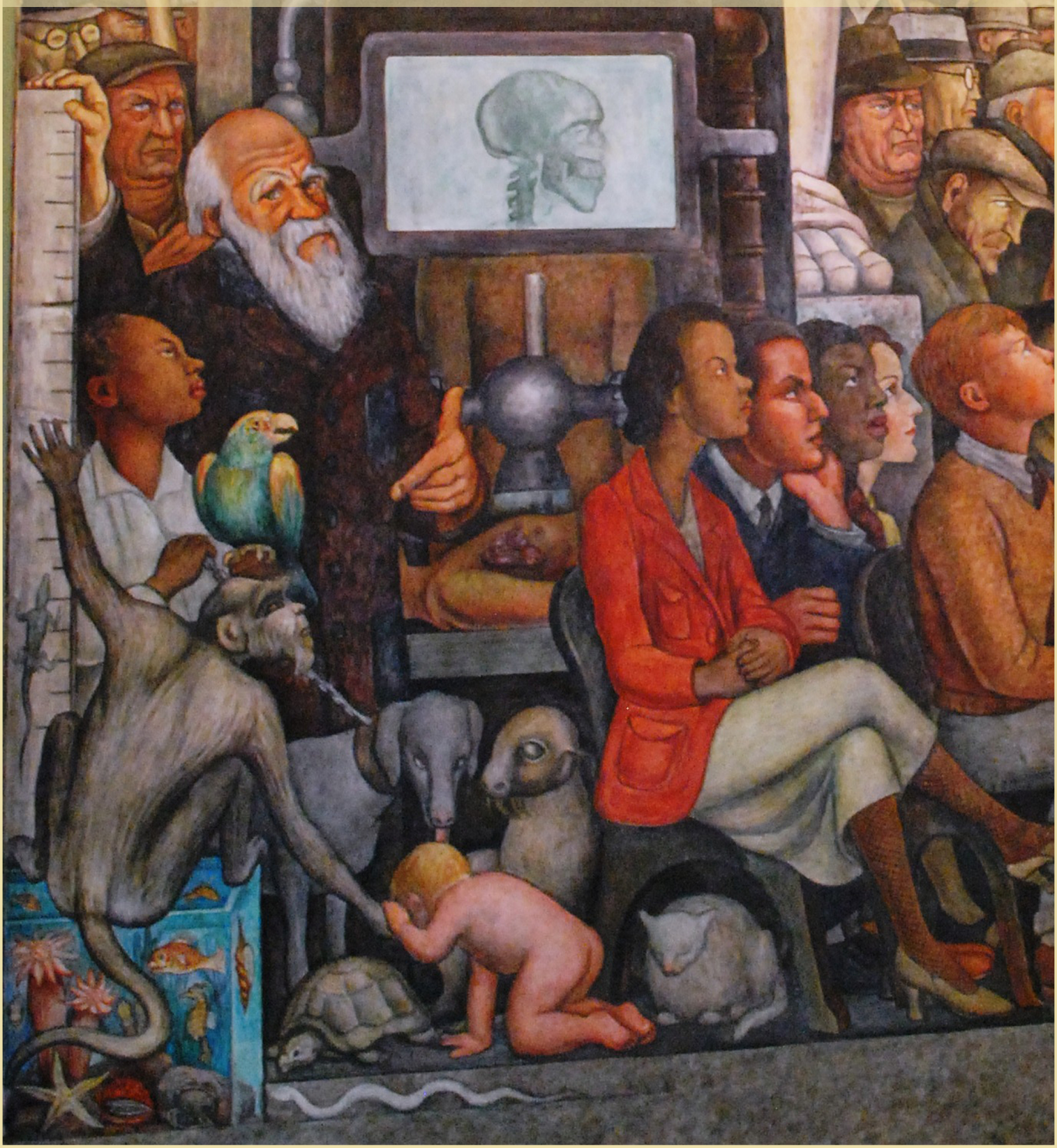
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
Programa de Pós-Graduação em Genética e Biologia Molecular

TESE DE DOUTORADO

PROCESSOS E FATORES EVOLUTIVOS ENVOLVIDOS NA GERAÇÃO E MANUTENÇÃO DA DIVERSIDADE E SEU SIGNIFICADO NA CONEXÃO GENÓTIPO-FENÓTIPO

Tábita Hünemeier

Orientadora: Profª Drª Maria Cátira Bortolini



Capa
El hombre en la encrucijada, de Diego Rivera
Palácio de Bellas Artes, Ciudad de México, México, 1934
Fotografia de Tábita Hünemeier, 2009.

Contra-capa
El Maiz, de Diego Rivera
Palácio Nacional, Ciudad de México, México, 1940.
Fotografia de Sol Kawage, 2006.

Figura interna
Modificado de El Caballero Águila, de Jesús Helguera, México.

Arte
Vanessa Rodrigues Paixão-Cortês

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

PROCESSOS E FATORES EVOLUTIVOS ENVOLVIDOS NA GERAÇÃO E MANUTENÇÃO DA DIVERSIDADE E SEU SIGNIFICADO NA CONEXÃO GENÓTIPO-FENÓTIPO

TÁBITA HÜNEMEIER

ORIENTADORA: Prof[a] Dr[a] MARIA CÁTIRA BORTOLINI

Tese submetida ao Programa de Pós-graduação em Genética e Biologia Molecular da Universidade Federal do Rio Grande do Sul como requisito para obtenção do título de Doutor em Ciências.

Porto Alegre, setembro de 2010.

Este trabalho foi desenvolvido no Laboratório de Genética de Populações Humanas e Evolução Molecular da Universidade Federal do Rio Grande do Sul, Galton Laboratory, University College London, Inglaterra e na Unidad de Biología Molecular y Medicina Genómica, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Cidade do México, México, entre outubro de 2006 e setembro de 2010, com o financiamento do Conselho Nacional de Pesquisa (CNPq) e da Coordenação Nacional de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

Aos povos indígenas da América.

Fundación Científica del Racismo
Eduardo Galeano
in Espejos: una historia casi universal (Editorial Siglo XXI, 2008).

Raza caucásica se llama, todavía, la minoria blanca que ocupa la cúspide de las jerarquias humanas. Así fue bautizada em 1775 por Johann Friedrich Blumenbach. Este zoólogo creia que el Cáucaso era la cuna de La humanidad y que de allí provenían la inteligência y la belleza. El término se sigue usando, contra toda evidencia, em nuestros dias. Blumenbach había reunido 245 cráneos que fundamentaban el derecho de los europeos a humillar a los demás. La humanidad formaba una pirâmide de cinco pisos. Arriba, los blancos. La pureza original había sido arruinada, pisos abajo, por las razas de piel súcia: los nativos australianos, los índios americanos, los asiáticos amarillos. Y debajo de todos, deformes por fuera y por dentro, estaban los negros africanos. La ciencia siempre ubicaba a los negros en el sótano. En 1863, la Sociedad Antropológica de Londres llegá a la conclusión de que los negros eran intelectualmente inferiores a los blancos, y solo los europeos tenían la capacidad de humanizarlos y civilizarlos. Europa consagro sus mejores energias a esta noble misión, pero no tuvo suerte. Casi um siglo y médio después, en el ano de 2007, el estadunidense James Watson, premio Nobel de Medicina, afirmo que está científicamente demostrado que los negros siguen siendo menos inteligente que los blancos.

Agradecimentos

À minha orientadora, Prof[a] Dr[a] Maria Cátira Bortolini, pelo constante estímulo, entusiasmo e pelo seu exemplo de dedicação à pesquisa e à vida acadêmica. Bem como pela confiança e compreensão ao longo destes seis anos de trabalho conjunto, durante os quais ela foi imprescindível para estabelecermos esta nova linha de pesquisa da qual este trabalho é um dos primeiros alicerces.

Ao Prof. Dr. Francisco Mauro Salzano, por ter me acolhido como se fosse sua orientada, por sua disposição em ajudar e cordialidade em me receber a qualquer momento, para discutir qualquer assunto, até mesmo as vitórias do Colorado! E, como não poderia deixar de ser mencionado, por seu pioneirismo e dedicação ao estudo das populações nativas da América, sem o qual hoje não haveria este trabalho de investigação científica.

Ao Prof. Dr. Andrés Ruiz-Linares pela co-orientação, amizade, atenção, confiança e por proporcionar, ao me receber em seu grupo de pesquisa, uma mudança na minha maneira de enxergar e fazer ciência.

Ao Dr. Samuel Canizales-Quinteros pela produtiva colaboração ao longo destes três anos de trabalho com a Universidad Nacional Autónoma do México.

Ao Dr. Rolando González-José por me receber no CENPAT, Argentina, pelas conversas esclarecedoras em várias etapas deste trabalho e por estar sempre disposto e interessado em colaborar com nossos projetos. E também por haver me apresentado o Fernet!

À Prof[a] Dr[a] Vera Valente Gaiesky, não somente pela ajuda na revisão da versão final deste trabalho, mas também por todo entusiasmo que certamente atuou como propulsor para o aperfeiçoamento da minha formação.

Ao Prof. Dr. Aldo Mellender de Araújo, pela amizade, pelos conselhos e pelo exemplo.

Ao Prof. Dr. Claiton D. Bau e à Verônica Contini pela válida colaboração, pois sem a boa disposição de ambos, várias partes deste trabalho não haveriam sido realizadas.

Ao meu bom amigo Víctor Acuña-Alonzo da Escuela Nacional de Antropología e História, México, uma pessoa fundamental para o êxito deste trabalho. É impossível transmitir em palavras meu agradecimento por seu companheirismo e pelo carinho com que sempre fui recebida por ele e por sua família ao longo destes anos; bem como pela importância de sua influência na forma de hoje eu pensar e estudar os povos nativo-americanos.

Ao amigo Jorge Gómez-Valdéz da Universidad Nacional Autónoma do México, por toda ajuda, amizade, carinho, e principalmente por sua paciência com uma bióloga geneticista tentando pensar como uma antropóloga física.

À Vanessa Rodrigues Paixão-Cortês pelo coleguismo, amizade, generosidade, e pela sempre disposição em ajudar, fora e dentro do ambiente acadêmico.

Aos meus colegas de laboratório e amigos, Caio César Cerqueira, Eduardo Guerra Amorim, Eli Vieira, Luana Maciel e Angelina García, pelo bom ambiente de trabalho, por sempre estarem prontos a cooperar, pelos cafés na tarde e pelas cervejas na noite.

Aos meus amigos do departamento, Luciana Tovo Rodrigues, Carla Martins Lopes, Daniel Graichen e Pedro Cordeiro Estrela, pela amizade e por serem pessoas honestas, competentes e dedicadas ao que fazem.

Aos meus amigos mexicanos, Mónica Ballesteros Romero, Alín Acuña-Alonzo, Celta Gómez-Trejo e Rodrigo Barquera por fazerem sempre o máximo possível para que eu me sentisse em casa em seu país.

