Os dados de posição geográfica das populações do 1.000 Genomas foram comparados à localização das populações com idioma

classificado em língua tonal ou não tonal no WALS, utilizando o *software* Tableau (2021.1). Foram selecionadas apenas populações próximas falantes de língua tonal e não-tonal de acordo com o WALS e com tamanho de amostra acima de 10 indivíduos representados no 1.000 Genomas.

3.2.2 Escolha das populações analisadas e da análise de seleção positiva

Para analisar se há genes sob seleção e se poderiam estar associados com o tipo de língua falada (se tonal ou não tonal), foi escolhido o método *pairwise* F_{ST} para os 22 pares de cromossomos autossômicos apenas.

A comparação do F_{ST} deve ser feita entre duas populações geograficamente próximas, e nesse caso, uma falante de língua tonal e outra falante de língua não tonal. Para uma análise inicial, escolhemos as populações de Vietnamitas (falante de língua tonal) e Cambojanos (falante de língua não tonal) com base na figura gerada pelo Tableau.

3.2.3 Obtenção dos dados genéticos e análise do *pairwise* F_{ST}

Os dados genéticos das populações foram obtidos no 1.000 Genomas na forma de CRAM. Os arquivos foram processados para geração do VCF através de um programa desenvolvido em parceria com o Instituto de Informática da UFRGS (grupo do prof. Cláudio Geyer), que desenvolveu um *script* baseado no programa SAMtools (Farah, 2021).

O cálculo do F_{ST} foi realizado no VCFtools e as etapas posteriores foram feitas no R (versão 3.6). O *Manhattan plot* foi feito com o auxílio da biblioteca ggplot2. O *cut-off* escolhido foi de 0.01%. Os SNPs com valores de F_{ST} entre os 0,01% mais altos foram anotados usando o “Variant Effect Predictor” (VEP) e as ontologias em que esses genes participam foram obtidas utilizando o FUMA GWAS. Foi utilizada correção de Benjamini-Hochberg (FDR), um p -valor de 0,05 e no mínimo 2 genes sobrepostos entre os conjuntos. Os fenótipos associados foram obtidos no *GWAS Catalog* (<https://www.ebi.ac.uk/gwas/>).

3.2.4 Levantamento dos genes candidatos

A seleção de genes candidatos para fala ocorreu através da busca de fenótipos associados à fala no *Human Phenotype Ontology* (HPO). O termo *speech* foi utilizado na busca. O nome dos genes foi obtido através de cada seção de cada fenótipo e uma lista com todos os genes foi gerada para análise nas etapas posteriores.

3.2.5 Análise de topologia de rede de interação dos genes candidatos

Para entender quais seriam os genes mais importantes da lista que foi gerada com a busca no HPO, foi montada uma rede de interação desses genes utilizando o STRING (versão WEB; disponível em: <https://string-db.org>). Escolheu-se o score de interação de 0,04 e tamanho da primeira e segunda camada da rede igual a 100. As fontes de interação, de onde são obtidas as informações para criação da rede, foram escolhidas removendo as fontes *text mining* e *neighborhood*. Após, o arquivo contendo informações da rede foi utilizado para a análise no software Cytoscape (versão 3.9.1). O *plugin* escolhido para a análise de topologia foi o Cytohubba (Chin *et al.*, 2014). Utilizamos o método MCC, que é mais efetivo para encontrar *hub nodes*, e buscamos pelos 20 genes mais importantes da rede.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., et al. (2010). A method and server for predicting damaging missense mutations. *Nature Methods*, 7(4), 248–249. doi:10.1038/nmeth0410-248.
- Albert, S., Blons, H., Jonard, L., Feldmann, D., Chauvin, P., Loundon, N., ... & Denoyelle, F. (2006). SLC26A4 gene is frequently involved in nonsyndromic hearing impairment with enlarged vestibular aqueduct in Caucasian populations. *European journal of human genetics*, 14(6), 773-779.
- Alonso, S., Izagirre, N., López, S., Smith-Zubiaga, I., Hervella, M., Boyano, M. D. et al. (2010). The Diversity Profile of TP53 Is Influenced by Positive Selection on the Immediately Upstream Locus WDR79. *Human Heredity*, 69(1), 34–44.
- Amorim, C. E., Nunes, K., Meyer, D., Comas, D., Bortolini, M. C., Salzano, F. M., & Hünemeier, T. (2017). Genetic signature of natural selection in first Americans. *Proceedings of the National Academy of Sciences*, 114(9), 2195–2199. doi:10.1073/pnas.1620541114.
- Arendt, J., Reznick, D. (2008a). Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends Ecol. Evol.* 23: 26–32.
- Arendt, J., Reznick, D. (2008b). Moving beyond phylogenetic assumptions about evolutionary convergence: response to Leander. *Trends Ecol. Evol.* 23: 483–484.
- Arikawa-Hirasawa, E., Le, A. H., Nishino, I., Nonaka, I., Ho, N. C., Francomano, C. A., et al. (2002). Structural and Functional Mutations of the Perlecan Gene Cause Schwartz-Jampel Syndrome, with Myotonic Myopathy and Chondrodysplasia. *The American Journal of Human Genetics*, 70(5), 1368–1375. doi:10.1086/340390.
- Arnold, K. & Zuberbühler, K. (2006a). Semantic combinations in primate calls. *Nature* 441:303–303.
- Arnold, K. & Zuberbühler, K. (2006b). The alarm-calling system of adult male putty-nosed monkeys, *Cercopithecus nictitans martini*. *Anim Behav* 72:643–653. doi: 10.1016/j.anbehav.2005.11.017.
- Au, W. W. L. (2009). Echolocation. *Encyclopedia of Marine Mammals*, 348–357. doi:10.1016/b978-0-12-373553-9.00085-7.
- Batissoco, A. C., Abreu-Silva, R. S., Braga, M. C. C., Lezirovitz, K., Della-Rosa, V., Alfredo Jr, T., ... & Mingroni-Netto, R. C. (2009). Prevalence of GJB2 (connexin-26) and GJB6 (connexin-30) mutations in a cohort of 300 Brazilian hearing-impaired individuals: implications for diagnosis and genetic counseling. *Ear and hearing*, 30(1), 1-7.

