

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

INSTITUTO DE BIOCÊNCIAS - DEPARTAMENTO DE GENÉTICA

**INVESTIGAÇÃO DA VARIABILIDADE MOLECULAR DO ENHANCER *HACNS1*
EM DIFERENTES POPULAÇÕES HUMANAS**

Aluno: Álvaro Silveira

Orientadora: Prof. Dr^a. Maria Cátira Bortolini

Co-orientadora: Dr^a. Tábita Hünemeier

Trabalho submetido à Comissão de Graduação em Ciências Biológicas como requisito parcial para obtenção do título de Bacharel em Ciências Biológicas, ênfase Molecular, Celular e Funcional.

Porto Alegre, novembro de 2011.

Financiamento: Conselho Nacional de Pesquisa (CNPq) e Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

Este trabalho constitui-se da versão integral¹do artigo,

**POPULATION DATA SUPPORT THE ADAPTIVE NATURE OF HACNS1
SAPIENS/NEANDERTAL-CHIMPANZEE DIFFERENCES IN A LIMB EXPRESSION DOMAIN.**

Hünemeier, T; Ruiz-Linares, Andres; **SILVEIRA, Álvaro**; Paixão-Côrtes, Vanessa Rodrigues; Salzano, Francisco M.; Bortolini, Maria Cátira.

American Journal of Physical Anthropology, p. 478-481, 2010.

¹ Conforme orientação da Comissão de Graduação em Ciências Biológicas da Universidade Federal do Rio Grande do Sul, o artigo, originalmente publicado na língua inglesa, foi traduzido para a língua portuguesa.

SUMÁRIO

1. Resumo.....	5
2. Introdução.....	6
3. Material e Métodos.....	9
4. Resultados e Discussão.....	12
5. Referências Bibliográficas.....	15
6. Tabelas e Figuras.....	18
7. Artigo Original.....	19

1. RESUMO

O *enhancer HACNS1* compreende uma região de 546 pares de base altamente conservada na maioria dos vertebrados terrestres. Em humanos esta região acumula 16 mutações específicas surgidas após a separação das linhagens humano-chimpanzé. Trabalhos recentes têm discutido se este acúmulo de mutações ocorreu por seleção positiva ou por uma conversão gênica enviesada. Neste trabalho, estudamos 83 ameríndios do sul, 11 esquimós e seis africanos subsaarianos, e nestes não foram encontrados variações nesta região do DNA. Este resultado favorece a idéia de que tenha ocorrido uma seleção positiva no passado e de uma seleção conservativa no presente, como o esperado para uma região com influências nos traços como o polegar opositor, destreza manual e locomoção bípede vista na linhagem humana.

2. INTRODUÇÃO

Como seres humanos, nós temos interesses especiais em identificar as modificações genéticas responsáveis por características específicas que nos distinguiram dos outros grandes primatas. Diferenças morfológicas podem ocorrer devido a poucas mutações de grande efeito em genes regulatórios ou estruturais chaves, assim como em sequências não codificantes com papel na regulação gênica (*major gene effect hypothesis*; Nei 1987, 2007). Prabhakar e cols. (2008) descreveram uma seqüência de 546 pares de base (pb) que atua como um *enhancer* da expressão de genes, o *HACNS1*, sendo esta altamente conservada em todos os genomas de vertebrados terrestres. No entanto, em humanos este *enhancer* acumula 16 mudanças específicas surgidas após a separação entre a linhagem humanos e a dos chimpanzés.

Treze destas 16 mutações são encontrados num segmento funcional de 81 pb desta região de 546 pb. Esses autores mostraram, usando um teste estatístico baseado em *log* de verossimilhança, que estas mudanças eram altamente inesperadas pelo acaso ($p=9.2 \times 10^{-12}$). Prabhakar e cols. (2008) também demonstraram que essas alterações de nucleotídeos promovem uma forte expressão de genes envolvidos nos membros do corpo, quando comparado com o elemento ortólogo encontrado em chimpanzés, tendo isto um provável impacto em nossa história evolutiva. Duret and Galtier (2009), por outro lado, argumentaram que o padrão que substituições encontrado no *HACNS1* poderia ser explicado por um processo neutro chamado de conversão gênica enviesada (*BGC – Biased Gene Conversion*) associado com eventos de recombinação aleatória, que favorece a fixação de transições AT-GC, e que também pode levar a fixação de mutações fracamente deletérias em elementos funcionais.