Aos grandes amigos com quem tive o prazer de trabalhar na University College London, Ana Victoria Valencia Duarte, Pau Corral-Montañés, Claudio Bravi, Magda Gaya Vidal, Barbara Kremeyer, Stephane Maziéres, Karla Sandoval Mendoza, Andrés Moreno Estrada, Matthew Nunes, Desmond Campbell, Mari-Wyn Burley e Sushma

Jansari. Se não fosse por eles, minha estada no Galton Laboratory não haveria sido tão significativa, pois além de conviver com excelentes colegas de trabalho, convivi com excelentes pessoas que se tornaram amigos presentes em todos momentos desde então. Gracias! Gràcies! Danke! Thanks! Merci! And cheers (with a pint of beer)!

Ao Elmo e à Ellen, pela eficiência e pela amizade, sem seu trabalho e dedicação seria impossível para nós alcançarmos o alto nível que temos hoje em nosso Programa de Pós-graduação.

Ao meu fiel escudeiro, Estevan Mocellin, por estar sempre comigo em todos os bons, e também nos não tão bons, momentos da minha vida.

À minha prima Alejandra Müller, por estar sempre por perto quando eu precisei.

Aos meus tios Arno e Lourdes Müller por todo incentivo e pelo exemplo de dedicação à pesquisa e à Universidade.

Às minhas famílias emprestadas, representadas aqui por Iraci Machado Moreira, Indiajara Simeoni Wermann e Tânia Bortoluzzi. Obrigada por me acolherem como se eu fosse parte de suas famílias.

Ao meu irmão Jefferson Hünemeier, por se fazer presente quando necessário.

E, por último, mas não menos importante, à minha mãe, Hiltrudes Maria Hünemeier, não só por ter me educado, me ensinado valores e ter me dado liberdade para que eu pudesse escolher meu caminho, mas também por nestes últimos anos ter deixado a sua casa para cuidar da minha, afim de que eu pudesse sair do Brasil para aperfeiçoar-me sempre que tivesse oportunidade. Mama, sem a tua dedicação e esforço nada disto seria possível. Muito obrigada!

SUMÁRIO

RESUMO

Um grande desafio para os geneticistas da atualidade é desvendar as relações que os genes ou regiões gênicas, bem como suas variáveis, têm com determinados fenótipos. Deste modo as abordagens apresentadas neste estudo buscaram ajudar a desvendar a conexão entre alguns destes genes/regiões gênicas com variações morfológicas e metabólicas em populações humanas e em outros mamíferos. Os resultados deste trabalho foram apresentados sob a forma de artigos científicos, sendo que os resultados podem ser resumidos como segue:

1) Hünemeier et al. 2009. *TCOF1 T/Ser* variant and brachycephaly in dogs. *Animal Genetics*; 40(3):357-358. Neste trabalho foi demonstrado que o polimorfismo do éxon 4 (*C396T, Pro117Ser*) do gene TCOF1, está envolvido nos mecanismos de variação morfológica da face, sendo encontrado com mais freqüência em cães braquicéfalos do que meso e dolicocéfalos, porém esta associação não é direta como descrita anteriormente. Foram encontrados cães dolicocéfalos homozigotos para o alelo *T/Ser*, bem como cães braquicéfalos homozigotos para o alelo *C/Pro.* Considerando a intricada rede de regulação e expressão deste gene, a mutação por nós estudada pode ser co-dependente de outras variantes na mesma rede de interação gênica;

2) Hünemeier et al. *FGFR1* gene haplotype tag-SNPs, linkage disequilibrium patterns, and their influence in normal craniofacial variation (em preparação para ser submetido para a revista *Human Heredity).* Neste trabalho dados de variação craniofacial foram obtidos e comparados através de análises uni e multivariadas com polimofismos do gene *FGFR1* (N = 333). O alelo rs46470905 C parece estar envolvido na variação normal relativo ao índice cefálico encontrado em populações humanas. Entretanto, claramente outros fatores da rede de desenvolvimento também parem estar envolvidos.

Os blocos de LD variam em cada grupo populacional investigado, influenciando também a relação de cada *background* genético com os seus respectivos fenótipos.

3) Hünemeier et al 2010. Population Data Support the Adaptive Nature of HACNS1 Sapiens/Neandertal-Chimpanzee Differences in a Limb Expression Domain. *American Journal of Physical Anthropology.* No prelo. DOI: 10.1002/ajpa.21378. Neste artigo envolvendo o sequenciamento do *enhancer HACNS1* em 194 indivíduos de vários grupos geográficos foi possível demonstrar que a ação da seleção positiva na linhagem *Homo* foi responsável pela fixação das mutações específicas encontradas nesta região quando comparada com o gênero *Pan*. A ausência de variação intra-específica no *Homo sapiens* e no *Neandertal* reforça o papel funcional do *enhancer*, mantido inalterado posteriormente pela ação extrema de seleção purificadora. Como discutido no artigo, é possível que o mesmo esteja envolvido na regulação de genes que caracterizam morfologicamente nosso gênero, tais como postura ereta e destreza das mãos.

4) Hünemeier et al. Gene-culture Dynamics: An Example Involving Native Americans (em preparação para ser submetido para a revista *American Journal of Physical Anthropology).* Foi possível com este estudo determinar que o alelo funcional autóctone nativo Americano *230Cys* do gene *ABCA1* teve uma origem a cerca de 8.268 ± 5.916 anos antes do presente, sendo que o aumento de sua freqüência na América Central/Mesoamérica, poderia estar relacionada com a transição de uma cultura caçadora-coletora para agriculturalista, baseada no cultivo do milho. Este caso seria o primeiro exemplo bem documentado de evolução gene-cultura envolvendo alelo autóctone nativo americano. Este cenário, no entanto, não poderia ser extrapolado para a América do Sul e América do Norte, visto que são duas regiões com distintos padrões culturais e ecológicos.

Conjuntamente estes achados permitem sugerir que as relações fenótipo-genótipo são influenciadas por diversos fatores biológicos/culturais/ambientais e que a variação existente dentro da espécie humana, é resultado de uma intricada rede de interações genéticas, que atuam em todos os níveis do desenvolvimento. Adicionalmente, as interações desta rede no *Homo sapiens sapie*ns se distingue daquelas de outros mamíferos, em especial pela ação catalisadora de processos derivados justamente da característica que torna nossa espécie singular, a cultura.

ABSTRACT

A major challenge for today geneticists is to unveil the relationships that genes or gene regions, and its variables, have over certain phenotypes. Thus the approaches presented in this study sought to help unravelling the connection between some of these genes / gene regions with metabolic and morphological variations in human populations and other mammals. The results of these works were presented in scientific papers, and the results can be summarized as follows:

1) Hünemeier et al. 2009. *TCOF1 T* / Ser variant and brachycephaly in dogs. Animal Genetics, 40 (3):357-358. This work demonstrated that the polymorphism of the *TCOF1* exon 4 (C396T, Pro117Ser) is involved in the mechanisms of morphological variation of the face, being found most often in brachycephalic dogs than meso-and dolichocephalic dogs. However, this association is not straightforward as previously described. Dolichocephalic dogs were found homozygous for the allele T / Ser, and brachycephalic dogs homozygous for the allele C / Pro. Considering the intricate network of regulation and expression of this gene, the mutation that we studied can be co-dependent on other variants on the same gene network;

2) Hünemeier et al. FGFR1 gene haplotype-tag SNPs, linkage disequilibrium patterns, and their influence in normal craniofacial variation (in preparation for Human Heredity journal). In this paper craniofacial variation data were obtained and compared by univariate and multivariate analyses with polymorphic gene FGFR1 (N = 333). Rs46470905 C allele seems to be involved in the normal range relative to the cephalic index found in human populations. However, other factors in the development network

also enjoy being involved. LD blocks vary in each population group investigated, also influencing the relationship of each genetic background with their respective phenotypes.

3) Hünemeier et al 2010. Population Data Support the Adaptive Nature of HACNS1 sapiens / Neanderthal-Chimpanzee Differences in the Limb Expression Domain. *American Journal of Physical Anthropology*. In press. DOI: 10.1002/ajpa.21378. This article involved the sequencing of the enhancer *HACNS1* in 194 individuals from several continents. It was possible to demonstrate that the action of positive selection in the *Homo* lineage was responsible for setting the specific mutations found in this region when compared with the genus *Pan*. The absence of intra- specific variation in *Homo sapiens* and *Neandertal* reinforces the functional role of the enhancer, remaining unchanged after due to extreme action of purifying selection. As discussed in the article, it is possible that it is involved in regulating genes that characterize our genus morphologically, such as upright posture and hands dexterity.

4) Hünemeier et al. Gene-culture Dynamics: An Example Involving Native Americans (in preparation for American Journal of Physical Anthropology Journal). It was possible to determine from this study that the functional allele indigenous Native American *230Cys ABCA1* gene had an origin about 8268 ± 5916 years before present, and the increase in its frequency in Central America / Mesoamerica could be related to the transition between the hunter-gatherer to agricultural cultures, the last based on the cultivation of maize. This case would be the first well documented example of gene-culture co-evolution involving and indigenous Native American functional allele. This scenario, however, could not be extrapolated to South America and North America,

since they are two regions with distinct cultural and ecological standards. Together these findings may suggest that the phenotype-genotype relationships are influenced by several biological / cultural / environmental factors and the variation within the species is the result of an intricate network of genetic interactions, which operate at all levels of development. Additionally, the interactions of this network in *Homo sapiens sapiens* is distinguished from those of others mammals, in particular by the action of catalytic process derived precisely from the characteristic that makes our species unique, the culture.

INTRODUÇÃO

Phenotypic variation is the raw material for natural selection, yet a century after Darwin, it is an almost unknown subject.
Leigh Van Valen, 1974

I. Variação Fenotípica

A variação fenotípica é tanto conceitualmente quanto historicamente um tópico central na Biologia Evolutiva. A observação da variação fenotípica foi o alicerce sob o qual Charles Darwin construiu sua teoria sobre a evolução das espécies através da seleção natural. Esta mesma variação, em nível inter e intra-específico, torna-se fundamento para mais tarde Ernst Mayr, Sewall Wright e outros pesquisadores, sustentarem uma corrente de pensamento que buscava explicar o surgimento e a manutenção da variabilidade sob a luz da genética. Surge a Síntese Evolutiva Moderna, que incorpora à Teoria da Evolução de Darwin os conhecimentos da Genética Mendeliana. Essencialmente, a Síntese Moderna introduziu a conexão entre dois importantes fenômenos, a mutação, como fonte primordial de variação e a seleção natural. No entanto, o cenário envolvendo fenômenos microevolutivos, genes e características físicas e metabólicas, que caracterizam as complexas relações entre genótipo e fenótipo só agora, passados 150 anos da publicação do livro seminal de Darwin e cerca de cem anos após a postulação da Síntese Evolutiva, começa a ser revelado (Rockman, 2008).

