

Associação entre Energia Livre de Sequências do Gene *Axin2* e Tamanho Relativo de Dentes Posteriores de Mamíferos Placentários

Palavras-Chave: AXIN2, DENTIÇÃO DE MAMÍFEROS, G-QUADRUPLEX

Autores:

Sara Garcia Azevedo - Faculdade de Odontologia de Piracicaba, Universidade Estadual de

Campinas. Contato: s165456@dac.unicamp.br

Prof. Dr. Sergio Roberto Peres Line - Faculdade de Odontologia de Piracicaba, Universidade

Estadual de Campinas. Contato: serglin@unicamp.br

INTRODUÇÃO:

O sistema mastigatório é uma das partes mais complexas dos mamíferos, cuja função depende de ações coordenadas de várias estruturas como a articulação temporomandibular, musculatura da mastigação e língua e o padrão da dentição com variações na forma, número de dentes e regularização do número de dentições. A dentição dos mamíferos é composta por quatro diferentes tipos morfológicos: incisivos, caninos, pré-molares e molares. Estes dois últimos possuem uma forma mais complexa, associada à habilidade de mastigação, característica que produziu um significativo ganho na eficiência digestiva, que é considerada necessária para o alto metabolismo associado à endotermia (Berkovitz et al., 1995). O controle de número, forma e estrutura tecidual dos dentes é feito durante a odontogênese, um processo complexo que envolve muitas interações moleculares. Mais de 300 genes envolvidos na odontogênese já foram descritos, e mutações em vários deles, como *PAX9*, *MSX1* e *AXIN2*, foram associadas à agenesia dentária em humanos. O gene *Axin2* (axis inhibitor 2) tem um papel importante na morfogênese da região craniofacial. É expresso no mesênquima ao redor do epitélio, nos nós de esmalte, na papila dentária e na camada de odontoblastos e pré-odontoblastos. Variações polimórficas nesse gene podem ocasionar perda permanente de molares, pré-molares, incisivos inferiores e incisivos laterais superiores. Nos humanos, o *AXIN2* está localizado no braço longo do cromossomo 17, na região 17q24.1, e possui 11 éxons, dos quais 10 são codificantes (Hloušková et al., 2017).

Quadruplexes (G4) são estruturas de DNA ou RNA que podem surgir sob condições fisiológicas. São constituídos por quatro bases de guanina unidas por ligações de hidrogênio e estabilizados por um cátion alcalino, formando um quadrado. São encontrados em maior quantidade nas regiões promotoras e introns de genes, o que sugere seu papel na regulação da expressão gênica. Além dessa função, várias outras já foram propostas, como manutenção da desmetilação das ilhas CpG e demarcação da terminação da transcrição. Nos íntrons, na fita não codificante, podem ter papel na regulação da transcrição, especialmente quando localizados na vizinhança de sítios de iniciação, ou no processamento do RNA, incluindo splicing alternativo (Robinson et al., 2021; Guiblet et al., 2021). O gene *Axin2* possui sequências de G4 que podem estar relacionadas a variações do fenótipo dental de mamíferos. Assim, o objetivo desse trabalho é mapear regiões no gene *Axin2* que estão associadas à

diversificação da dentição posterior em mamíferos, por meio da comparação da energia livre do mRNA deste gene.

METODOLOGIA:

Foram incluídas 34 espécies de mamíferos placentários (como mostrado na **Tabela 2**), pertencentes a 9 ordens (Primata, Artiodactyla, Perissodactyla, Soricomorpha, Scandentia, Tubulidentata, Rodentia, Chiroptera e Afrosoricida), cujas medidas mesiodistais (md) e bucolinguais (bl) dos últimos pré-molares (P4), primeiros (M1), segundos (M2) e terceiros molares (M3) foram retiradas do estudo de Jara-Espejo et al. (2020) - a distância mesiodistal foi obtida por meio do desenho de uma linha reta entre os pontos de contato distal e mesial, na face oclusal, e a bucolingual, por uma linha reta que abrange desde a face vestibular até a lingual.

As sequências do gene *Axin2* das 34 espécies foram obtidas do Ensembl (<https://www.ensembl.org/index.html>) e do GenBank (<https://www.ncbi.nlm.nih.gov/>). Utilizamos a ferramenta BLAT do UCSC Genome Browser Gateway (<http://genome.ucsc.edu/>) para adquirir a sequência de cada espécie em coincidência aos 2 primeiros éxons do gene humano, que corresponde à posição 65557806 até 65561768 bp da fita reversa do cromossomo 17. As sequências foram, então, alinhadas com o software MAFFT (Kato e Standley 2013) e escaneadas de 5' para 3' utilizando janelas de 30, 35, 40, 45, 50, 55 e 60 bp, com diferenças de 1 bp. As sequências de cada janela foram analisadas usando o software RNAfold para calcular a energia livre (Mfe) de quadruplex (G4) e dentro de cada uma delas foi escolhido o fragmento de maior estabilidade de pareamento da fita de RNA para cada espécie. Estes valores foram usados como variável genética nos testes de correlação.