Prabhakar *e cols.* (2009), depois de uma avaliação ampla do genoma, demonstraram que o *HACNS1* é incomum com relação aos eventos de recombinação, e que apesar do excesso de mudanças AT-GC encontradas neste *enhancer*, mudanças deletérias, ou neutras, não seriam esperadas por fortalecer os padrões de expressão ancestrais introduzindo um novo e robusto domínio de expressão na linhagem humana. Estes autores também evidenciaram a acelerada evolução no *HACNS1* devido a um mecanismo comum: a sinergia entre a *BGC* e a seleção positiva produzindo um *cluster* de substituições AT-GC neste sítio funcional (Prabhakar *e cols.*, 2009). Por fim, Katzman *e cols.* (2010) sugeriram que a *BGC* poderia ser um importante fator guiando a evolução do *HACNS1*, mas indicando que nenhum dado estava disponível com o qual fosse possível distinguir se as mutações humano-específicas refletiriam um processo que foi essencialmente uma batalha para evitar o ataque da *BGC* e deste modo continuar num contexto adaptativo, ou se as mutações pressionaram este elemento levando-o a uma nova configuração que permitiu a ação da seleção positiva, num cenário adaptativo para os humanos. Até o momento, nenhuma discussão havia focado e considerado dados populacionais.

Nativo-americanos possuem um padrão bem conhecido de variação molecular, com diferenças marcantes em relação a outras populações humanas, quando sistemas genéticos neutros são considerados (Wang *e cols.*, 2007). Considerando que análises com populações humanas revelam que os padrões de variação sob seleção natural são diferentes daqueles esperados sob neutralidade (Meyer *e cols.* 2006), este estudo foi desenhado para avaliar as duas hipóteses sobre a evolução do *HACNS1*.

Neste trabalho são apresentados dados do seqüenciamento de uma região de 710 pb, que inclui o *enhancer HACNS1* e regiões adjacentes, em populações sul-americanas amplamente distribuídas e com distintas histórias demográficas. Para uma maior cobertura étnica e continental, esquimós siberianos, europeus, africanos (bantus e não-bantus) e brasileiros miscigenados foram incluídos no estudo. Adicionalmente foi realizada uma busca nos bancos de genomas humanos completos e no genoma do Neandertal.

3. MATERIAL E MÉTODOS

Amostras de DNA

Um total de 194 amostras foi analisado, incluindo 83 nativo-americanos de 12 populações distintas (Apalai, Arara, Galibi, Kuben-Kran-Keng, Mundurucu, Karitiana, Xavante, Ache, Guarani, Ticuna, Wayuu e Zenu), amplamente distribuídas por todo continente. Onze esquimós siberianos, 37 africanos sub-Saarianos, 35 europeus, e 28 brasileiros miscigenados também foram incluídos como amostras não-ameríndias.

Os 37 africanos foram obtidos de indivíduos bantus e não-bantus que vivem na República Democrática do Congo, Camarões e Costa do Marfim, enquanto os europeus são espanhóis. Informações adicionais sobre as populações africanas e européia podem ser encontradas em Silva *e cols.* (2006) e Bortolini *e cols.* (2004), respectivamente. Os esquimós da amostra falam a língua Yupic e vivem em uma comunidade no extremo leste da Sibéria. Vinte e oito gaúchos do sul do Brasil também foram amostrados. Os mesmos podem ser caracterizados como miscigenados a partir do encontro de ameríndios, africanos e europeus (Marrero *e cols.*, 2007).

A aprovação ética para este estudo foi provida pela Comissão Nacional de Ética na Pesquisa (CONEP – Resoluções número 123/1998 e 133/2002), bem como pelos comitês de ética dos países onde as amostras não-brasileiras foram coletadas.