- Beaumont, M. A., & Balding, D. J. (2004). Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, 13(4), 969–980.
- Bee, M. A. (2015). Treefrogs as animal models for research on auditory scene analysis and the cocktail party problem. *International Journal of Psychophysiology*, 95(2), 216–237. doi:10.1016/j.ijpsycho.2014.01.004.
- Belyantseva, I. A., Boger, E. T., & Friedman, T. B. (2003). Myosin XVa localizes to the tips of inner ear sensory cell stereocilia and is essential for staircase formation of the hair bundle. *Proceedings of the National Academy of Sciences*, 100(24), 13958–13963. doi:10.1073/pnas.2334417100.
- Bentz, C., Dediu, D., Verkerk, A., Jager, G. (2018) The evolution of language families is shaped by the environment beyond neutral drift. *Nature Human Behaviour* Springer US, 2(11), pp. 816–821.
- Bernatchez, L., & Landry, C. (2003). MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *Journal of Evolutionary Biology*, 16(3), 363–377.
- Berthelsen, S. G., Horne, M., Shtyrov, Y., & Roll, M. (2020). Different neural mechanisms for rapid acquisition of words with grammatical tone in learners from tonal and non-tonal backgrounds: ERP evidence. *Brain Research*, 1729, 146614.
- Best, C. T. (2019). The Diversity of Tone Languages and the Roles of Pitch Variation in Non-tone Languages: Considerations for Tone Perception Research. *Frontiers in Psychology*, 10. doi:10.3389/fpsyg.2019.00364.
- Blasi, D. E., Moran, S., Moisiuk, S. R., Widmer, P., Dediu, D., & Bickel, B. (2019). Human sound systems are shaped by post-Neolithic changes in bite configuration. *Science*, 363(6432), eaav3218. doi:10.1126/science.aav3218.
- Bolhuis, J. J., Beckers, G. J. L., Huybregts, M. A. C., Berwick, R. C., & Everaert, M. B. H. (2018). Meaningful syntactic structure in songbird vocalizations? *PLOS Biology*, 16(6), e2005157. doi:10.1371/journal.pbio.2005157.
- Bradbury, J. W. & Vehrencamp, S. L. (1998). *Principles of Animal Communication*. Sunderland, Massachusetts: Sinauer.
- Bromham, L., Hua, X., Fitzpatrick, T. G., & Greenhill, S. J. (2015). Rate of language evolution is affected by population size. *Proceedings of the National Academy of Sciences*, 112(7), 2097–2102. doi:10.1073/pnas.1419704112.
- Brundin, L. (1976). A Neocomian chironomid and Podonominae–Aphroteniinae (Diptera) in the light of phylogenetics and biogeography. *Zool. Scr.* 1: 107–120.
- Campbell, L. & Poser, W. J. (2008) *Language Classification: History and Method.*, Cambridge: Cambridge University Press. Cambridge University Press (CUP).

- Capella-Gutierrez, S., Silla-Martinez, J. M., & Gabaldon, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, 25(15), 1972–1973. doi:10.1093/bioinformatics/btp348.
- Carneiro, P., de Freitas, M. V., Matte, U. (2022). In silico analysis of potential off-target sites to gene editing for Mucopolysaccharidosis type I using the CRISPR/Cas9 system: Implications for population-specific treatments. *PLoS One*. 24;17(1):e0262299. doi: 10.1371/journal.pone.0262299. PMID: 35073349; PMCID: PMC8786118.
- Carroll, S. P., Dingle, H., & Famula, T. R. (2003). Rapid appearance of epistasis during adaptive divergence following colonization. *Proceedings of the Royal Society B: Biological Sciences*, 270(Suppl_1), S80–S83. doi:10.1098/rsbl.2003.0019.
- Cavalli-Sforza, L. L. (1966). Population Structure and Human Evolution. *Proceedings of the Royal Society B: Biological Sciences*, 164(995), 362–379. doi:10.1098/rspb.1966.0038.
- Chandra, S. R., Issac, T. G., Gayathri, N., Shivaram, S. (2015). Schwartz-Jampel syndrome. *J Pediatr Neurosci.* ;10(2):169-71. doi: 10.4103/1817-1745.159202.
- Chen, P., Wu, W., Zhang, J., Chen, J., Li, Y., Sun, L., ... & Yang, J. (2022). Pathological mechanisms of connexin26-related hearing loss: Potassium recycling, ATP-calcium signaling, or energy supply?. *Frontiers in Molecular Neuroscience*, 15, 976388.
- Cheng, J., Nguyen, T., Cygan, K. et al. MMSplice: modular modeling improves the predictions of genetic variant effects on splicing. *Genome Biol* 20, 48 (2019).
- Chien, P., Friederici, A. D., Hartwigsen, G., & Sammler, D. (2020). Neural correlates of intonation and lexical tone in tonal and non-tonal language speakers. *Human Brain Mapping*. doi:10.1002/hbm.24916.
- Chin, C.-H., Chen, S.-H., Wu, H.-H., Ho, C.-W., Ko, M.-T., & Lin, C.-Y. (2014). cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Systems Biology*, 8(Suppl 4), S11. doi:10.1186/1752-0509-8-s4-s11.
- Chomsky, N. (2017). The language capacity: architecture and evolution. *Psychonomic Bulletin & Review*, 24(1), 200–203. doi:10.3758/s13423-016-1078-6.
- Clay Z & Zuberbühler K (2011) Bonobos Extract Meaning from Call Sequences. *PLOS ONE* 6:e18786. doi: 10.1371/journal.pone.0018786.
- Clulow, S., Mahony, M., Elliott, L., Humfeld, S., & Gerhardt, H. C. (2016). Near-synchronous calling in the hip-pocket frog *Assa darlingtoni*. *Bioacoustics*, 26(3), 249–258. doi:10.1080/09524622.2016.1260054.
- Cooper, D. L. (2006). Broca's arrow: Evolution, prediction, and language in the brain. *The Anatomical Record Part B: The New Anatomist*, 289B(1), 9–24. doi:10.1002/ar.b.20088.