Estudos genéticos baseados na filogenia e na homologia têm procurado desvendar como o genótipo afeta o fenótipo. O pressuposto básico dessas investigações

estaria fortemente correlacionado à seleção natural, pois homologia de fenótipos implicaria na subjacente homologia de genótipo. Muitos exemplos envolvendo regiões muito conservadas têm sido usados para sustentar este pressuposto. No entanto, existem situações que fogem deste contexto, pois fenótipos idênticos ou similares pode ser resultado de mutações em genes ou regiões gênicas diferentes, já que normalmente muitos genes estão conectados através de redes funcionais e regulatórias. Da mesma forma, mutações idênticas num determinado gene não necessariamente levam ao mesmo fenótipo, justamente porque o *background* envolvendo os demais genes da rede funcional e regulatória pode ser diferente. Por exemplo, nosso grupo de pesquisa demonstrou recentemente (Paixão-Côrtes *et al*., 2010a, 2010b) que a variante *240Pro* localizada no exon 3 do gene *PAX9* seria a responsável pelo padrão de agenesia do terceiro molar em populações do sul do Saara. Porém, fora da África o alelo variante não explicaria sozinho altos níveis deste tipo de agenesia tanto em europeus quanto em asiáticos. Os autores também investigaram cerca de 6.000 SNPs em 101 outros genes do desenvolvimento dentário e constataram que o *background* genético desta rede era diferente na África e fora dela, o que poderia justificar os resultados encontrados com o polimorfismo *Ala240Pro* do *PAX9*, bem como a existência de um complexo dentário africano há muito tempo reconhecido na literatura. Paixão-Côrtes *et al*. (2010a, 2010b) concluem que os diferentes complexos genéticos estariam estavam sujeitos a ação da seleção natural, provavelmente relacionados com a dispersão dos humanos modernos para fora da África.

O contexto do trabalho apresentado acima salienta que para o entendimento da conexão genótipo e fenótipo é fundamental uma abordagem multidisciplinar. A aferição do fenótipo por métodos como a morfometria, por exemplo, tem um papel importante em estudos desta natureza à medida que fornece métodos para quantificação da variação fenotípica de forma e tamanho. A biologia do desenvolvimento, de forma complementar dá suporte para o melhor entendimento das relações temporais e espaciais da expressão gênica durante o desenvolvimento de determinado organismo. Outro pilar importante para a compreensão da variação fenotípica e genotípica deve ser firmemente embasado na Genética de Populações, pois esta busca desvendar o papel da seleção natural, deriva genética, padrões de fluxo gênico, estruturação e tamanho populacional, geografia e cultura, no surgimento e manutenção de determinado genótipo e consequentemente de determinada e correspondente característica fenotípica.

Finalmente, ao se estudar o *Homo sapiens sapiens* devemos ainda acrescentar ao já complexo contexto descrito acima mais um fator adicional: cultura. Como veremos nesta tese, hábitos culturais podem ser catalisadores de fatores evolutivos envolvidos na geração e manutenção da diversidade, com significante significado na conexão genótipo-fenótipo.

II. Genótipo-Fenótipo

Do ponto de vista genético, existem várias teorias em relação aos efeitos de variações no genoma sobre o surgimento e manutenção de determinado fenótipo. Nei

(2007) ao analisar a expressão de genes que possuem expressão importante nos primeiros estágios do desenvolvimento, e que alterações em estes genes levariam a mudanças fenotípicas importantes, propõe que a variação fenotípica morfológica inter e intra-específica seria resultado de mudanças em poucos genes, que atuando sozinhos ou interagindo entre si, produziriam grande efeito sobre algumas características fenotípicas. Dada a complexidade da diversidade morfológica existente, é provável que genes de maior efeito estejam associados a muitos outros genes de pequeno efeito, atuando em diferentes etapas do desenvolvimento (e.g. Cor de pelo e olhos em várias espécies; tamanho das patas em cães), ainda que os genes de grande efeito possam ser responsáveis sozinhos pela expressão de caracteres em alguns casos (e.g. *IGF1* em cães, Sutter et al. 2007).

Deste modo não se pode pretender que desvendar a conexão entre genótipo e fenótipo seja uma tarefa fácil, ainda mais quando estas estariam relacionadas a variações normais dentro de uma mesma espécie. Neste contexto, estudar os padrões de variação de genótipos e fenótipos dentre e entre populações humanas é fundamental Além disso, estudos em nível populacional permitem desvendar as forças evolutivas que atuam/atuaram na diferenciação destas populações, e se estas diferenças seria produto de fatores estocásticos ou não.

III. Variação Morfológica Normal e Genes de Maior Efeito

Considerando o que foi exposto, é esperado que vários genes estejam envolvidos na morfologia craniofacial humana, assim como outros fatores mecânicos, ambientais, e epigenéticos (Enlow, 1990; Roseman, 2004; Coussens e Dall, 2005). Um exemplo da complexa interação desses vários fatores pode ser visto na extensa variação de fenótipos craniofaciais existentes nas diversas populações humanas. Além disso, é bem conhecido a relativamente rápida mudança que ocorreu no crânio na evolução da linhagem hominidae (Lawson et al., 2006). Até o momento, vários genes foram relacionados às variações craniométricas normais em vertebrados (Kim *et al.*, 1998; Lawson et al., 2006; Richsmeieret al., 2007). Porém, os mesmos teriam uma significativa redundância entre si, resultando no fato de que mutações em genes diferentes podem freqüentemente levar a fenótipos similares (Kim *et al.*, 1998), o que torna mais complicado desvendar a conexão entre genótipo e fenótipo. Os últimos autores também evidenciaram o grande número de substituições nucleotídicas em genes envolvidos no desenvolvimento craniofacial. O papel que esta variabilidade genética desempenha na diversidade normal, entretanto, permanece desconhecida.

O controle genético do desenvolvimento craniofacial tem sido o foco de vários estudos nos últimos anos envolvendo modelos animais, mas em humanos, as investigações estão particularmente concentradas em indivíduos afetados por diversas alterações craniofaciais sindrômicas e não-sindrômicas (Howard *et al.*, 1997; Francis-West *et al.*, 2003; Albuisson *et al.*, 2005). Poucos estudos têm proposto identificar variações que poderiam estar relacionadas à diversidade craniofacial normal. Para que

um gene seja indicado como um bom candidato para estudos que visam desvendar a relação deste com características morfológicas, é avaliar se há alguma evidência funcional de que o mesmo está envolvido com o traço fenotípico em consideração.

Com base nisto e na teoria de genes de grande efeito (Nei, 2007) é possível especular que genes envolvidos em situações patológicas sejam bons candidatos para estudos que visam identificar as causas moleculares que levam as variações normais encontradas dentro e entre espécies. Essa estratégia, por exemplo, tem sido utilizada com sucesso em estudos com genes para fatores de transcrição e sua relação com a fala (Enard *et al*., 2002) e variações dentais dentro da nossa espécie (Pereira *et al.*, 2006; Paixão-Côrtes *et al*., 2010a, 2010b).

O complexo craniofacial pode ser dividido em três regiões: a base craniana, a face craniana e caixa craniana (calvária). Os ossos dessas regiões se desenvolvem e crescem por meio de dois processos diferentes: ossificação endocondral e ossificação intramembranosa (Junqueira e Carneiro, 2004). Os ossos da base craniana são formados por ossificação endocondral, enquanto os ossos da calvária, juntamente com os ossos da mandíbula, ocipital e temporal são formados por ossificação intramembranosa.

Na calvária, as suturas são os sítios de crescimento no desenvolvimento embrionário tardio e pós-natal. Na base craniana há um tipo de crescimento pós-natal mais lento, chamado sincondrose, onde há substituição gradual das células, processo fundamental para todo o desenvolvimento do crânio e do posicionamento da mandíbula e maxila superior (Enlow, 1990).

Mecanismos de sinalização celular precisos governam esses processos de crescimento ósseo, e assim como em outros sistemas do corpo, eles envolvem genes para fatores de crescimento altamente conservados na trajetória evolutiva dos vertebrados, como por exemplo, os FGF (*Fibroblast Growth Factor*) e os MSXs (Wilkie *et al*., 1995).

III. 1. Bases moleculares e genéticas do desenvolvimento morfológico

Os fatores de crescimento de fibroblastos (FGFs) e seus respectivos receptores (FGFRs - *Fibroblast Growth Factor Receptors*) estão envolvidos em vários estágios da diferenciação dos osteoblastos, sendo fundamentais na sinalização que leva a ativação de outros genes, entre os quais homeogenes (e.g. *Twist* e *MSXs*; Rice *et al.*, 2000). Eles também têm papel crucial na formação e crescimento dos ossos da calvária e base craniana, regulando a substituição de tecidos ao longo do desenvolvimento e crescimento pós-natal (Helms *et al.*, 2005).

Mutações em genes que codificam receptores desses fatores de crescimento de fibroblastos (*FGFRs*) têm sido identificadas como causas de condrodisplasias, acondroplasias, hipocondroplasias, assim como craniosinostoses sindrômicas e não sindrômicas (Wilkie, 1997; Passos-Bueno *et al.*, 1999). Todas essas condições causam anomalias no desenvolvimento e crescimento da região craniofacial. A craniosinostoses que é resultado da fusão prematura das suturas calvárias, leva a assimetrias faciais,

extrema dolicocefalia ou braquicefalia (dependendo da mutação), achatamento da face e possível retardo mental, dentre outras alterações (Wilkie, 1997).

Os genes dos receptores de fatores de crescimento de fibroblastos desempenham múltiplos e fundamentais papéis na morfogênese craniofacial em várias fases do desenvolvimento (Helms *et al.*, 2005). Por exemplo, O *FGFR1* tem um papel crucial no desenvolvimento da caixa craniana, enquanto o *FGFR3* parece estar associado ao crescimento dos ossos na base do crânio. Muitos estudos têm sido realizados com estes dois genes, sendo que mutações nos mesmos levam a fenótipos com severas anomalias craniofaciais (Wilkie, 1997; Rice *et al*.,2003).

A etapa do desenvolvimento em que ocorre o fechamento das suturas da calvária parece determinar o formato da caixa craniana, de modo que o fechamento parece ser dependente do local no gene onde ocorre a mutação. Como citado anteriormente, evidências disso são as mutações que levam aos fenótipos braquicéfalos e dolicocéfalos patológicos (Wilkie, 1997). Vale lembrar que em populações normais o índice cefálico (IC = comprimento lateral/comprimento antero-posterior, onde IC>80=braquicéfalo, 75<IC<80=mesocéfalo e IC<75=dolicocéfalo; Coussens e Dall, 2005) apresenta distintas e variadas distribuições. Ainda que exista heterogeneidade dentro de uma população ou região geográfica bem como superposição entre populações/ no que diz respeito a estas e outras medidas, existem extremos craniométricos que podem ser associados a populações específicas (Roseman, 2004).