Genotipagem e Sequenciamento

O fragmento de 710 pb incluindo o *HACNS1* foi amplificado com primers específicos (F-5'TCTCGTCGGCATTACTCATCGTCA3'; R- 5'CAAATGGAGGCTTTTCTGCA3'), desenhados especificamente para este estudo. As reações de *PCR* continham 20 to 50

ng de DNA genômico, 5µL de Hot Start Master Mix Kit (Qiagen, Hilden, Germany), 4 µL água e 0.5 µL (10 pmol/ µL) de cada par de primer, expostos a 94°C por 15min, 33 ciclos a 94°C por 30s, 57°C por 40s, 72°C por 30s, seguido de 10 min a 72°C. O produto de PCR foi verificado em eletroforese em gel de agarose 1.0% corado com brometo de etídio, e depois purificado com o *kit* caseiro Microclean. Os fragmentos de DNA foram sequenciados em um sequenciador ABI 3730xl e analisados com os programas BioEdit v7.0.9 e Sequence Scanner v1.0.

Testes Evolutivos

O programa *EvoNC* (Wong and Nielsen, 2004) foi utilizado para testar se o padrão de variação no “*enhancer*” *HACNS1* poderia ser explicado por um modelo neutro ou não. As seqüências de seis espécies de primatas não-humanos (*Otolemur garnettii*, *Callithrix jacchus*, *Macaca mulatta*, *Pongo pygmaeus*, *Gorilla gorilla*, *Pan troglodytes*), bem como de *Homo sapiens*, *Homo neanderthalensis* e um hominídeo sem classificação definida encontrado na caverna de Denisova foram obtidas por buscas nos bancos de genomas *Ensembl* (<http://www.ensembl.org>) e UCSC (<http://genome.ucsc.edu>).

As seqüências analisadas incluíram o “*enhancer*” *HACNS1* (659 bp) mais o *exon* 1 do gene *AGP1/CENTG2* (163 bp), que codifica a proteína ativadora da GTPase e está localizado imediatamente após o “*enhancer*”. Estes alinhamentos em conjunto com uma árvore filogenética dos primatas foram então utilizados para identificar o padrão evolutivo do “*enhancer*” *HACNS1*.

O programa *EvoNC* compara as taxas de substituições em regiões não-codificadoras em relação às taxas de substituições sinônimas nas regiões codificadoras.

Com base nisto, o parâmetro ζ (zeta) foi então calculado. Três modelos foram testados: Um modelo nulo é comparado com dois modelos alternativos com duas e três categorias de ζ , respectivamente. O modelo neutro assume dois conjuntos de sítios com $\zeta_0 < 1$ e $\zeta_1 = 1$. O modelo de duas categorias também assume dois conjuntos com $\zeta_0 \leq 1$ e $\zeta_1 \geq 1$; e o modelo de três categorias assume três conjuntos: $\zeta_0 < 1$, $\zeta_1 = 1$ e $\zeta_2 > 1$. Basicamente, $\zeta = 1$ quando os sítios na região não-codificadora estão evoluindo de maneira neutra, $\zeta > 1$ quando há seleção natural, e $\zeta < 1$ quando a seleção negativa está presente. Também foi feita a análise da probabilidade a posteriori (Teorema de Bayes) para classificar cada sítio individual como pertencente a uma das classes de ζ mencionadas acima (Wong and Nielsen, 2004).

4. RESULTADOS E DISCUSSÃO

As 13 substituições agrupadas na região de 81 pb identificadas por Prabhakar *e cols.* (2008), assim como três outras que nos distinguem dos chimpanzés, foram observadas em todos os indivíduos estudados, sugerindo que elas estão fixadas na nossa espécie. Uma mutação adicional (cromossomo 2: 236774267; Figura 1) não reportada por Prabhakar *e cols.* (2008), dado que eles seqüenciaram somente 546bp (236773657 to 236774202) enquanto nós seqüenciamos 710bp (236773577 to 236774286), também foi encontrada distinguindo as duas espécies (ver Figura 1).