- Cram, Dominic L., et al. (2022). The ecology and evolution of human-wildlife cooperation. *People and Nature* 4.4 841-855.
- Creanza, N., Ruhlen, M., Pemberton, T. J., Rosenberg, N. A., Feldman, M. W., & Ramachandran, S. (2015). A comparison of worldwide phonemic and genetic variation in human populations. *Proceedings of the National Academy of Sciences*, 112(5), 1265–1272. doi:10.1073/pnas.1424033112.
- Dasgupta, W. (2014). *Clinical Chemistry, Immunology and Laboratory Quality Control: A Comprehensive Review for Board Preparation, Certification and Clinical Practice*. Elsevier. <https://doi.org/10.1016/C2012-0-06507-6>.
- Deardorff, M. A., Kaur, M., Yaeger, D., Rampuria, A., Korolev, S., Pie, J. et al. (2007). Mutations in Cohesin Complex Members SMC3 and SMC1A Cause a Mild Variant of Cornelia de Lange Syndrome with Predominant Mental Retardation. *The American Journal of Human Genetics*, 80(3), 485–494.
- Dediu, D., & Ladd, D. R. (2007). Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, ASPM and Microcephalin. *Proceedings of the National Academy of Sciences*, 104(26), 10944–10949. doi:10.1073/pnas.0610848104.
- Dediu, D., Janssen, R., & Moisik, S. R. (2019). Weak biases emerging from vocal tract anatomy shape the repeated transmission of vowels. *Nature Human Behaviour*. doi:10.1038/s41562-019-0663-x.
- Del Castillo, I., Moreno-Pelayo, M. A., Del Castillo, F. J., Brownstein, Z., Marlin, S., Adina, Q., ... & Moreno, F. (2003). Prevalence and evolutionary origins of the del (GJB6-D13S1830) mutation in the DFNB1 locus in hearing-impaired subjects: a multicenter study. *The American Journal of Human Genetics*, 73(6), 1452-1458.
- Der Sarkissian, C., Balanovsky, O., Brandt, G., Khartanovich, V., Buzhilova, A., et al. (2013). Ancient DNA Reveals Prehistoric Gene-Flow from Siberia in the Complex Human Population History of North East Europe. *PLoS Genetics*, 9(2), e1003296. doi:10.1371/journal.pgen.1003296
- Dryer, M. S. & Haspelmath, M. (2013) *WALS Online*. Max Planck Institute for Evolutionary Anthropology, Leipzig.
- Duellman, W.E., Trueb, L. (1986). *Biology of Amphibians*. Mc Graw - Hill Book Company, 1221 Avenue of the Americas, New York, New York 10020, U.S.A.; Mc Graw - Hill Book Company GmbH, Postfach 630520, D-2000 Hamburg 63, Bundesrepublik Deutschland. 670 pp.
- Eberhard, D. M., Simons, G. F., & Fennig, C. D. (eds.) (2020). *Ethnologue: Languages of the World*. Twenty-third edition. Dallas, Texas: SIL International. Online version: <http://www.ethnologue.com>.

- Ekblom, R., Sæther, S. A., Fiske, P., Kålås, J. A., & Höglund, J. (2010). Balancing selection, sexual selection and geographic structure in MHC genes of Great Snipe. *Genetica*, 138, 453-461.
- Enard, W., Przeworski, M., Fisher, S. E., Lai, C. S. L., Wiebe, V., Kitano, T., ... Pääbo, S. (2002). Molecular evolution of FOXP2, a gene involved in speech and language. *Nature*, 418(6900), 869–872. doi:10.1038/nature01025.
- Evans, P. D., Gilbert, S. L., Mekel-Bobrov, N., Vallender, E. J., Anderson, J. R., Vaez-Azizi, L. M., Tishkoff, S. A., Hudson, R. R., Lahn, B. T. (2005). Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans. *Science* 309(5741), pp. 1717–1720.
- Everett, L.A. (2006). New insights into the role of pendrin (SLC26A4) in inner ear fluid homeostasis. *Novartis Found Symp.* 2006;273:213-25; discussion 225-30, 261-4.
- Everett, C. (2017) Languages in drier climates use fewer vowels. *Frontiers in Psychology* 8(JUL), pp. 1–15.
- Everett, C., Blasi, D. E., & Roberts, S. G. (2015). Climate, vocal folds, and tonal languages: Connecting the physiological and geographic dots. *Proceedings of the National Academy of Sciences*, 112(5), 1322–1327. doi:10.1073/pnas.1417413112.
- Everett, C., Blasí, D. E., & Roberts, S. G. (2016). Language evolution and climate: the case of desiccation and tone. *Journal of Language Evolution*, 1(1), 33–46. doi:10.1093/jole/lzv004.
- Everett, C. (2013) Evidence for Direct Geographic Influences on Linguistic Sounds: The Case of Ejectives. *PLoS ONE* 8(6), p. e65275.
- Fay, J. C., & Wu, C.-I. (2000). Hitchhiking Under Positive Darwinian Selection. *Genetics*, 155(3), 1405–1413. doi:10.1093/genetics/155.3.1405.
- Feder, M. E., Burggren, W. W. (1992). *Environmental physiology of the Amphibians*. Chicago:University of Chicago Press.
- Ferrannini, E., Perniola, T., Krajewska, G., Serlenga, L., & Trizio, M. (1982). Schwartz-Jampel Syndrome with Autosomal-Dominant Inheritance. *European Neurology*, 21(3), 137–146. doi:10.1159/000115471.
- Foll, M., & Gaggiotti, O. (2008). A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. *Genetics*, 180(2), 977–993. doi:10.1534/genetics.108.092221.
- Frankenberg, E. (1982). Vocal Behavior of the Mediterranean House Gecko, *Hemidactylus turcicus*. *Copeia*, 1982(4), 770. doi:10.2307/1444085.