O fato de mutações no gene *FGFR1* resultarem em fenótipos anormais que apresentam pelo menos alguns traços morfológicos similares com aqueles encontrados

na população normal, levou Coussens e Dall (2005) a desenvolverem o único estudo com este gene em populações humanas saudáveis. Os autores analisaram fragmentos envolvendo todo o gene *FGFR1* através de denaturação por cromatográfica líquida de alta resolução (dHPLC) para identificar SNPs que poderiam estar relacionados à variação craniométrica normal dentro da nossa espécie. As medidas morfológicas foram obtidas através de fotografias. Na investigação que envolveu 44 indivíduos de quatro populações distintas identificadas como afro-americanos, asiáticos, aborígines australianos e caucasóides (australianos, descendentes de britânicos) foram descritos 17 SNPs. Considerando apenas os 8 SNPs que apresentavam maior variação, foi possível identificar 9 haplótipos comuns nos grupos investigados. Destes, três estariam presentes em mais do que 87% da amostra como um todo. Foi encontrada ainda uma correlação negativa entre o índice cefálico e o polimorfismo rs4647905 do *FGFR1*. Esse trabalho, além de mostrar a associação entre um genótipo e uma característica craniométrica humana pela primeira vez, também corroborou a idéia de que genes envolvidos em situações patológicas são bons candidatos para estudos que visam esclarecer a natureza das variações craniométricas normais.

O gene *TCOF1* codifica uma fosfoproteína que parece estar envolvida no processamento ribossomal, mas ainda não está totalmente clara a função e estrutura da mesma. A importância do *TCOF1* na morfologia do craniofacial é evidenciada por estudos de mutações nesse gene, correlacionando-o com a chamada Síndrome Treacher-Collins (TCS). A TCS é caracterizada por más-formações mandibulo-faciais. A maioria

das mutações no *TCOF1* que causam a TCS são mutações que levam a uma proteína truncada não funcional.

Além de uma série de mutações diretamente associadas à TCS, já foram descritos pelo menos outros 20 SNPs (*Single Nucleotide Polymorphisms*) na seqüência codificadora de *TCOF1* presentes em populações normais (Splendore *et al*., 2000; 2005). Isso implica que há um SNP a cada 210 pares de bases (pb), o que é superior à média encontrada para genes ativos (um a cada 346 pb). Nas investigações dos últimos autores, não foram detectadas diferenças nas distribuições dos alelos entre controles e pacientes com Síndrome de Treacher-Collins, com exceção de um SNP, localizado no éxon 8. Entretanto, os autores não descartaram a possibilidade da associação ser devido a problemas amostrais. Além disso, não foi possível estabelecer a relação de algum(ns) deste(s) SNPs com a variabilidade clínica encontrada na Síndrome (Splendore *et al*., 2000).

Os indivíduos controles nos estudos de associação realizados nas investigações com o gene *TCOF1*, são europeus ou euro-descendentes, o que inviabiliza inferir qualquer panorama mais geral sobre a variação deste gene em pessoas normais dentro da nossa espécie. Além disso, os dados disponíveis até este momento não permitem inferências sobre a natureza desta variação, se é seletivamente neutra ou se é resultado da ação da seleção natural.

Estudos com *TCOF1* em outros mamíferos também mostraram resultados interessantes. Shows *et al.* (2006) após investigarem um espécime de macaco do velho mundo (Rhesus; *Macaca mulatta*) afetado pela TCS, concluíram que a doença estava

relacionada com a expressão reduzida do gene e não com uma mutação especifica em regiões codificadoras ou sítios de *splicing*. Logo, alterações em alguma rota regulatória pareceriam estar envolvidas com a patologia, e não com alguma modificação na estrutura molecular original do gene *TCOF1*. Os autores também especulam a possibilidade do *TCOF1* estar associado à presença do *sinus* maxilar (cavidade no osso da face acima da maxila e abaixo da órbita ocular) em Rhesus. O *sinus* maxilar encontra-se também presente nos grandes macacos e nos humanos. Porém, essa estrutura não está presente nos demais macacos do velho mundo. Sendo assim, essa característica craniofacial teria sido perdida após a separação dos macacos do velho mundo e dos grandes macacos/homem, sendo readquirida posteriormente exclusivamente no ramo filogenético que deu origem ao *Rhesus*.

Haworth *et al.* (2001), por sua vez, analisaram esse mesmo gene em treze diferentes raças de cães, agrupados de acordo com o formato da cabeça. Nesta investigação foi encontrada uma forte associação entre uma substituição não-sinônima (mutação de *C396T, Pro117Ser*) no éxon 4 e a diferenciação craniofacial nos cães. O alelo variante *T/117Ser* só estava presente em raças que apresentam braquicefalia (e.g. Bulldog Francês e Boxer). Ou seja, em cães, uma única troca de aminoácidos foi associada à marcante retração dos ossos da face e conseqüente arredondamento da sua caixa craniana.

A haploinsuficiência do gene *TCOF1* leva à deficiência das células da crista neural, o que resulta em severas anomalias craniofaciais. As células da crista neural são populações de células migratórias que dão origem à maioria dos tecidos cartilaginosos e

ósseos. O *TCOF1* apresenta variação temporal de expressão nos dois primeiros arcos faríngeos e nas células da crista neural que dão origem aos ossos da face, e é expresso também de uma forma variada em diversos tecidos adultos, como fígado pâncreas e pulmão (Dixon *et al*. 2006). Essa variação temporal de expressão possivelmente sugere um mecanismo regulatório específico nas células embrionárias dos arcos e da crista neural, provavelmente influenciado pela presença de um elemento ativador, ausente nos demais tecidos onde ele é expresso na vida adulta (Masotti *et al*. 2006).

Considerando essa expressão diferencial temporal durante o desenvolvimento, pode-se sugerir que mutações no promotor do *TCOF1* podem alterar os níveis de expressão durante o desenvolvimento, podendo esse mecanismo estar envolvido na diversificação das estruturas craniofaciais normais.

Tendo-se em vista a possível influência de genes de maior efeito na variação morfológica normal, tanto na diversidade craniofacial quanto em outras características morfológicas, dentro e entre espécies, uma alternativa interessante seria em vez de se estudar diretamente as alterações nucleotídicas nas seqüências codificadores do gene de interesse, estudar alterações nas seqüências de elementos reguladores do gene. Neste contexto, encontra-se o recém descrito *enhancer HACNS1* (Prabhakar et al., 2008), que atua como regulador de mais de uma dezena de genes importantes no desenvolvimento, dentre eles, *PAX9* e *ZNF423*. O *HACNS1* abrange uma região de 546 pares de base que contém 16 variações em humanos, 13 delas localizadas numa região pequena de 81pb. Estas mutações não estão presentes em outros primatas, indicando que podem ser humano-específicas. Em camundongos transgênicos, a introdução das substituições

presentes nos humanos neste *enhancer*, leva a um aumento da expressão do mesmo nas extremidades dos membros, e principalmente nas regiões dos polegares das mãos. Estes indícios funcionais levaram os autores a aventar a possibilidade de que este *enhancer* poderia estar envolvido não só nas modificações morfológicos e estruturais dentro dos primatas, mas também em características exclusivas humanas, tais como polegares opositores, habilidade manual e locomoção bípede (Prabhakar et al., 2008). Com base nisto, tem sido sugerido que essas variantes foram alvo de seleção positiva na linhagem humana, e que estariam fixadas em nossa espécie, não obstante a ausência completa de estudos em nível populacional que poderiam indicar se estas variações estariam fixadas em todas diferentes populações humanas.

Ambos genes *TCOF1* e *FGFR1*, bem como o *enhancer HACNS1* estarão presentes nos estudos conduzidos para a presente tese.

III. 2. Influência dos Processos Micro-evolutivos sobre a Diversidade Craniofacial em Nativos Americanos

Diversos estudos indicam que as diferenças craniofaciais encontradas entre populações humanas poderiam ser o produto de adaptações a condições climáticas, sugerindo que a base genética que sustenta esta diversidade não seja seletivamente neutra (Roseman, 2004; Havarti e Weaver, 2006), muito embora esta sugestão esteja ainda longe de ser um consenso. Recentemente, Relethford (2010) analisou 57 medidas obtidas para indivíduos de 22 populações ao redor do mundo e constatou que para pelo

menos alguns grupos as variações no crânio não podiam ser explicadas por um modelo neutro. Os Buriat (Sibéria) e esquimós, por exemplo, teriam modificações no crânio compatíveis com adaptação ao frio extremo. Betti *et al.* (2010) fizeram sugestões similares ao analisar mais de 6000 indivíduos de 105 populações e interpretar que processos neutros estariam mais envolvidos às mudanças na forma nos crânios, porém sinais de adaptação climática poderiam ser encontrados em populações de climas extremos, em especial para traços específicos. O índice nasal (proporção da largura pelo comprimento X100) assim como a largura da face, por exemplo, ambos vistos em crânios braquicéfalos como o dos Esquimós, seriam traços sujeitos a seleção natural visto estarem diretamente envolvido com a termoregulação (Betti *et al.*, 2010).

Independentemente se a variação craniométrica normal é o produto do equilíbrio entre mutação e deriva genética ou o resultado de seleção, é reconhecido que essa questão tem implicação relevante num tema que há muito vem despertando o interesse dos pesquisadores, incluindo de nosso grupo de investigação: povoamento pré-histórico da América (ver exemplos recentes em Wang *et al*., 2007; González-José *et al*., 2008). Estes últimos autores buscaram integrar resultados de várias áreas do conhecimento, tais como a genética, a morfologia e a lingüística. Pelo modelo proposto populações vindas da Ásia permaneceram na Beríngia por tempo suficiente para que ali surgissem alelos nativos-americanos autóctones. Essa população beringiana ancestral era caracterizada por grande variação na morfologia craniana. Após a submersão da Beríngia, houve fluxo gênico bidirecional baixo, mas constante entre os habitantes do extremo oeste do Alasca com aqueles do extremo leste da Sibéria, sendo um mito a idéia de que haveria

um isolamento reprodutivo completo entre o Velho e o Novo Mundo (González-José et al., 2008).

IV. **Gene-Cultura e sua importância na conexão genótipo-fenótipo**

Praticas culturais humanas tem alterado drasticamente as condições ambientais e de comportamentos, promovendo rápidas e marcantes mudanças em nível genotípico e fenotípico bem como em suas interconexões (dinâmica da co-evolução gene-cultura; Laland et al., 2010; Richerson et al., 2010). Um evento particularmente importante na história do *Homo sapiens sapiens*, que definitivamente foi um gatilho poderoso no estabelecimento de um novo ciclo de co-evolução gene-cultura, foi o desenvolvimento da agricultura e a domesticação de animais na era Neolítica (~10,000 anos atrás). Porém, o impacto do surgimento de civilizações, guiada pela revolução agrícola e domesticação de animais, tem sido amplamente estudada. A maioria destes estudos, no entanto, está focada nas enfermidades infecciosas e má-nutrição que ocorreram quando sociedades menos complexas foram incorporadas à civilizações mais complexas (i.e. formação do Império Inca nos Andes; conquistas Aztecas no México, etc). Detalhes relacionados as mudanças ocorridas na transição de bandos de caçadores-coletes para sociedades agriculturalistas na América, e suas conseqüências, tem sido relativamente pouco

explorados quando comparados aquela ocorrida em outros continentes (Bocquet-Appel e Bar-Yosef, 2008).