Não foram encontradas variações intra ou inter-populacionais, sugerindo que a região é conservada. Vale a pena mencionar que uma amostra de 4.000 anos do recém seqüenciando Palaeo-eskimo, bem como um Khoisan, um Bantu, dois Yorubas, um chinês Han, um coreano e cinco europeus ou europeu-derivados, que possuem seu genoma completo seqüenciado, também apresentam todas as substituições humano-específicas (Rasmussen *e cols.*, 2010; Schuster *e cols.*, 2010; <http://www.ensembl.org/>; <http://www.galaxy.org/>).

A comparação com os recém publicados genomas do Neandertal e o espécime de Denisova (Green *e cols.*, 2010, Reich *e cols.*, 2010) mostra que 8 das 13 substituições humano-específicas do *HACNS1* presentes entre 236774003 e 236774084 também estão presentes no neandertal. Infelizmente, o genoma do Neandertal não oferece informação completa para toda a região do *HACNS1*, mas outras mutações *sapiens*-específicas localizadas fora da região de 81pb, incluindo a que nós descrevemos aqui (236774267), também estão presentes no genoma do Neandertal

(ver figura 1). O espécime de Denisova possui uma maior cobertura do genoma, e foi possível verificar que todas as 13 substituições humano-específicas estão presentes (ver figura 1).

Prabhakar *e cols.* (2008) demonstraram usando o teste estatístico de *log* de verossimilhança que o *HACNS1* está sobre seleção funcional em humanos, considerando que a rápida divergência é altamente inesperada dada à forte conservação desta região em outras espécies. Aqui nós realizamos testes adicionais para checar se a variação *H. sapiens*/Neandertal pode ser explicada pelo modelo neutro. Comparações entre os modelos mostraram um evidência significativa de seleção positiva quando as seqüências de *H. sapiens*/Neandertal são comparadas com as de outros primatas, incluindo chimpanzé (*Pan troglodytes*) (Tabela 1). O teste A de verossimilhança indicou que cerca de 18% dos sítios naquela região estão sob seleção positiva. Interessantemente, todos os 13 sítios reconhecidos como exclusivos na linhagem *Homo* mostram sinais de seleção positiva com probabilidade posterior maior que 99% (Figura 1). Nossas análises adicionais fornecem evidência de que um modelo neutro não explica a variação que ocorreu depois da separação nos gêneros *Homo* e *Pan*.

Nativos americanos apresentam uma pequena diversidade genética (calculado por π e índices de heterozigidade) e altos níveis de estruturação populacional (como determinado por F_{ST} e outras estatísticas similares) quando comparados com outras populações/grupos continentais (Cavalli-Sforza *e cols.*, 1994; Wang *e cols.*, 2007). O oposto é visto quando estudamos populações africanas sub-Saarianas (Rosenberg *e cols.*, 2002). Estes padrões de diversidade/divergência derivam principalmente de

processos demográficos (efeito-fundadores sucessivos com posterior expansão) relacionado com a dispersão do *Homo sapiens* moderno da África para outros continentes (Alonso and Armour, 2001; Ramachandran *e cols.*, 2005; Fagundes *e cols.*, 2007; Santos-Lopes *e cols.*, 2007; Wang *e cols.*, 2007). Como os efeitos demográficos afetam todo o genoma, os desvios deste modelo clássico de diversidade/divergência, assim como o observado neste estudo são esperados quando porções do genoma estão sofrendo pressão por seleção natural (Bamshad and Wooding, 2003; Barreiro *e cols.*, 2008).

Nossos resultados em nível populacional dão suporte à idéia de que estas 13 substituições conferem alguma vantagem adaptativa à linhagem *sapiens*, provavelmente à linhagem *Homo*, e em algum momento da história evolucionária dos homínídeos elas foram fixadas por seleção positiva. Após a fixação, uma forte seleção purificadora tem mantido a conservação intra-específica/intra-gênero do *cluster* de 81pb.

O real impacto deste achado deve ainda ser explorado através de mais estudos funcionais e populacionais, mas as implicações destas mutações podem indicar uma provável influência em traços comuns a espécie *Homo sapiens* ou ao gênero *Homo*, tais como polegares opositores, destreza manual e adaptações do calcanhar ou pé para o bipedalismo.