- Friederici, A. D. (2012). The cortical language circuit: from auditory perception to sentence comprehension. *Trends in Cognitive Sciences*, 16(5), 262–268. doi:10.1016/j.tics.2012.04.001.
- Fu, L., Shi, Z., Luo, G., Tu, W., Wang, X., Fang, Z., & Li, X. (2014). Multiple microRNAs regulate human FOXP2 gene expression by targeting sequences in its 3' untranslated region. *Molecular Brain*, 7(1). doi:10.1186/s13041-014-0071-0
- Galeotti, P., Sacchi, R., Fasola, M., & Ballasina, D. (2005). Do mounting vocalizations in tortoises have a communication function? A comparative analysis. *Herpetological Journal*, 15(2), 61–71.
- Garstang, M. (2004). Long-distance, low-frequency elephant communication. *Journal of Comparative Physiology A*, 190, 791-805.
- Gerhardt, H. C., Huber, F., & Simmons, A. M. (2003). Acoustic Communication in Insects and Anurans: Common Problems and Diverse Solutions. *The Journal of the Acoustical Society of America*, 114(2), 559–559. doi:10.1121/1.1591773.
- Gettelfinger, J. D., & Dahl, J. P. (2018). Syndromic hearing loss: a brief review of common presentations and genetics. *Journal of pediatric genetics*, 7(01), 001-008.
- Gillam, E. (2011) An Introduction to Animal Communication. *Nature Education Knowledge* 3(10):70.
- Gould, S. J. (2002). *The Structure of Evolutionary Theory*. The Belknap Press, Cambridge, MA.
- Greenhill, S. J., Hua, X., Welsh, C. F., Schneemann, H., & Bromham, L. (2018). Population Size and the Rate of Language Evolution: A Test Across Indo-European, Austronesian, and Bantu Languages. *Frontiers in Psychology*, 9. doi:10.3389/fpsyg.2018.00576.
- Guerrero-Ferreira, R., Taylor, N. M., Arteni, A. A., Kumari, P., Mona, D., Ringler, P., ... & Stahlberg, H. (2019). Two new polymorphic structures of human full-length alpha-synuclein fibrils solved by cryo-electron microscopy. *Elife*, 8, e48907.
- Haas, O., and Simpson, G. G. (1946). Analysis of some phylogenetic terms, with attempts at redefinition. *Proc. Am. Philos. Soc.* 90: 319–349.
- Hall, B. G. (2008). *Phylogenetic Trees Made Easy*. Sinauer Associates, Sunderland, MA.
- Hammarström, H. (2016). Linguistic diversity and language evolution. *Journal of Language Evolution* 1(1), pp. 19–29. doi: 10.1093/jole/lzw002.
- Haque, M., Chatterjee, S., Rahman, M. S., Jamil, H. M., Akhtar, N. (2016). SNPs of ASPM suggested population differentiation and Genographic patterns among diverse population. *Enliven: J Genet Mol Cell Biol* 3(1), 001.

- Harari, Y. N. (2015). *Sapiens*. Harper.
- Hedrick, P.W. (1998). Balancing selection and MHC. *Genetica* 104, 207–214.
- Hennig, W. (1966). *Phylogenetic Systematics*. University of Illinois Press, Urbana.
- Hoelzel, A. R. (Ed.) (2002). *Marine mammal biology: an evolutionary approach*. Oxford: Blackwell Publishing.
- Hettiarachchi, D. S., Dissanayake, V. H. W., Bonnard, C., & Yu-Jin, A. N. G. (2018). Schwartz-Jampel Syndrome (SJS1), Type 1 a Novel Variant in the HSPG2 Gene. *J Clin Case Rep*, 8(1182), 2.
- Hua, X., Greenhil, S. J., Cardillo, M., Schneemann, H., Bromham, L. (2019) The ecological drivers of variation in global language diversity. *Nature Communications*, Springer US, 10(1), pp. 1–10.
- Hughes, K. A., & Leips, J. (2017). Pleiotropy, constraint, and modularity in the evolution of life histories: insights from genomic analyses. *Annals of the New York Academy of Sciences*, 1389(1), 76-91.
- Hurst, J. A., Baraitser, M., Auger, E., Graham, F., Norell, S. (1990). An extended family with a dominantly inherited speech disorder. *Dev Med Child Neurol*. Apr;32(4):352-5. doi: 10.1111/j.1469-8749.1990.tb16948.x. PMID: 2332125.
- Iakoucheva L. M., Radivojac, P., Brown, C. J., O'Connor, T. R., Sikes, J. G., Obradovic, Z., Dunker, A. K. (2004) The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res* 32:1037-1049.
- Jacovas, V. C., Couto-Silva, C. M., Nunes, K., Lemes, R. B., de Oliveira, M. Z., Salzano, F. M. et al. (2018). Selection scan reveals three new loci related to high altitude adaptation in Native Andeans. *Scientific Reports*, 8(1). doi:10.1038/s41598-018-31100-6.
- Jones, G. (2005). Echolocation. *Current Biology* Vol 15 No 13.
- Jones, D. (1944) Chronemes and tonemes. *Acta Linguistica Hafniensia* 4(1), pp. 11–10. doi:10.1080/03740463.1944.10410902.
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O. et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583–589. doi:10.1038/s41586-021-03819-2.
- Kenneson, A., Braun, K. V. N., & Boyle, C. (2002). GJB2 (connexin 26) variants and nonsyndromic sensorineural hearing loss: a HuGE review. *Genetics in Medicine*, 4(4), 258-274.
- Kimura, M. (1979). The Neutral Theory of Molecular Evolution. *Scientific American*, 241(5), 98–129.