A transição para a agricultura tem sido amplamente considerada como um ponto inicial para o surgimento das civilizações e os elementos dela decorrentes (Johnson e Earle, 1987; Bocquet-Appel e Bar-Yosef, 2008), trazendo benefícios nutricionais que se refletem em mudanças em características fenotípicas tanto morfológicas quanto metabólicas. No entanto, ainda que estas vantagens sejam observadas em longo prazo após a estabilização dos meios de produção de alimento, durante um largo período o processo de transição para agricultura levou a déficit nutricionais, diminuindo os índices de saúde nas populações que passavam de estratégias caçadoras-coletoras para agriculturalistas (Blake, 2006).

Kennet et al. (2006) propuseram modelos ecológicos para a origem da produção de milho na Mesoamérica, comparando modelos ecológicos de modo de vida caçador coletor com agriculturalistas incipientes por meio da taxa de retorno em cada modo de vida. A taxa de retorno seria a relação entre o dispêndio de energia de um indivíduo para buscar comida pelo valor energético adquirido pelos alimentos obtidos. Caçadores-coletores tenderiam a maximizar o ganho de energia, selecionando fontes de alimento de bom retorno energético, tendo uma variedade maior de alimentos a disposição e alterando sazonalmente as fontes de recursos naturais. Além disto, as populações nativo-americanas caçadora-coletoras dificilmente enfrentavam período irregulares de obtenção de comida, sendo populações na maioria nômades, ou pequenos grupos, que se deslocavam em busca

de alimento, ou se estabeleciam em regiões ecológicas suficientemente abundantes para suprir o grupo.

Em contrapartida, o investimento na agricultura não era um processo direcional de ganho incondicional, como poderia ser facilmente dedutível partindo do grande sucesso obtido pelas sociedades agriculturalistas do velho e novo mundo. De acordo com Blake (2006), para que a agricultura fosse benéfica em termos de taxa de retorno energético, foi necessário um período de cerca de 4000 anos de desenvolvimento e melhoramento da agricultura do milho pelas sociedades Mesoamericanas. Estudos embasados na análise de ossos e dentes provenientes de várias sociedades pré-colombianas do México (Steckel et al. 2002), corroboram os estudos de modelos ecológicos de taxa de retorno, mostrando um stress nutricional acompanhando o surgimento das cidades e decorrente crescimento populacional (Figura 1).

Quando a comparação de índices de saúde é realizada usando ossos e dentes de sociedades caçadoras-coletoras nativos americanas que evoluíram para sociedades agriculturalistas nativo-americanas (Figura 2 e 3), o declínio nestes índices ao longo de milhares de anos é mais evidente, mostrando claramente que a mudança nos hábitos alimentares e o investimento na manutenção de uma nova dieta foi durante muito tempo desvantajoso (Steckel and Rose 2002).

Do ponto de vista evolutivo, alguns dos mais interessantes sinais de seleção natural tem sido atribuídos a adaptações à dieta em humanos. Estas adaptações a especializações na dieta são exemplos da interação gene-cultura na história

evolutiva de diferentes populações humanas. Ainda que a cultura tenha direcionado a nossa diversificação de cultivares, por meio da necessidade de implementação da agricultura e pecuária para manutenção das cidades, ela age sobre uma variabilidade genética pré-existente que habilita ou não certos indivíduos dentro de uma população não somente a diferentes dietas, mas também capazes de suportar o estresse nutricional proveniente da transição à agricultura.

Hancock et al. (2010) correlacionaram dados de variação em frequencias alelicas de alguns polimorfismos com sua eco-regiões. Populações com dietas ricas em tubérculos e cereais apresentaram freqüências mais altas de algumas variantes alélicas relacionadas em estudos de associação anteriores à metabolismo de amido e sucrose, enquanto populações polares apresentavam altas freqüências de variantes alélicas que haviam sido relacionadas com outros metabolismos energéticos.

Outro estudo importante que torna evidente o papel da cultura no processo adaptativo de diferentes grupos humanos à agricultura e a pecuária, diz respeito a possível seleção do alelo *47His* (rs12299840 do gene ADH1B (*Class I alcohol dehydrogenase*) em populações leste-asiáticas em virtude do cultivo de arroz (Peng et al. 2010) e a adaptação independente à cultura do leite de europeus e árabes, cada população como um mecanismo independente levando ao mesmo fenótipo metabólico de intolerância a lactose (Ennattah et al. 2008).

Em um trabalho recente de nosso grupo (Acuna-Alonzo et al. 2010; apêndiceI), foi descrito o primeiro alelo nativo-americano exclusivo, funcional e positivamente selecionado, o alelo *230Cys* do gene *ABCA1*. Esta variante autóctone nativo-

americana tem sido associada, em vários estudos, com sobrepeso e doenças correlacionadas (diabetes e doenças cardiovasculares) em populações ameríndias e mestiças. Indivíduos homozigotos para o alelo variante (*230Cys*) apresentarem uma taxa de efluxo de colesterol 30% menor do que indivíduos que possuem o alelo ancestral, o que indicaria que este fenótipo metabólico poderia favorecer uma maior resistência a estresses nutricionais. Os autores também encontraram sinais de seleção positiva no haplótipo onde o alelo variante é encontrado. Seria o primeiro exemplo bem fundamentado da chamada hipótese do alelo/genótipo frugal (Neel, 1962) em nativos americanos.

O estudo de Acuna-Alonzo et al. (2010), no entanto, não explorou as razões para a ocorrência da varredura adaptativa (*seletive sweep*) relacionada ao alelo variante. Tendo em vista que este caso pode representar um exemplo de como os fatores evolutivos interconectados como hábitos culturais estão envolvidos na geração e manutenção da diversidade e seu significado na conexão genótipo-fenótipo, um capítulo desta tese será dedicado a isso.

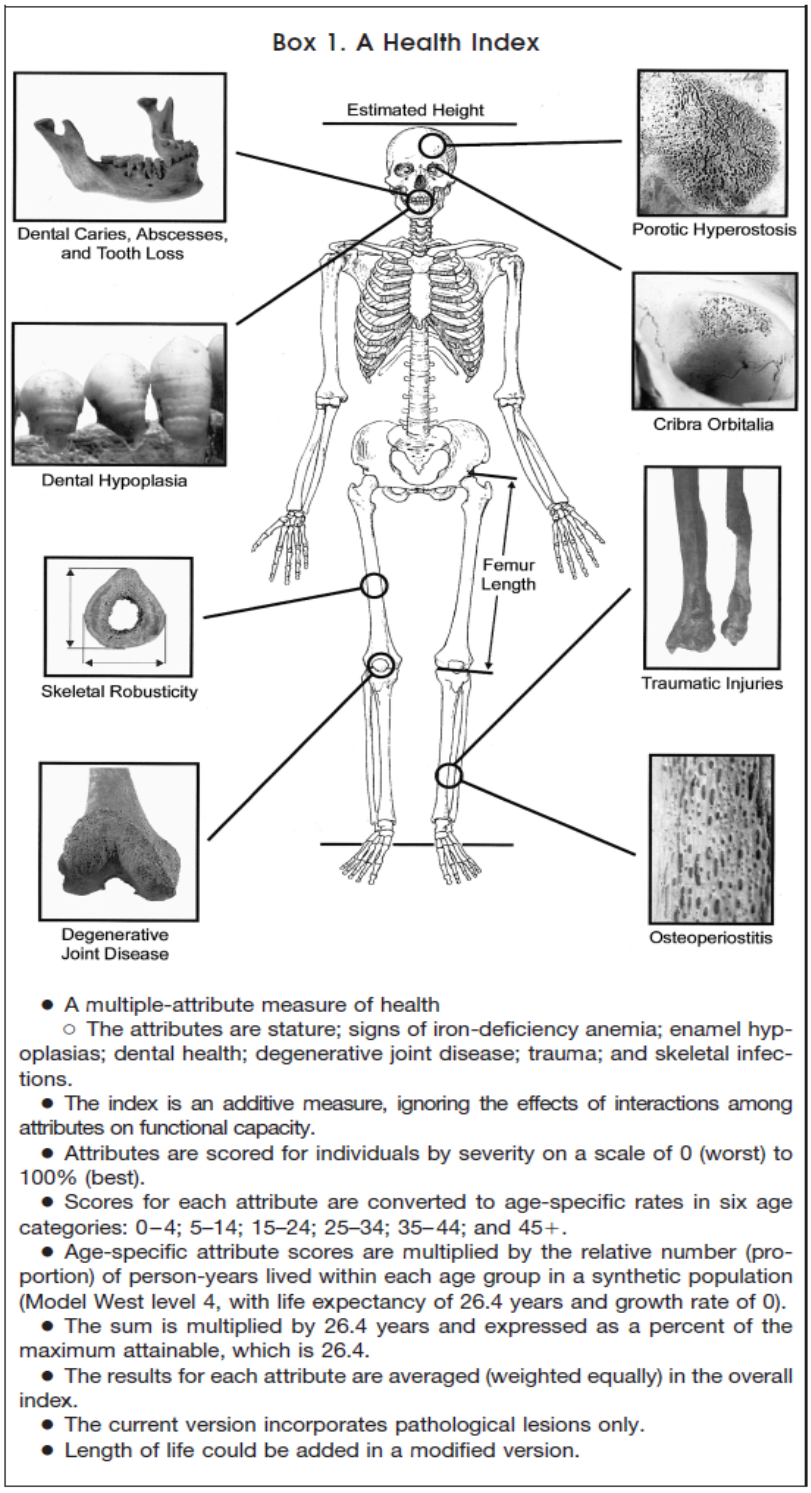


Figura 1. Índices nutricionais em cidades mexicanas pré-colombianas baseados em dados dentais e osteológicos (Steckel et al. 2002).