As análises estatísticas foram realizadas utilizando-se a versão 3.4.4 do software R (<https://www.r-project.org/>, R Core Team, 2019). Correlações de Spearman foram feitas com os dados de energia livre e proporções entre dentes posteriores. As razões escolhidas foram: M2md/M1md, M3md/M1md, M3md/M2md, P4md/M1md, P4md/M2md, P4md/M3md, M2bl/M1bl, M3bl/M1bl, M3bl/M2bl, P4bl/M1bl, P4bl/M2bl e P4bl/M3bl. Foram selecionadas regiões onde o valor absoluto da correlação foi maior do que 0.45 e valor de $p < 0.01$.

As sequências de RNA mitocondrial de todos os mamíferos placentários analisados foram obtidas do banco de dados Nucleotide (<https://www.ncbi.nlm.nih.gov/nucleotide/>) e alinhadas usando o algoritmo de alinhamento MAFFT. Por fim, obtivemos um objeto de classe filo que foi usado para computar contrastes filogeneticamente independentes (pic) usando o método descrito por Felsenstein (1985). Para isso, usamos a função `pic()` do pacote R "phytools" (Revell 2012).

Dentre as sequências selecionadas, foram escolhidas aquelas que se repetiram em pelo menos 4 *sliding windows* dentro da mesma razão, e dentre estas, as que apresentaram o maior valor de rho e/ou número de espécies com G4.

RESULTADOS E DISCUSSÃO:

Quatro regiões localizadas no intron 1 foram identificadas em associação com as razões que envolviam distâncias mesiodistais, as duas primeiras estão no início do intron. As correlações obtidas com as medidas bucolinguais foram menores e menos significativas do que as obtidas com as mesiodistais, mas uma sequência localizada no exon 2 se destacou. Os resultados estão resumidos na **Tabela 1**.

	Razão	Chr17	Posição	G4	rho C1	rho C2
Seq 1	P4md/M3md	65561335-65561368 bp	Início do intron 1	16	0.66	0.54

Seq 2	P4md/M3md	65561081-65561091 bp	Início do intron 1	21	0.54	0.49
Seq 3	P4md/M1md	65559580-65559597 bp	Intron 1	22	0.53	0.52
Seq 4	M3md/M2md	65559515-65559546 bp	Intron 1	0	-0.75	-0.54
Seq 5	P4bl/M1bl	65558079-65558095 bp	Exon 2	0	-0.70	-0.50

Tabela 1. As 5 sequências selecionadas. Chr17: localização na fita reversa do cromossomo 17. G4: quantidade de espécies com G4, dentre as 34. Rho C1: correlação de Spearman com correção filogenética. Rho C2: correlação de Spearman sem correção filogenética.

Das 5 sequências, a primeira foi a que apresentou maior valor de energia livre (Mfe) de quadruplex na espécie humana (-25.59 kcal/mol); a **Figura 1** mostra a conformação do RNA para esse fragmento no humano. A **Tabela 2** apresenta o fragmento de maior energia das 34 espécies para a sequência 1.

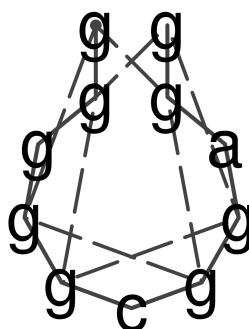


Figura 1. Sequência 1 - conformação do RNA na espécie *Homo sapiens*, formando sequência G4.

Para as razões mesiodistais que envolviam somente molares, todas as correlações foram negativas e as que envolviam pré-molares, todas foram positivas. É sabido que quanto menor a energia livre, mais estável é a molécula de RNA. Desta maneira, G4 mais estáveis (mais negativos) parecem estar relacionados com um maior comprimento dos molares posteriores em relação aos anteriores, bem como um menor comprimento do último pré-molar em comparação aos molares. Esses resultados indicam que quanto maior a dimensão dos molares, menor a dos pré-molares e vice-versa. Isso é observado, por exemplo, nas dentições do cão e do gato, em que a área dos maxilares ocupada pelos pré-molares é maior que pelos molares, sendo que, nos primatas, a área ocupada pelos molares é maior que a dos pré-molares (Berkovitz et al., 1992). Já para as razões bucolinguais não houve um padrão como para as mesiodistais. As razões que envolviam somente molares foram todas negativas, mas as que envolviam pré-molares variaram, algumas sendo positivas e outras negativas.