- Kline, A. D., & Deardorff, M. (2020). CORNELIA DE LANGE SYNDROME. Cassidy and Allanson's Management of Genetic Syndromes, 207–223.
- Kline, A. D., Krantz, I. D., Sommer, A., Kliever, M., Jackson, L. G., FitzPatrick, D. R. et al. (2007). Cornelia de Lange syndrome: Clinical review, diagnostic and scoring systems, and anticipatory guidance. *American Journal of Medical Genetics Part A*, 143A(12), 1287–1296.
- Knornschild, M., Jung, K., Nagy, M., Metz, M., & Kalko, E. (2012). Bat echolocation calls facilitate social communication. *Proceedings of the Royal Society B: Biological Sciences*, 279(1748), 4827–4835. doi:10.1098/rspb.2012.1995.
- Korver, A. M., Smith, R. J., Van Camp, G., Schleiss, M. R., Bitner-Glindzicz, M. A., Lustig, L. R., ... & Boudewyns, A. N. (2017). Congenital hearing loss. *Nature reviews Disease primers*, 3(1), 1-17.
- Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H., & Frost, S. D. W. (2006). Automated Phylogenetic Detection of Recombination Using a Genetic Algorithm. *Molecular Biology and Evolution*, 23(10), 1891–1901. doi:10.1093/molbev/msl051.
- Krause, J., Lalueza-Fox, C., Orlando, L., Enard, W., Green, R.E., Burbano, H.A., Hublin, J.J., Hanni, C., Fortea, J., de la Rasilla, M., et al. (2007). The derived FOXP2 variant of modern humans was shared with Neandertals. *Curr. Biol.* 17, 1908–1912.
- Kumar, A. (2003). Acoustic communication in birds. *Resonance*, 8(6), 44–55. doi:10.1007/bf02837868.
- Lai, C. S. L., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F., & Monaco, A. P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature*, 413(6855), 519–523. doi:10.1038/35097076.
- Lausch, M., Seebach, J., Schnittler, H., & Jessberger, R. (2013). Imbalance of SMC1 and SMC3 Cohesins Causes Specific and Distinct Effects. *PLoS ONE*, 8(6), e65149.
- Leander, B. (2008). Different modes of convergent evolution reflect phylogenetic distances: a reply to Arendt and Reznick. *Trends in Ecology & Evolution*, 23(9), 481–482. doi:10.1016/j.tree.2008.04.012.
- Li, X. (James), Valadez, A. V., Zuo, P., & Nie, Z. (2012). Microfluidic 3D cell culture: potential application for tissue-based bioassays. *Bioanalysis*, 4(12), 1509–1525.
- Li, C., Kato, M., Shiue, L., Shively, J. E., Ares Jr, M., & Lin, R. J. (2006). Cell type and culture condition-dependent alternative splicing in human breast cancer cells revealed by splicing-sensitive microarrays. *Cancer Research*, 66(4), 1990-1999.
- Lieberman, P. (1992). Speech production, syntax comprehension, and cognitive deficits in Parkinson's disease. *Brain and Language*, 43(2), 169–189. doi:10.1016/0093-934x(92)90127-z.

- Liu, J., & Krantz, I. D. (2008). Cohesin and human disease. *Annual review of genomics and human genetics*, 9, 303–320.
- López-Bigas, N., Rabionet, R., de Cid, R., Govea, N., Gasparini, P., Zelante, L., Arbonés, M. L., Estivill, X. (1999). Splice-site mutation in the PDS gene may result in intrafamilial variability for deafness in Pendred syndrome. *Human Mutation*, 14(6), 520–526. doi:10.1002/(sici)1098-1004(199912)14:6<520::aid-humu11>3.0.co;2-k.
- Mabry, K. M., Payne, S. Z., & Anseth, K. S. (2016). Microarray analyses to quantify advantages of 2D and 3D hydrogel culture systems in maintaining the native valvular interstitial cell phenotype. *Biomaterials*, 74, 31–41.
- Maddieson, I. (2011). Phonological Complexity in Linguistic Patterning. In *ICPhS* (pp. 28-34).
- Manczinger, M., Boross, G., Kemény, L., Müller, V., Lenz, T. L., Papp, B., & Pál, C. (2019). Pathogen diversity drives the evolution of generalist MHC-II alleles in human populations. *PLOS Biology*, 17(1), e3000131.
- Mandell, J., Schulze, K., Schlaug, G. (2007) Congenital amusia: An auditory-motor feedback disorder? *Restor Neurol Neurosci* 25:323–334.
- McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., et al. (2016). The Ensembl Variant Effect Predictor. *Genome Biology*, 17(1). doi:10.1186/s13059-016-0974-4.
- McKenna, A., Findlay, G. M., Gagnon, J. A., Horwitz, M. S., Schier, A. F., & Shendure, J. (2016). Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science*, 353(6298), aaf7907.
- Mekel-Bobrov, N., Gilbert, S. L., Evans, P. D., Vallender, E. J., Anderson, J. F., Hudson, R. R., Tishkoff S. A., Lahn, B. T. (2005). Ongoing adaptive evolution of ASPM, a brain size determinant in *Homo sapiens*. *Science* 309(5741), pp. 1720–1722.
- Meher, P. K., Sahu, T. K., & Rao, A. R. (2016). Prediction of donor splice sites using random forest with a new sequence encoding approach. *BioData Mining*, 9(1). doi:10.1186/s13040-016-0086-4.
- Meher, P. K., Sahu, T. K., Rao, A. R., & Wahi, S. D. (2016). Identification of donor splice sites using support vector machine: a computational approach based on positional, compositional and dependency features. *Algorithms for Molecular Biology*, 11(1). doi:10.1186/s13015-016-0078-4.
- de Mendoza, A., Sebé-Pedrós, A., Šestak, M. S., Matejčić, M., Torruella, G., Domazet-Lošo, T., & Ruiz-Trillo, I. (2013). Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular lineages. *Proceedings of the National Academy of Sciences*, 110(50), E4858-E4866.