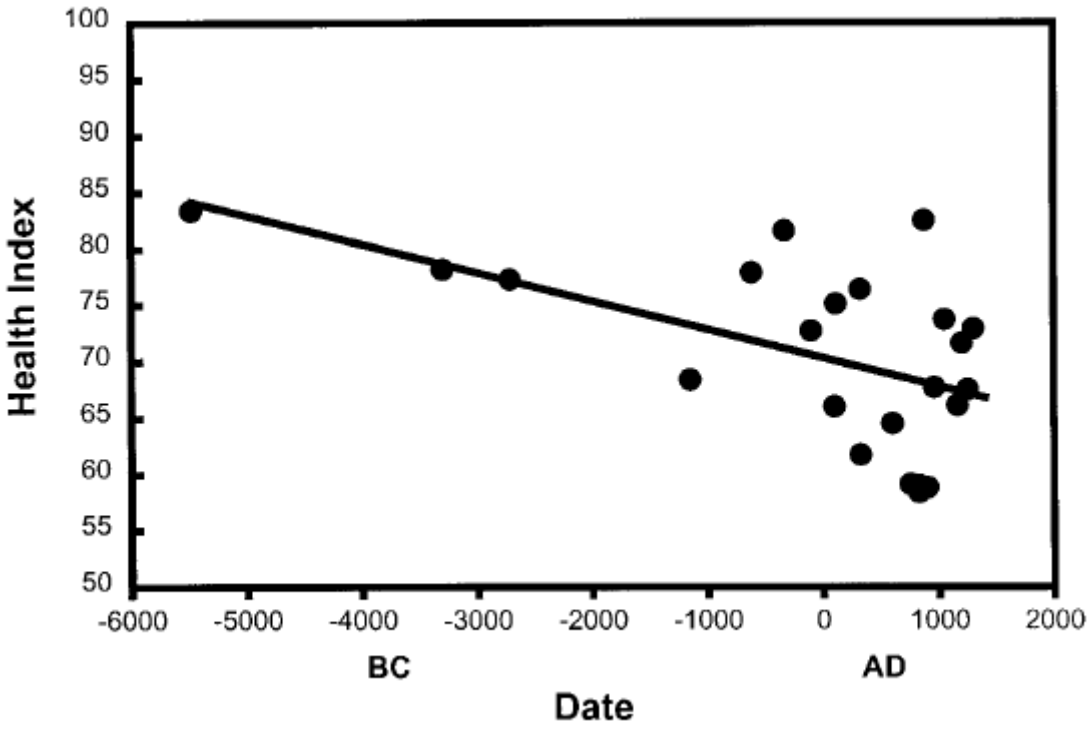


Figura 2. Tendência de diminuição do índice nutricional em populações pré-colombianas ao longo do tempo, baseado em dados osteológicos (Steckel et al. 2002)

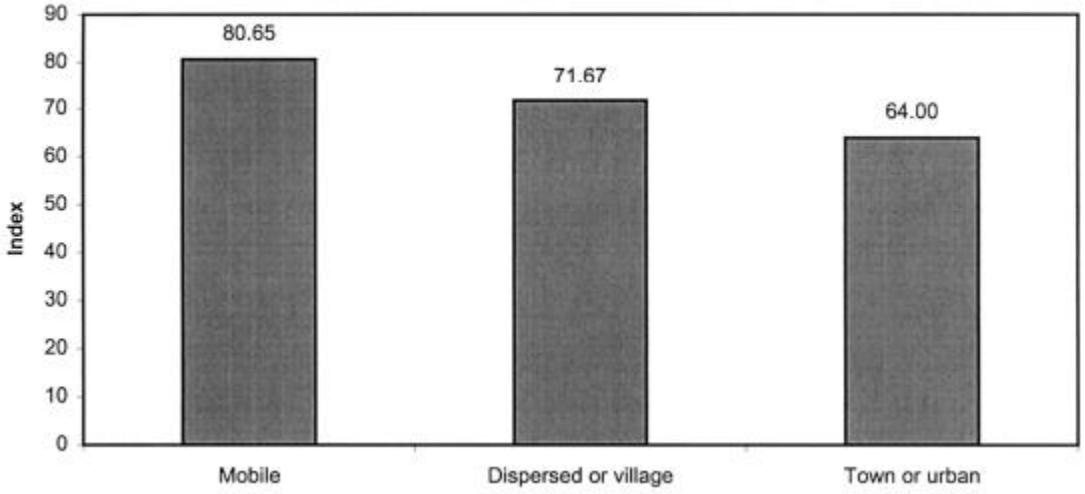


Figura 3. Índice nutricional e modo de socialização (Steckel et al. 2002).

RESULTADOS

Os resultados deste trabalho serão apresentados na forma de quatro artigos, apresentados de acordo com a seqüência de temas introduzidos acima. Adicionalmente um quinto artigo será apresentado no apêndice I.

ARTIGO I

Hünemeier T, Salzano FM, Bortolini MC. 2009. *TCOF1 T/Ser* variant and brachycephaly in dogs. *Animal Genetics*; 40(3):357-358.

BRIEF NOTES

doi:10.1111/j.1365-2052.2008.01841.x

Radiation hybrid mapping of six porcine genes of the matrix metalloproteinase family

X. Wu and Y. C. Pan

Department of Animal Science, School of Agriculture and Biology, Shanghai Jiao Tong University, 800 Don Chuan Road, Shanghai 200240, China

Accepted for publication 7 November 2008

Source/description: Proteins of the matrix metalloproteinase (MMP) family are involved in the breakdown of extracellular matrix (ECM) in normal physiological processes such as embryonic development, reproduction and tissue remodelling. *MMP9*, *MMP12*, *MMP13*, *MMP14*, *MMP15* and *MMP20* genes are members of MMP family, and the *MMP9* gene has been implicated as the major facilitator of ECM degradation.[1,2]

Primer designation, PCR condition and sequencing: Primers for porcine *MMP13*, *MMP15* and *MMP20* genes were designed according to cDNA sequences of human homologous genes (GenBank accession numbers are NM_002427, NM_002428 and NM_004771) (Table S1). The primers for *MMP9*, *MMP12* and *MMP20* genes were designed based on pig sequences from NCBI (GenBank accession numbers are NM_001038004, NC_000011 and NM_214239) (Table S1). The amplified products were obtained in 20 μl reaction volume consisting of 50 ng of porcine genomic DNA, 1x polymerase chain reaction (PCR) buffer, 0.3 μM of each primer, 75 μM of each dNTP, 2.0 mM $MgCl_2$ and 2 U *Taq* DNA polymerase (Tiangen). The PCR conditions were as follows: 95 °C for 5 min and 30 cycles of 94 °C for 20 s, 56–58 °C for 20 s and 72 °C for 20 s, followed by a further 8-min extension at 72 °C. The PCR products were purified with a PCR purification system (Tiangen) and sequenced. The identities of the PCR products amplified from genomic DNA were confirmed by sequence analysis.

Chromosomal location: The six genes were mapped (Table 1) using a whole genome porcine radiation hybrid panel (IMpRH),[3,4] and mapping results of the six genes were analysed using the IMpRH mapping tool[5] (http://imprh.toulouse.inra.fr/). The PCR typing of each gene was carried out in duplicate.

Comments: Our mapping results are consistent with the known human–pig comparative map (http://www.toulouse.inra.fr/lgc/pig/compare/SSC.htm). Locations of the six genes are reported by choosing the closest markers. In human, *MMP12*, *MMP13* and *MMP20* are parts of a cluster of MMP genes, which localize to chromosome 11q22.3. However, the porcine *MMP13* and *MMP20* genes were mapped on SSC9, while the porcine *MMP12* gene was mapped on SSC2.

Acknowledgements: The authors would like to thank Dr Martin Yerle for kindly providing IMpRH. This work was supported by the National Basic Research Program of China (2006CB102102) and National Basic Research Program of China (2004CB117502).

References

1 Fini M. E. *et al.* (1998) In: *Matrix Metalloproteinases* (Ed. by W.C. Parks & R.P. Mecham), pp. 299–356. Academic Press, New York.
2 Whiteside E. J. *et al.* (2001) *Biol Reprod* **64**, 1331–7.
3 Yerle M. *et al.* (1998) *Cytogenet Cell Genet* **82**, 182–8.
4 Milan D. *et al.* (2000) *Bioinformatics* **16**, 558–9.
5 Marklund L. *et al.* (1996) *Anim Genet* **27**, 255–69.

Correspondence: Y. C. Pan (panyuchun1963@yahoo.com.cn)

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 The primers used for isolating and mapping the six genes.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors.

doi:10.1111/j.1365-2052.2008.01838.x

TCOF1 T/Ser variant and brachycephaly in dogs

T. Hünemeier, F. M. Salzano and M. C. Bortolini

Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Caixa Postal 15053, 91501-970 Porto Alegre, RS, Brazil

Accepted for publication 12 November 2008

Background: The central goal of genetics is to understand how heritable information encoded in the genome determines the

Table 1 Radiation hybrid mapping of six porcine matrix metalloproteinase genes.

Gene	GenBank Acc. no. (porcine)	Retention (%)	LOD score	Closest marker	Dist (cR)	Chr	Human localization[1]
MMP9	FJ263936	22	6.92	*SW1031*	0.57	17q23	20q11.2–q13.1
MMP12	FJ263937	33	6.05	*ADM*	0.62	2q21	11q22.3
MMP13	FJ263938	34	5.13	*SWR1848*	0.74	9p13–p21	11q22.3
MMP14	FJ263939	29	7.72	*TCRA*	0.51	7q21	14q11–q12
MMP15	FJ263940	25	8.59	*SW2525*	0.45	6p14–p15	16q13–q21
MMP20	FJ377370	34	5.26	*SW539*	0.73	9p11–p13	11q22.3

[1]The locations of genes of the human map were obtained from http://www.ncbi.nlm.nih.gov/.

phenotype of an organism,[1] and the search for functional variants has been intense. Dogs evolved through mutually beneficial relationships with humans, and artificial selection has generated >350 distinct pure breeds, most of them over the past 250 years. Despite their recent origin, modern breeds present large differences in morphological and behavioural traits, larger than those observed among many genera of wild canids. This notable diversity contrasts with a low variation within individual breeds, suggesting that a limited number of genes are responsible for some of the phenotypic differences between breeds.

In a recent review,[2] the authors suggested that these and other characteristics make the domestic dog an excellent model for the investigation of the genetic basis of disease susceptibility, morphological variation and behavioural traits. Association studies with candidate genes have been one of the strategies frequently used to identify functional variants between breeds.[2–5] One example is the *TCOF1* gene, which functions in craniofacial development. Mutations in this gene cause Treacher Collins–Franceschetti Syndrome 1 in humans and in other primates.[6,7] Considering the possibility that some variant of this gene could be responsible for 'normal' skull/face variation in dogs, Haworth *et al.*[3] cloned and sequenced the canine homologue of the human *TCOF1* (GeneID in *Canis familiaris*: 403592 and location on CFA4 at 61.902–61.942 Mb) in several breeds. Their investigation revealed that a single C→T substitution at position 396 of *TCOF1* exon 4, which results in a p.Pro117Ser amino acid change, displayed a highly significant association with brachycephaly (broad skull/short face, as found in the Boxer, English Bulldog and Pekingese breeds).