Em resumo, as correlações entre a energia livre das sequências e as razões foram mais significativas para as medidas mesiodistais. Resultados similares foram encontrados no estudo de Jara-Espejo et al. (2020), onde os autores explicam que a medida mesiodistal mandibular posterior dos primatas é influenciada primariamente por fatores genéticos, enquanto que a medida bucolingual parece ter uma maior influência de fatores externos, como tamanho do corpo, sexo e limitação imposta pelo maxilar. Deste modo, a energia livre pode recapitular o potencial efeito do gene *Axin2* nas medidas bucolinguais, mas foi parcialmente contido por fatores externos (Jara-Espejo et al., 2020).

No teste aplicado neste estudo, os valores mais significativos de correlações foram encontrados nas razões P4mdM1md e P4mdM3md, diferentemente dos resultados obtidos no estudo de Jara-Espejo et al. (2020) sobre o gene *Pax9*, em que os valores de maior significância foram para

M2md/M1md. Isso sugere que, de maneira distinta ao gene *Pax9*, o gene *Axin2* tem uma participação importante na formação de pré-molares em mamíferos.

<i>Espécies</i>	<i>Seq 1</i>	<i>Mfe</i>
<i>Nomascus leucogenys</i>	CCGAGCCGCGTCGTGGTCGCATGGGGCGGGAGG	-16
<i>Gorilla gorilla</i>	GGCGTGGGCGCGGGGGGCGGGAGGGGGCCCGGCC	-25.59
<i>Homo sapiens</i>	GGCGTGGGCGCGGGGGGCGGGAGGGGGCCCGGCC	-25.59
<i>Pan troglodytes</i>	GGCGTGGGCGCGGGGGGCGGGAGGGGGCCCGGCC	-25.59
<i>Pan paniscus</i>	GGCGTGGGCGCGGGGGGCGGGAGGGGGCCCGGCC	-25.59
<i>Pongo abelii</i>	GCCGCGGCGTGGGCGCGGGGACGGGAGGGGGCCCGGC	-29.8
<i>Macaca fascicularis</i>	-	0
<i>Macaca mulatta</i>	GCTGGGGGCGGGAGGGGGCCCGGC	-18.89
<i>Chlorocebus sabaeus</i>	GCTGGGGGCGGGAGGGGGCCCGGC	-18.89
<i>Papio anubis</i>	GCTGGGGGCGGGAGGGGGCCCGGC	-18.89
<i>Saimiri sciureus</i>	-	0
<i>Bos taurus</i>	CCGAGGGCGCGCCCGGCGGCGTCGCGGGCGCCGGGGGCGGGAGGGGCTCCGG	-34.09
<i>Capra hircus</i>	GGCCGCGTCGCGGGCGCCGGGGACGGGAGGGGCTCCGGCC	-25.3
<i>Ovis aries</i>	GCGAGGGCGCGGCCCGGCCGCGCGCGGGC	-18.5
<i>Equus caballus</i>	CCGGGGGCGAGAGGGGGTCCGCCG	-12.7
<i>Ceratotherium simum</i>	GGTCGCGCCGAGTCGCGTCGCGGGCGCCGGGGGCGAGAGGGGGTCCGGCC	-23
<i>Sus scrofa</i>	GGGCTGGAAGGGGTCCAGCCC	-14.5
<i>Vicugna pacos</i>	-	0
<i>Camelus bactrianus</i>	-	0
<i>Sorex araneus</i>	-	0
<i>Tupaia belangeri</i>	GGTCGCGCCGAGCCTCCTAGCGGGCGTCCGGGACGGGAGAGGGTCCGGCC	-20.7
<i>Orycteropus afer</i>	-	0
<i>Spermophilus tridecemlineatus</i>	CGAGGTCGTGTCGGGTTGTGTAGTGGGCGCGGGCGACAGGAGGGGGCCAGCG	-15.9
<i>Tarsius syrichta</i>	CCGTCGCGGGCTCCGGGGGCGGAAGGGGCTCAATGG	-18.69
<i>Eptesicus fuscus</i>	GGGCGCGCCAAGCCGCATCGCGTGC CGGGGGGCGGGAGAGGGTCC	-20.2
<i>Cavia porcellus</i>	CGAGGTCGTGCCCCACCCCGTCGTGGGCGTAGGGGGGGCGCAAAGGGCCCG	-23.6
<i>Chinchilla lanigera</i>	-	0
<i>Heterocephalus glaber</i>	-	0
<i>Mus musculus</i>	CGGGGTGCCCCGCGCGGGGCCGGGAGAGGCGCGGGCGCGGGAAGGGGCCCCG	-31.9
<i>Rattus norvegicus</i>	CGGTTGGCAGCGGGGCGCGGGGCCGCGGGAGGGGGCGACCG	-25.25
<i>Microtus ochrogaster</i>	GGGCCGCGGGAGGGGGCGGCCGGGAGCAGGCC	-32.2
<i>Mesocricetus auratus</i>	GCGGCTGGCAGCGGGGTACGGGGCCACGGAAGGGGGCAGGAGGGTGCAAGTCGT	-23.99
<i>Echinops telfairi</i>	CGGCTGGCAGCGGGGTACGGGGCTACGGGAGGGGACGCGAGGGTGCAAGTCG	-25.45
<i>Jaculus jaculus</i>	GGCAGCGACGCGGGCGCGGGCCGCGGGAAGGGGCCCGCC	-20.65