- Mitchell, M., Krakauer, D. C. (2022). The Debate Over Understanding in AI's Large Language Models. arXiv preprint arXiv:2210.13966.
- Moisik, S. R., & Dediu, D. (2017). Anatomical biasing and clicks: Evidence from biomechanical modeling. *Journal of Language Evolution*, 2(1), 37–51. doi:10.1093/jole/lzx004.
- Moore, D. (2005) Classificação interna da família lingüística Mondé. *Estudos Lingüísticos XXXIV*, pp. 515–520.
- Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K., & Kosakovsky Pond, S. L. (2012). Detecting Individual Sites Subject to Episodic Diversifying Selection. *PLoS Genetics*, 8(7), e1002764. doi:10.1371/journal.pgen.1002764.
- Ng, P. C., & Henikoff, S. (2001). Predicting deleterious amino acid substitutions. *Genome research*, 11(5), 863-874.
- Norris, K. S., Prescott, J. H., Asa-Dorian, P. V., & Perkins, P. (1961). An Experimental Demonstration of Echo-Location Behavior in the Porpoise, *Tursiops truncatus* (Montagu). *Biological Bulletin*, 120(2), 163–176.
- Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, 18(3), 375–402.
- Nunes, K., Maia, M. H. T., dos Santos, E. J. M., dos Santos, S. E. B., Guerreiro, J. F., Petzl-Erler, M. L. et al. (2021). How natural selection shapes genetic differentiation in the MHC region: A case study with Native Americans. *Human Immunology*, 82(7), 523–531.
- Pfeiffer RA, Bauer H, Petersen C. (1977). The Schwartz-Jampel syndrome (myotonia chondrodystrophica). *Helvetica Paediatrica Acta*. Sep;32(3):251-261. PMID: 618079.
- Oliver, M. K., & Piertney, S. B. (2006). Isolation and characterization of a MHC class II DRB locus in the European water vole (*Arvicola terrestris*). *Immunogenetics*, 58(5-6), 390–395. doi:10.1007/s00251-006-0121-6.
- Orsini, L., Vanoverbeke, J., Swillen, I., Mergeay, J., & De Meester, L. (2013). Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Molecular Ecology*, 22(24), 5983–5999.
- Ouattara, K., Lemasson, A. & Zuberbühler, K. (2009a). Campbell's Monkeys Use Affixation to Alter Call Meaning. *PLOS ONE* 4:e7808. doi: 10.1371/journal.pone.0007808.
- Ouattara, K., Lemasson, A. & Zuberbühler, K. (2009b). Campbell's monkeys concatenate vocalizations into context-specific call sequences. *Proc Natl Acad Sci* 106:22026–22031.
- Ovaska, K., & Rand, A. S. (2001). Courtship and Reproductive Behavior of the Frog *Eleutherodactylus diastema* (Anura: Leptodactylidae) in Gamboa, Panama. *Journal of Herpetology*, 35(1), 44. doi:10.2307/1566021.

- Pallares-Ruiz, N., Blanchet, P., Mondain, M., Claustres, M., & Roux, A. F. (2002). A large deletion including most of GJB6 in recessive non syndromic deafness: a digenic effect?. *European Journal of Human Genetics*, 10(1), 72-76.
- Park, H.-J. (2003). Origins and frequencies of SLC26A4 (PDS) mutations in east and south Asians: global implications for the epidemiology of deafness. *Journal of Medical Genetics*, 40(4), 242–248. doi:10.1136/jmg.40.4.242.
- Parreiras-e-Silva, L. T., Vargas-Pinilla, P., Duarte, D. A., Longo, D., Espinoza Pardo, G. V., Dulor Finkler, A., et al. (2017). Functional New World monkey oxytocin forms elicit an altered signaling profile and promotes parental care in rats. *Proceedings of the National Academy of Sciences*, 114(34), 9044–9049.
- Patterson, C. (1982). Morphological characters and homology. In K. A. Joysey and A. E. Friday (eds.). *Problems of Phylogenetic Reconstruction*. Academic Press, London, pp. 21–74.
- Patterson, C. (1988). Homology in classical and molecular biology. *Mol. Biol. Evol.* 5: 603–625.
- Pera, A., Dossena, S., Rodighiero, S., Gandía, M., Bottà, G., Meyer, G., ... & Paulmichl, M. (2008). Functional assessment of allelic variants in the SLC26A4 gene involved in Pendred syndrome and nonsyndromic EVA. *Proceedings of the National Academy of Sciences*, 105(47), 18608-18613.
- Perrakis, A., Sixma, T. K. 2021 AI revolutions in biology. *EMBO reports*.. vol. 22 e54046 p.<https://doi.org/10.15252/embr.202154046>.
- Petit, C., Levilliers, J., & Hardelin, J. P. (2001). Molecular genetics of hearing loss. *Annual review of genetics*, 35(1), 589-645.
- Pierce, G. W., & Griffin, D. R. (1938). Experimental Determination of Supersonic Notes Emitted by Bats. *Journal of Mammalogy*, 19(4), 454–455. doi:10.2307/1374231.
- Pineda, E. T., Nerem, R. M., & Ahsan, T. (2013). Differentiation Patterns of Embryonic Stem Cells in Two- versus Three-Dimensional Culture. *Cells Tissues Organs*, 197(5), 399–410.
- Pischedda, S., Barral-Arca, R., Gómez-Carballa, A., Pardo-Seco, J., Catelli, M. L., Álvarez-Iglesias, V., et al. (2017). Phylogeographic and genome-wide investigations of Vietnam ethnic groups reveal signatures of complex historical demographic movements. *Scientific Reports*, 7(1). doi:10.1038/s41598-017-12813-6.
- Rasmussen, L., Schmidt, M., Henneous, R., Groves, D., & Daves, G. (1982). Asian bull elephants: Flehmen-like responses to extractable components in female elephant estrous urine. *Science*, 217(4555), 159–162. doi:10.1126/science.7089549.
- Reidl, R. (1979). *Order in Living Organisms*. John Wiley, New York.