Here we test the hypothesis that the presence of a serine (or loss of proline) at position 117 of the TCOF1 protein could be a significant predisposing factor for canine brachycephaly, as Haworth *et al.*'s result is frequently cited as an instance where a functional variant was identified.[2,8–10]

Materials and methods: Our data were obtained from a sample of 95 dogs of 16 different pedigree breeds collected at a dog clinic, which were classified independently by two trained veterinarians according to head shape as dolichocephalic (narrow skull and a long face), mesocephalic (moderate skull with medium length face) or brachycephalic (Table S1). All the dogs came from the Porto Alegre area and were genetically unrelated. Genotype *CC/ProPro*, *CT/ProSer* or *TT/SerSer* identification was obtained by RFLP using primers, the HpaII enzyme, and other methodological conditions described in Haworth *et al.*[3] In addition, DNA sequence of all brachycephalic dogs that did not present the *T/Ser* allele was performed to confirm the genotype, as well as of two dolichocephalic dogs used as controls. The MegaBace500 (Amersham Biosciences) or ABI3730 (Applied Biosystem) sequencing machines were used to obtain the results according to standard procedures.

Results and discussion: Our data reveal that *C/Pro* is found in homozygosity in several brachycephalic breeds such as Boxer, French Bulldog, Pekingese and Shih Tzu, while *T/Ser* is observed in homozygosity in dolichocephalic breeds like Dachshund (see also Fig. S1). The *T/Ser* frequencies obtained by us in the three categories of animals were: dolichocephalic, 0.28; mesocephalic, 0.21; brachycephalic, 0.32, which are not statistically different ($P > 0.05$). These results contradict the findings of Haworth *et al.*[3] The reasons for the discrepancy may be that only five brachychephalic animals had been tested by them. Results based on small samples and several comparisons can easily show associations simply due to chance. Note also that although dog breeds are genetically quite homogeneous, the three French Bulldogs we tested were *CC/ProPro* homozygotes, not a *TT/SerSer* homozygote, as found in the French Bulldog tested by Howarth *et al.*[3] It is probably that other genetic and/or epigenetic factors also act to define face shapes in dogs. Dog breeds show clear diversity in morphology, but the extent to which they differ in their genomes as a whole remains to be established. Any association study involving them should be carefully evaluated and subjected to replication.

Acknowledgements: Special thanks are due to veterinarians Joseane Salve and Márcia Regina Cordeiro, who enabled this study to be carried out. We would also like to thank Drs Claiton D. Bau, Eliane Bandinelli and Sídia M. Callegari-Jacques for constructive comments. Our research is financed by the Institutos do Milênio and Apoio a Núcleo de Excelência Programs, Conselho Nacional de Desenvolvimento Científico e Tecnológico and Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul.

References

1 Lee I. *et al.* (2008) *Nat Genet* **40**, 181–8.
2 Wayne R. K. & Ostrander E. A. (2007) *Trends Genet* **23**, 557–67.
3 Haworth K. E. *et al.* (2001) *Mamm Genome* **12**, 622–9.
4 Haworth K. E. *et al.* (2007) *DNA Seq* **18**, 209–19.
5 Lark K. G. *et al.* (2006) *Trends Genet* **22**, 537–44.
6 The Treacher Collins Collaborative Group (1996) *Nat Genet* **12**, 130–6.
7 Shows K. H. *et al.* (2006) *Mamm Genome* **17**, 168–77.
8 Coussens A. K. & van Daal A. (2005) *Genomics* **85**, 563–73.
9 Michell A. R. (2005) *Vet J* **170**, 153–62.
10 Ostrander E. A. & Wayne R. K. (2005) *Genome Res* **15**, 1706–16.

Correspondence: M. C. Bortolini (maria.bortolini@ufrgs.br)

Supporting information

Additional supporting information may be found in the online version of this article.

Figure S1 *HpaII* restriction digest fragments of c.396C>T *TCOF1* exon 4 segment from different dog breeds: (a) Boxer, (b) French Bulldog, (c) Boxer, (d) Dachshund, (e) Labrador, (f) Dachshund, (g) Shih Tzu, (h) Collie and (i) Shi Tzu.

Table S1 Genotype and allele distributions of the c.396C>T (p.Pro117Ser) polymorphism in dog breeds with different head shapes.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors.

SUPLEMENTARY DATA

Figure S1. *HpaII* restriction digest fragments of c.396C>T *TCOF1* exon 4 segment from different dog breeds: a) Boxer, B) French Bulldog, c) Boxer, d) Dachshund, e) Labrador, f) Dachshund, g) Shih Tzu, h) Collie, and i) Shi Tzu.

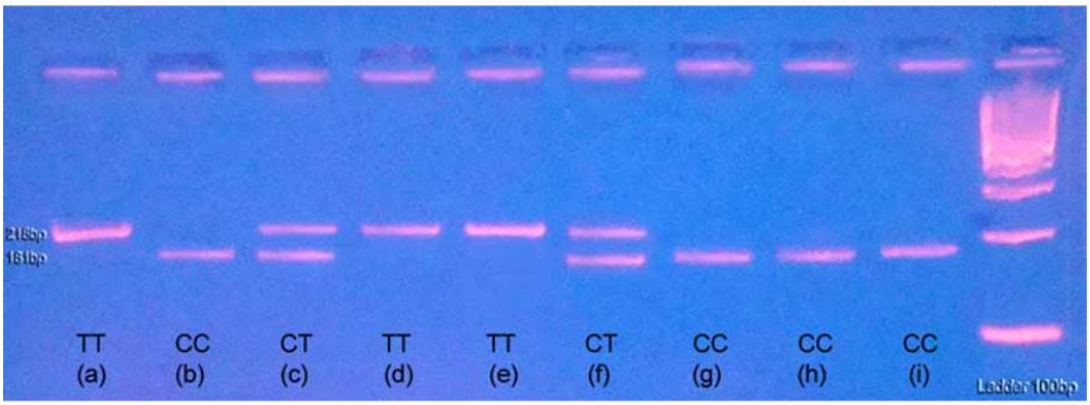


Table S1. Genotype and allele distributions of the c.396C>T (p.Pro117Ser) polymorphism in dog breeds with different head shapes

	Dog breed	n[1]	Head shape[1]	Genotype			Allele (%)	
				CC	CT	TT	C	T
1	Boxer	17	B	4	5	8		
2	French Bulldog	3	B	3	0	0		
3	Bulldog	5	B	2	3	0		
4	Pequingese	5	B	4	1	0		
5	Shi-Tzu	11	B	9	2	0		
	Total brachycephalic	41		22	11	8	67	33
6	Akita	9	M	8	1	0		
7	Bichon Frise	1	M	1	0	0		
8	Dalmatian	3	M	2	1	0		
9	English Sheepdog	1	M	1	0	0		
10	Labrador Retriever	7	M	4	2	1		
11	Rottweiler	6	M	4	2	0		
12	Mini Schnauzer	4	M	0	3	1		
	Total mesocephalic	31		20	9	2	79	21
13	Bull Terrier	1	D	0	1	0		
14	Collie	4	D	4	0	0		
15	Daschund	17	D	7	8	2		
16	Whippet	1	D	1	0	0		
	Total dolichocephalic	23		12	9	2	72	28

[1]n = number of animals tested: B=brachycephalic, M=mesocephalic, D=dolichocephalic.

ARTIGO II

FGFR1 GENE HAPLOTYPE TAG-SNPS, LINKAGE DISEQUILIBRIUM PATTERNS, AND THEIR INFLUENCE IN NORMAL CRANIOFACIAL VARIATION

Hünemeier T[1*], Gomez-Valdez J[2*], Contini V[1], Macin G[3], Ballesteros M[3], Acuña-Alonzo V[3,4], Corral P[6], Sánchez-Mejorada G[2], Salzano FM[1], Ruiz-Linares A[7], Canizales-Quinteros S[2,5] and Bortolini MC[1]

[1] Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

[2] Universidad Nacional Autonoma de Mexico, Mexico City, Mexico

[3] Escuela Nacional de Antropologia e Historia, Mexico City, Mexico

[4]Instituto Nacional de Medicina Genómica, Mexico City, Mexico

[5]Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico

[6]Universidad Pompeu Fabra, Barcelona, Spain

[7] University College London, London, UK

*These authors contributed equally to this work

Manuscrito em preparação (a ser submetido para a revista *Human Heredity*)

Introduction

Nei (1987, 2007) proposed the so-called "Major Gene Effect Hypothesis", where morphological evolution would occur by the action of a small number of mutations of large effect on structure (s) and / or regulatory (s) gene (s). However, most of the normal morphological variation can be attributable to several genes, as well as to other mechanical, environmental, and epigenetic factors. All these elements have been related to normal vertebrate craniometric variation (Kim et al., 1998, Lawson et al., 2006; Richsmeier et al., 2007). Additionally, genes would have a significant redundant effect, since mutations in different genes can often lead to similar phenotypes (Kim et al., 1998), making even more complicated the genotype-phenotype connections.

Several genes are involved in human craniofacial morphology (Enlow, 1990; Roseman, 2004; Coussens and Dall, 2005), among them the *Fibroblast Growth Factor Receptors (FGFR)*genes, with an important role in various stages of craniofacial development (Helms et al., 2005). For example, *FGFR1* is crucial in skull development, while *FGFR3* seems to be associated with bones growth at the skull base. Many medical studies have been conducted with both *FGFR1* and *FGFR3*, since mutations in them lead to phenotypes with severe craniofacial anomalies (Wilkie et al., 1997, Rice et al., 2003). A large number of mutations have been identified in pathologic cases, suture closure/fusion patterns heterogeneity, leading to a wide range of facial morphologic differences.

To test the influence of these suture-closure/fusion processes in normal craniofacial variation, Coussens and Dall (2005) developed one *FGFR1* study in healthy human

populations. They sequenced the entire gene in African-American, Asian, Caucasian and Australian Aborigine individuals. Forty four of them had their craniofacial traits also measured. A total of 17 SNPs were identified, but only eight of them showed high variation between the four investigated populations. Nine common haplotypes could be inferred from these eight SNPs, three being the most common in all populations. Coussens and Dall (2005) also identified two haplotype tag SNPs (htSNPs) (rs4647905 and rs3213849) responsible for > 85% of the diversity in each population. As a consequence, those two *htSNPs* could be very useful for population and association studies. In addition, a negative correlation between rs4647905 and the cephalic index (CI), one of the classical measures of head shape, was found, suggesting that the variant allele (C) would be associated with a decrease in this index (Coussens and Dall, 2005).

It is known that a significant portion of the differences in human physical appearance, susceptibility to diseases, and other traits found within and between human populations can be attributable to genetic variants (Bamshad and Wooding, 2003). For Instance, allele frequencies can be significantly different between Asians and Native Americans, and variants which exist in Asians may be absent in Native Americans or vice-versa (see examples of private Native American alleles in Rosenberg et al., 2009; Acuña-Alonzo et al., 2010). This sort of information is fundamental to help detect signs of association between one allele/haplotype and a specific phenotype, since it is expected that part of these alleles are functional. Furthermore, associations or linkage disequilibrium (LD) levels found in one population cannot be extrapoled to others. Africans generally have lower levels of LD (Lonjon et al. 2003), whereas Native

Americans have high LD values. This happens for several reasons, and is related to the evolutionary histories of each population (Amorim et al. 2010).