Tabela 2. Sequência 1, razão P4md/M3md - energia mínima livre (Mfe, kcal/mol) das 34 espécies incluídas no estudo.

As sequências que apresentaram maior valor de rho e, simultaneamente, maior número de espécies com G4 foram encontrados no intron 1, especialmente no seu início (**Tabela 1**). Quadruplexes apresentam papel de relevância no funcionamento das células e, quando presentes em introns, podem afetar o splicing e a expressão de genes. Quase metade dos genes humanos têm regiões ricas em guanina na extremidade 5' do primeiro intron e já foi mostrado que essas sequências no começo do

intron 1 do gene *Pax9* formam quadruplexes estáveis na estrutura de RNA e têm um papel chave na eficiência de splicing desse gene (Ribeiro et. al, 2015). Nossos resultados se somam a esses na confirmação da importância das regiões ricas em guanina com conformação de G4 na expressão gênica.

CONCLUSÕES:

Quatro regiões localizadas no intron 1 do gene *Axin2* foram identificadas em associação com as razões que envolviam distâncias mesiodistais e uma região no exon 2 foi associada às distâncias bucolinguais. Esses resultados sugerem que variações nas sequências do *Axin2* contribuem para um fenótipo dental diversificado nas espécies de mamíferos.

BIBLIOGRAFIA

BERKOVITZ, Barry K B; HOLLAND, G R; MOXHAM, B J. **Oral Anatomy, Histology and Embryology**. Wolfe Publishing Ltd., 1995.

FELSENSTEIN, J; Phylogenies and the Comparative Method. *American Naturalist*, v. 125, p. 1-15. Disponível em: <https://www.journals.uchicago.edu/doi/10.1086/284325>.

GUIBLET, Wilfried M; DEGIORGIO, Michael; CHENG, Xiaoheng; CHIAROMONTE, Francesca; ECKERT, Kristin A; HUANG, Yi-Fei; MAKOVA, Kateryna. **Selection and Thermostability Sugget G-Quadruplexes are Novel Functional Elements of The Human Genome**. *Genome Research*, v. 31, p. 1136-1149, 2021. Disponível em: <https://genome.cshlp.org/content/31/7/1136>.

HLOUSKOVA, Alena; BIELIK, Peter; BONCZEK, Ondrej; BALCAR, Vladimir J; SEREY, Omar. **Mutation in AXIN2 as a Risk Factor for Tooth Agenesis and Cancer: a Review**. *Neuroendocrinology Letters*, v. 38, p. 131-137, 2017.

JARA-ESPEJO, Manuel; HAWKINS, Melissa T R; FOGALI, Giovani B; LINE, Sergio R P. **Folding Stability of Pax9 Intronic G-Quadruplex Correlates with Relative Molar Size in Eutherians**. *Molecular Biology and Evolution*, v. 38, p. 1860-1873, 2020. Disponível em <https://doi.org/10.1093/molbev/msaa331>.

KATOH, Kazutaka; STANDLEY, Daron M. **MAFFT multiple sequence alignment software version 7: improvements in performance and usability**. *Mol Biol Evol.*, v. 30, p. 772-780, 2013. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3603318/>.

REVELL, Liam J. **Phytools: an R package for phylogenetic comparative biology (and other things)**. *Methods Ecol Evol.*, v. 3, p. 217–223, 2012.

RIBEIRO, Mariana M; TEIXEIRA, Gleidson S; MARTINS, Luciane; MARQUES, Marcelo R; DE SOUZA, Ana Paula; LINE, Sergio R P. **G-Quadruplex Formation Enhances Splicing Efficiency of PAX9 intron 1**. *Hum Genet*, v. 134, p. 37-44, 2015. Disponível em: <https://link.springer.com/article/10.1007/s00439-014-1485-6>.

ROBINSON, Jenna; RAGUSEO, Federica; NUCCIO, Sabrina P; LIANO, Denise; DI ANTONIO, Marco. **DNA G-Quadruplex Structures: More than Simple Roadblocks to Transcription**. *Nucleic Acids Research*, v. 49, p. 8419-8431, 2021. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8421137/>.