Agradecimentos do artigo

Nós agradecemos a Sandro L. Bonatto, Maria Luiza Petzl-Erler, Kim Hill, Ana Magdalena Hurtado and Wilson A. da Silva Júnior pelas respectivas amostras de esquimós, Guarani, Ache e africanos sub-Saarianos.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- Alonso S, Armour JA. 2001. A highly variable segment of human subterminal 16p reveals a history of population growth for modern humans outside Africa. *Proc Natl Acad Sci USA* 98: 864-869.
- Bamshad M, Wooding SP. 2003. Signatures of natural selection in the human genome. *Nat Rev Genet* 4: 99-111.
- Barreiro LB, Laval G, Quach H, Patin E, Quintana-Murci L. 2008. Natural selection has driven population differentiation in modern humans. *Nat Genet* 40: 340-345.
- Cavalli-Sforza LL, Menozzi P, Piazza A. 1994. *The history and geography of human genes*. Princeton: Princeton University Press.
- Duret L, Galtier N. 2009. Comment on "Human-specific gain of function in a developmental enhancer". *Science* 323: 714.
- Fagundes NJ, Ray N, Beaumont M, Neuenschwander S, Salzano FM, Bonatto SL, Excoffier L. 2007. Statistical evaluation of alternative models of human evolution. *Proc Natl Acad Sci USA* 104: 17614-17619.
- Meyer D, Single RN, Mack SI, Erlich HA, Thomson G. 2006. Signatures of demographic history and natural selection in the human major histocompatibility complex loci. *Genetics*. 173: 2121-2142.
- Nei M. 1987. *Molecular evolutionary genetics*. New York: Columbia University Press.
- Nei M. 2007. The new mutation theory of phenotypic evolution. *Proc Natl Acad Sci USA* 104: 12235-12242.
- Prabhakar S, Visel A, Akiyama JA, Shoukry M, Lewis KD, Holt A, Plajzer-Frick I, Morrison H, Fitzpatrick DR, Afzal V, Pennacchio LA, Rubin EM, Noonan JP. 2008. Human-specific gain of function in a developmental enhancer. *Science* 321: 1346-1350.
- Prabhakar S, Visel A, Akiyama JA, Shoukry M, Lewis KD, Holt A, Plajzer-Frick I, Morrison H, FitzPatrick DR, Afzal V, Pennacchio LA, Rubin EM, JP Noonan1 JP. 2009. Response to Comment on "Human-Specific Gain of Function in a Developmental Enhancer". *Science* 323: 714.
- Ramachandran S, Deshpande O, Roseman CC, Rosenberg NA, Feldman MW, Cavalli-Sforza LL. 2005. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proc Natl Acad Sci USA* 102: 15942-15947.
- Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen A, Moltke I, Metspalu M, Metspalu E, Kivisild T, Gupta R, Bertalan M, Nielsen K, Gilbert MT, Wang Y, Raghavan M, Campos PF, Kamp HM, Wilson AS, Gledhill A, Tridico S, Bunce M, Lorenzen ED, Binladen J, Guo X, Zhao J, Zhang X, Zhang H, Li Z, Chen M, Orlando L, Kristiansen K, Bak M, Tommerup N, Bendixen C, Pierre TL, Grønnow B, Meldgaard M, Andreasen C, Fedorova SA, Osipova LP, Higham TF, Ramsey CB, Hansen TV, Nielsen FC, Crawford MH, Brunak S, Sicheritz-Pontén T, Villems R, Nielsen R, Krogh A, Wang J, Willerslev E. 2010. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* 463:757-762.

- Reich D , Green RE , Kircher M, Krause J, Patterson N, Durand EY, Viola B, Briggs AW, Stenzel U, Johnson PLF et al (2010) Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* 468:1053-1060.
- Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW. 2002. Genetic structure of human populations. *Science* 298: 2381-2385.
- Santos-Lopes SS, Pereira RW, Wilson IJ, Pena SD. 2007. A worldwide phylogeography for the human X chromosome. *PLoS One* 2: e557.
- Wang S, Lewis CM, Jakobsson M, Ramachandran S, Ray N, Bedoya G, Rojas W, Parra MV, Molina JA, Gallo C, Mazzotti G, Poletti G, Hill K, Hurtado AM, Labuda D, Klitz W, Barrantes R, Bortolini MC, Salzano FM, Petzl-Erler ML, Tsuneto LT, Llop E, Rothhammer F, Excoffier L, Feldman MW, Rosenberg NA, Ruiz-Linares A. 2007. Genetic variation and population structure in Native Americans. *PLoS Genet* 3: e185.