- Rentzsch, P., Schubach, M., Shendure, J., & Kircher, M. (2021). CADD-Splice—improving genome-wide variant effect prediction using deep learning-derived splice scores. *Genome medicine*, 13(1), 1-12.
- Riggan, L., Hildreth, A. D., Rolot, M., Wong, Y.-Y., Satyadi, W., Sun, R., et al. (2020). CRISPR-Cas9 Ribonucleoprotein-Mediated Genomic Editing in Mature Primary Innate Immune Cells. *Cell Reports*, 31(7), 107651.
- Roll P, Rudolf G, Pereira S, Royer B, Scheffer IE, Massacrier A, Valenti M-P, Roeckel-Trevisiol N, Jamali S, Beclin C et al. (2006) SRPX2 mutations in disorders of language cortex and cognition. *Hum Mol Genet* 15:1195–1207. doi: 10.1093/hmg/ddl035.
- Roll, P., Vernes, S. C., Bruneau, N., Cillario, J., Ponsolle-Lenfant, M., Massacrier, A., Rudolf, G., Khalife, M., Hirsch, E., Fisher, S. E. et al. (2010) Molecular networks implicated in speech-related disorders: FOXP2 regulates the SRPX2/uPAR complex. *Human Molecular Genetics* 19:4848–4860.
- Rosen, S. (2011). *Signals and Systems for Speech and Hearing* (2nd ed.). BRILL. p. 163.
- Rotman-Pikielny, P., Hirschberg, K., Maruvada, P., Suzuki, K., Royaux, I. E., Green, E. D., Kohn, L. D., Lippincott-Schwartz, J., Yen P. M. (2002). Retention of pendrin in the endoplasmic reticulum is a major mechanism for Pendred syndrome. *Human Molecular Genetics*, 11(21), 2625–2633. doi:10.1093/hmg/11.21.2625.
- Ryan, M. J. (2001). *Anuran Communication*. Smithsonian Institution Press, Washington, London, 252p.
- Scott, W. B. (1896). Palaeontology as a morphological discipline. *Biological Letters of the Marine Biology Laboratory, Woods Hole Summer Session, 1895*, pp. 43–61.
- Seikel, J. A., Konstantopoulos, K., Drumright, D. G. (2020). *Neuroanatomy and Neurophysiology for Speech and Hearing Sciences*, 385p. Plural Publishing ISBN13: 978-1-63550-071-4.
- Shen, J., Oza, A. M., Del Castillo, I., Duzkale, H., Matsunaga, T., Pandya, A., ... & Abou Tayoun, A. N. (2019). Consensus interpretation of the p. Met34Thr and p. Val37Ile variants in GJB2 by the ClinGen Hearing Loss Expert Panel. *Genetics in medicine*, 21(11), 2442-2452.
- Smith, J. (1977). *The behavior of communicating. An ethnological approach*. Cambridge, Mass. and London, England: Harvard University Press, 1977. Pp. viii+545. *Language in Society*, 7(03), 435. doi:10.1017/s0047404500005844.
- Sjögren, C., & Nasmyth, K. (2001). Sister chromatid cohesion is required for postreplicative double-strand break repair in *Saccharomyces cerevisiae*. *Current Biology*, 11(12), 991–995.

- Xu, S. X., Ren, W. H., Li, S. Z., Wei, F. W., Zhou, K. Y., & Yang, G. (2009). Sequence Polymorphism and Evolution of Three Cetacean MHC Genes. *Journal of Molecular Evolution*, 69(3), 260–275. doi:10.1007/s00239-009-9272-z
- Spurgin, L. G., & Richardson, D. S. (2010). How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proceedings of the Royal Society B: Biological Sciences*, 277(1684), 979–988. doi:10.1098/rspb.2009.2084
- Storz, J. F. (2016). Causes of molecular convergence and parallelism in protein evolution. *Nature Reviews Genetics*, 17(4), 239–250. doi:10.1038/nrg.2016.11.
- Stringer, C., & Andrews, P. (1988). Genetic and fossil evidence for the origin of modern humans. *Science*, 239(4845), 1263–1268.
- Svensson, R. U., Parker, S. J., Eichner, L. J., Kolar, M. J., Wallace, M., Brun, S. N., ... & Shaw, R. J. (2016). Inhibition of acetyl-CoA carboxylase suppresses fatty acid synthesis and tumor growth of non-small-cell lung cancer in preclinical models. *Nature medicine*, 22(10), 1108-1119.
- Takeda, J. I., Fukami, S., Tamura, A., Shibata, A., & Ohno, K. (2021). IntSplice2: Prediction of the Splicing Effects of Intronic Single-Nucleotide Variants Using LightGBM Modeling. *Frontiers in genetics*, 1232.
- Talarico, L., Marta, S., Rossi, A. R., Crescenzo, S., Petrosino, G., Martinoli, M., & Tancioni, L. (2021). Balancing selection, genetic drift, and human-mediated introgression interplay to shape MHC (functional) diversity in Mediterranean brown trout. *Ecology and Evolution*.
- Thanapattheerakul, T., Engchuan, W., & Chan, J. H. (2020). Predicting the effect of variants on splicing using Convolutional Neural Networks. *PeerJ*, 8, e9470.
- Theocharopoulou, G., Vlamos, P. (2020). Modeling the Critical Activation of Chaperone Machinery in Protein Folding. In: Vlamos, P. (eds) *GeNeDis 2018. Advances in Experimental Medicine and Biology*, vol 1194. Springer, Cham.
- Townsend, S. W., Deschner, T., & Zuberbühler, K. (2008). Female Chimpanzees Use Copulation Calls Flexibly to Prevent Social Competition. *PLoS ONE*, 3(6), e2431. doi:10.1371/journal.pone.0002431.
- Van Hauwe, P., Everett, L. A., Coucke, P., Scott, D. A., Kraft, M. L., Ris-Stalpers, C., Bolder, C., Otten, B., de Vijlder, J. J. M., Dietrich, N. L., Ramesh, A., Srisailapathy, S. C. R., Parving, A., Cremers, C. W. R. J., Willems, P. J., Smith, R. J. H., Green, E. D., Van Camp, G. (1998) Two Frequent Missense Mutations in Pendred Syndrome, *Human Molecular Genetics*, Volume 7, Issue 7, Pages 1099–1104.
- Vargas-Pinilla, P., Paixão-Côrtes, V. R., Paré, P., Tovo-Rodrigues, L., Vieira, C. M. de A. G., Xavier, A., et al. (2014). Evolutionary pattern in the OXT-OXTR system in primates:

- Coevolution and positive selection footprints. *Proceedings of the National Academy of Sciences*, 112(1), 88–93.
- Velichko, A. A., Kurenkova, E. I., & Dolukhanov, P. M. (2009). Human socio-economic adaptation to environment in Late Palaeolithic, Mesolithic and Neolithic Eastern Europe. *Quaternary International*, 203(1-2), 1–9.
- Venkatesh, M. D., Moorchung, N., & Puri, B. (2015). Genetics of non syndromic hearing loss. *medical journal armed forces india*, 71(4), 363-368.
- Vergne, A. L., Pritz, M. B., & Mathevon, N. (2009). Acoustic communication in crocodylians: from behaviour to brain. *Biological Reviews*, 84(3), 391–411. doi:10.1111/j.1469-185x.2009.00079.x.
- Viscardi, L. H., Tovo-Rodrigues, L., Paré, P., Fagundes, N. J. R., Salzano, F. M., Paixão-Côrtes, V. R., Bau, C. H. D., & Bortolini, M. C. (2017). FOXP in Tetrapoda: Intrinsically Disordered Regions, Short Linear Motifs and their evolutionary significance. *Genetics and molecular biology*, 40(1), 181–190.
- Walkowiak, W. (2006). “Call production and neural basis of vocalization,” in *Hearing and Sound Communication in Amphibians*, edited by P. M. Narins, A. S. Feng, R. R. Fay, and A. N. Popper (Springer, New York), pp. 87–112.
- Wang, W. H., Liu, Y. F., Su, C. C., Su, M. C., Li, S. Y., & Yang, J. J. (2011). A novel missense mutation in the connexin30 causes nonsyndromic hearing loss. *PloS one*, 6(6), e21473.
- Watanabe, K., Taskesen, E., van Bochoven, A., & Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. *Nature Communications*, 8(1). doi:10.1038/s41467-017-01261-5.
- Weir, B. S. (2005). Measures of human population structure show heterogeneity among genomic regions. *Genome Research*, 15(11), 1468–1476.
- Wells, K. D. (2010). *The ecology and behavior of amphibians*. University of Chicago Press, In.
- Wettschureck, N., Moers, A., Hamalainen, T., Lemberger, T., Schutz, G., & Offermanns, S. (2004). Heterotrimeric G Proteins of the Gq/11 Family Are Crucial for the Induction of Maternal Behavior in Mice. *Molecular and Cellular Biology*, 24(18), 8048–8054. doi:10.1128/mcb.24.18.8048-8054.2004.
- Wilson, E.O. (1975) *Communication: Origins and Evolution*. In: *Sociobiology: The New Synthesis*, Belknap Press of Harvard University Press, Cambridge, US, 225 p.
- Wong, P. C. M., Chandrasekaran, B., & Zheng, J. (2012). The Derived Allele of ASPM Is Associated with Lexical Tone Perception. *PLoS ONE*, 7(4), e34243. doi:10.1371/journal.pone.0034243.

- Wong, P. C. M., Kang, X., Wong, K. H. Y., So, H.-C., Choy, K. W., & Geng, X. (2020). ASPM-lexical tone association in speakers of a tone language: Direct evidence for the genetic-biasing hypothesis of language evolution. *Science Advances*, 6(22), eaba5090. doi:10.1126/sciadv.aba5090.
- Yang, P.-C., & Mahmood, T. (2012). Western blot: Technique, theory, and trouble shooting. *North American Journal of Medical Sciences*, 4(9), 429.
- Yang, Z. (2007). PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution*, 24(8), 1586–1591. doi:10.1093/molbev/msm088.
- Yang, E.-J., Kelley, D. B. (2008). Hormones and the regulation of vocal patterns in amphibians: *Xenopus laevis* vocalizations as a model system. In: *Hormones, Brain, and Behavior*, second edition (eds. Pfaff D, Arnold A, Etgen A, Fahrbach S, Rubin R). Academic Press, New York, pp. 693 – 706.
- Yang, Z. (1998). Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Molecular biology and evolution*, 15(5), 568-573.
- Zhang, H. (2018) *Second Language Acquisition of Mandarin Chinese Tones*. Brill | Rodopi xii, 146 pp.
- Zhou, J., Troyanskaya, O. (2015) Predicting effects of noncoding variants with deep learning–based sequence model. *Nature Methods* 12, 931–934.
- Wendt, K. S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., et al. (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature*, 451(7180), 796–801. doi:10.1038/nature06634.
- Whitlow W.L. Au. (2009) *Echolocation*, Editor(s): William F. Perrin, Bernd Würsig, J.G.M. Thewissen, *Encyclopedia of Marine Mammals (Second Edition)*, Academic Press, Pages 348-357,
- Zhu, Z., Li, Q. V., Lee, K., Rosen, B. P., González, F., Soh, C.-L., & Huangfu, D. (2016). Genome Editing of Lineage Determinants in Human Pluripotent Stem Cells Reveals Mechanisms of Pancreatic Development and Diabetes. *Cell Stem Cell*, 18(6), 755–768.
- Zuallaert, J., Godin, F., Kim, M., Soete, A., Saeys, Y., & De Neve, W. (2018). SpliceRover: interpretable convolutional neural networks for improved splice site prediction. *Bioinformatics*. doi:10.1093/bioinformatics/bty497.