6. TABELAS E FIGURAS

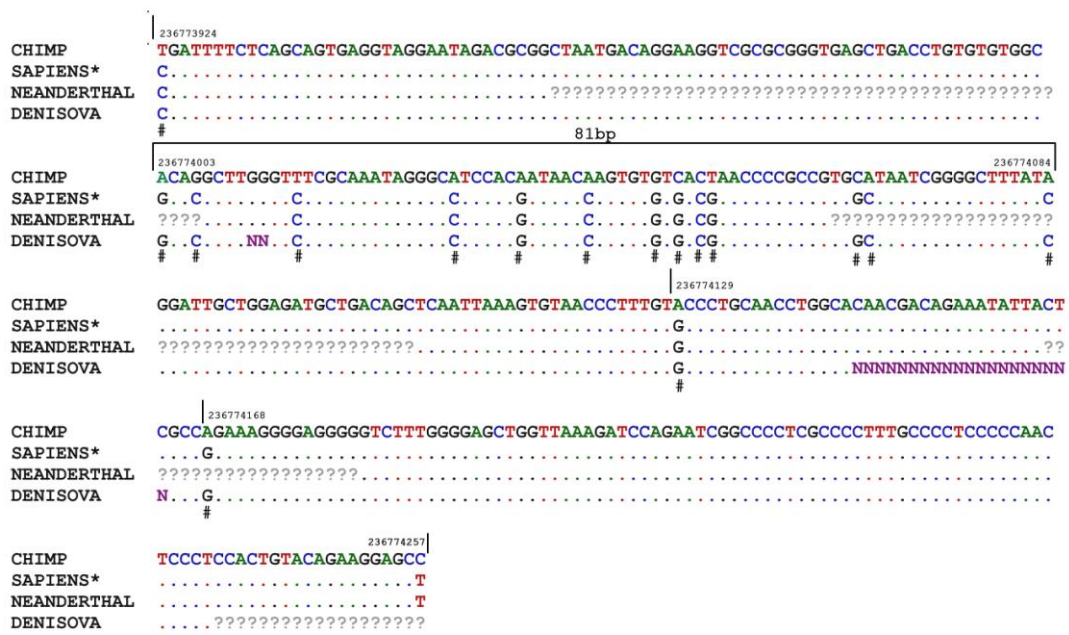


Figura 1 (modificada de Hünemeier e cols.2010). Identificação das 13 substituições humano-específicas no *enhancer HACNS1* (região de 81bp) e 4 substituições humano-específicas encontradas adicionalmente dentro dos 710 pb sequenciados.

* sem variação intraespecífica

sítios com seleção positiva com probabilidade de superior a 99%

Tabela 1. Parâmetros e escores de verossimilhança sob diferentes modelos considerando a variável ζ entre conjuntos de nucleotídeos avaliados

	Model	Estimated parameters	ℓ	<i>P-value</i>
Test 1	M1: Neutral model	Zeta[0] = 0.001000, p[0] = 0.699713; Zeta[1] = 1.000000, p[1] = 0.300287	-1737.911095	<0.001
	M3: 3 category	Zeta[0] = 0.001000, p[0] = 0.803405; Zeta[1] = 1.000000, p[1] = 0.019007 Zeta[2] = 5.044662, p[2] = 0.177588	-1728.952010	
Test 2	M1: Neutral model	Zeta[0] = 0.001000, p[0] = 0.699713; Zeta[1] = 1.000000, p[1] = 0.300287	-1737.911095	<0.001
	M2: 2 category	Zeta[0] = 0.001000, p[0] = 0.807452; Zeta[1] = 4.028553, p[1] = 0.192548	-1728.720487	

7. ARTIGO ORIGINAL