In this context, we performed a study to evaluate the distribution of the *FGFR1* haplotypes in a large healthy native sample (N= 333) from different continents (Africa, Asia, Europe, and America). Additionally, Mexican mestizos were also investigated. Several craniofacial parameters were obtained in a subset of our Native American and Mexican sample to detect possible associations between *FGFR1* polymorphisms and craniofacial indexes.

Subjects and Methods

Population Samples

For this purpose, craniofacial data were obtained from 136 individuals belonging to three groups: Mexican mestizos from the city of Tepango and Mexico city (Central Mexicans), and Totonaco, a Native American people from Sierra de Puebla, Mexico. We also collected blood from each individual to extract DNA via standard methods. Additionally 107 Native Americans from five populations (Kayapo, Kaingang, Xavante, Yanomama and Baniwa), 39 individuals from two European populations, 26 South Saharan Africans and 25 Siberian Eskimos were studied at the DNA level, without craniofacial information. All the subjects gave informed consent to participate in the study. Ethical approval for the present study was provided by the Brazilian National Ethics Commission (CONEP; resolution number: 1333/2002) and by ethics committees in the countries where the non-Brazilian samples were collected.

Craniofacial indices

Craniofacial indices (Table 1 and 2) were obtained from the 136 individuals. Using sliding and spreading calipers nine direct craniofacial metrical measures were performed; linear lengths, widths and heights were achieved considering two major components (neural and facial). In addition, based on these values we calculated eight standard craniofacial indices. Indices values were grouped into categories (Comas, 1976). In order to minimize any accumulative error the anthropometric technique was made following the international agreements and by only one observed (JAGV).

***FGFR1* polymorphisms**

Genotyping for the five *FGFR1* polymorphisms was performed using TaqMan assays (ABI Prism 7900HT Sequence Detection System; Applied Biosystems); the two htSNPs (rs4647905 and rs3213849) that represent >85% of the *FGFR1* gene haplotype variability, plus three other SNPs (rs2304000, rs2293971, and rs930828) were investigated to evaluate population diversity.Haplotype phases are inferred with Beagle 3.2.1 (Browning and Browning 2007), and LD analysis was performed with Haploview software (Barret et al. 2005).

Data Analysis

The statistical analyses were carried out in three stages; first we cross-tabulated the observed and expected frequencies for the total number of cases classified by indices considering *FGFR1* genotypes for each population. Association between common htSNPs and craniofacial phenotypes was tested using Chi-square tests separately by population data. Secondly bar graphs representing population mean indices values were obtained and the Shapiro-Wilk's Goodness of Fit test calculated us to determine whether or not the variables a normal distribution of the variables. The minimum, maximum, mean, standard deviation, skewness, and kurtosis of each measurement were calculated grouped according to *FGFR1* genotype, considering the different populations.

As was done by Coussens and van Daal (2005), we normalized (using Blom's transformation) and ranked the indices values to obtain new variables for the linear regression analysis. The indices were ranked using sequential ranks for unique values procedure implemented in the SPSS software to assume that the critical assumptions underlying conventional regression models were not violated. To test the proportion of variability due to the htSNPs, the coefficient of determination R^2 was used; whereas *F* statistics determined the likelihood ratio of the explained variance by the residual sum of squares.

A set of categorical and continuous multivariate analyses were also preformed. Multiple correspondence (MC) analysis was applied to analyze the pattern of

relationships of several categorical dependent variables at the same time (htSNPs and indices classifications); considering all the possible Chi-square combinations. This procedure can be seen as a generalization of the principal component analysis when a Chi-square matrix is used instead of correlation or variance-covariance matrices.

A Principal Component Analysis (PCA) was also performed to reduce the metric data to a smaller number of dimensions (Figure 1). Four PCs were obtained reducing 50% of the original variables with cumulative variance > 80%. The PC scores were obtained and plotted separately; bars represent the contribution of every craniofacial index to each PC. Principal component values were grouped according to genotype and were graphically illustrated through box-plots. Finally, to analyze PCs and htSNPs correlations we computed Spearman's *rho* coefficient between all possible pairs including the original indices. Sex differences were considered by independent-sample t-tests. In all cases the significance was placed at $\alpha = 0.05$. All the statistical analysis was carried out using the SPSS (v. 17.0) software (http://www.spss.com/media/collateral/statistics/ProgDataMgmt_SPSS17.pdf).

Results and Discussion

Regarding the values observed using Chi-square and the Shapiro-Wilk's Goodness of Fit tests (Table 3), Allele C of the htSNP rs4647905 was associated with a decrease of the cephalic index (CI) in the Tepango population (χ^2 test with p=0.029), and with the jugo-frontal (JFI) index in the Central Mexican population (p=0.022);

htSNP rs3213849 allele T is associated with the prosoptic (PI) index in Tepango (χ^2 test, p=0.022) and with the jugo-mandibular (JMI) index (χ^2 test, p=0.030) in Central Mexicans. Additional results based on MC analysis (data not shown) corroborated the association of JFI and JMI with the rs3213849 *FGFR1*.

There is a reduction in cephalic index in individuals carrying rs4647905 CC in Central Mexicans and Totonaco, (Figure 2 and Figure 3), although the values are not significant. The coefficient of determination indicated that 16.7% of the variance in CI was explainable by *FGFR1* rs4647905 variant. For the prosopic and jugo-frontal indices, the coefficient of determination estimated that 23% and 25% of the variances, respectively could be explained by *FGFR1* rs3213849 SNP. The regression analysis for the three correlations tested furnished a probability of $p < 0.001$. There was no significant correlation among PCs and htSNPs. PC1 shows that 28% of the total head size variability is related with JFI and JMI, whereas PC2 (23%) explains cephalic indexes. The residual variation is relative to nasal and nasofacial indexes (PC3 = 16%), and PI (PC4 = 13%). These results indicate that PC1 is correlated with facial components, whereas PC2 explains the variability of the neural size (Figure 4, Table 5).

Coussens and Daal (2005) reported an association between CI and the rs4647905 htSNP (allele C would be associated with a decrease in this CI), and they also found no differences in LD pattern among the populations studied. Our results point to a different way. When Europeans, Native Americans, Eskimos, Africans and admixed populations are compared for LD patterns, the two LD blocks found by previous by the two above-indicated authors are present in Europeans and Central Mexicans only. South Amerindians have a higher LD level, Tepango Mestizos and Totonacos present just one

LD block, while Africans showed the lowest the linkage disequilibrium (Figure 5). Haplotype distributions are found in Table 5.

The two htSNPs showed some LD degree in all three phenotyped populations. They are physically located at the gene extremes, rs4647905 in intron 14 (3'region) and rs3213849 at 5'UTR; there seems to have been recombination event(s) inside the gene (Coussens and Daals 2005), leading the two blocks to be inherited independently. PC analysis, as well as association tests, seem to show a slightly different pattern, with rs3213849 htSNP associating with upper facial features, and rs4617905 htSNP with cranial shape (due to its association with CI) (Figure 4). This pattern is seen despite the LD differences among populations. When other populations, such as South Amerindians are considered, the Xavante, for example, who present high percentages of hyperdolicocephalic and dolicocephalic individuals, characterized by extreme and low CI, respectively (Rodriguez-Delfin et al. 1997), did not present a single rs4617905 C allele (Table 5). This finding could be interpreted as an LD effect of the alleles in the blocks involved in that morphological feature. The rs46470905 C allele could be not directly responsible for the cephalic index variation, but it could be in LD with other(s) variant(s) that would lead to a decrease in this index.

FGFR1 is involved in a complex gene network that compresses several *FGFs* genes and others important genes related to expression in bone junction (Figure 6). Each *FGFRs* binds or is activated by a unique subset of *FGFs*, the specificity of which is further regulated by the alternative splicing of *FGFRs* genes (Itoh et al. 2004); therefore

SNPs in a non-coding region (i.e. htSNPs rs4647905 and rs 3213849) could per se or by *hitchhiking* contribute to phenotype differentiation.

Kidd et al (2000) demonstrated, working with a *Phenylalanine Hydroxylase Locus* (*PHA*), that the physical extent of linkage disequilibrium can differ substantially among populations from different regions of the world, due to both ancient genetic drift in the common ancestor or to recent genetic drift affecting individual populations. They also inferred that haplotypes with different *PHA* mutations might be involved with selection (Kidd et al., 1987) or, what seems more probable based on LD studies, with stochastic aspects of mutation and random genetic drift. As morphological features are always in the top discussions of selection *versus* drift as a process to generate and maintain diversity (Kimura et al, 2009; Betti et al. 2010), it is important to evaluate all possibilities of genetic interaction that would contribute to this variation, as well to try to identify the biological patterns involved.

Summarizing, an understanding of the evolutionary histories of *FGFR1* haplotypes in healthy populations may be important in determine whether their distribution might be influenced by selection or neutral processes.

Table 1. Measurements used in this study.

Code	Measurement	Description
GOL	Head length (g-op)	Glabella opisthocranium length
XCB	Head breadth (eu-eu)	Eurion eurion breadth
VTH	Head height (v-t)	Vertex tragion height
MFB	Minimum frontal breadth (ft-ft)	Right - left frontotemporal breadth
ZYB	Bizygomatic breadth (zy-zy)	Right - left zygion breadth
GOB	Bigonial breadth (go-go)	Right - left gonion breadth
NLH	Nose length (n-sn)	Nasal subnasal length
NLB	Nose breadth (al-al)	Right - left alare
NGH	Facial length (n-gn)	Nasion gnation length

Table 2. Indices used in this study

Code	Index		
CI	Cephalic index	a	XCB / GOL * 100
TI	Transverse frontoparietal index	b	MFB / XCB * 100
PI	Prosopic index	c	NGH / ZYB * 100
JMI	Jugomandibular index	d	GOB / ZYB * 100
JFI	Jugofrontal index	e	MFB / ZYB * 100
CFI	Transverse cephalofacial index		ZYB / XCB * 100
NI	Nasal index	f	NLB / NLH * 100
NFI	Nasofacial index	g	NLH / NGH * 100

Indices classification

a Dolichocepahlic (75.9 or less), Mesocephalic (76 to 80.9), Brachycephalic (81 or greater)

b Estenometopic (68.9 or less), Metriometopic (69 to 70.9), Eurimetopic (71 or greater)

c Euriprosopic (83.9 or less), Mesoprosopic (84 to 87.9), Leptoprosopic (88 or greater)

d Narrow (74.9 or less), Medium (75 to 79.9), Broad (80 or greater)

e Narrow (74.9 or less), Medium (75 to 79.9), Broad (80 or greater)

f Leptorrhine (69.9 or less), Mesorrhine (70 to 84.9), Platyrrhine (85 or greater)

g Short (37.9 or less), Intermediate (38 to 45.9), Long (46 or greater)

Table 3. Shapiro-Willk's Goodness to Fit test result.

Ethnicity		Shapiro-Willk's Goodness-Fit test							
			rs4647905				rs3213849		
			Shapiro-Wilk	df	Sig.		Shapiro-Wilk	df	Sig.
Mexican	Cephalic index	GG	0,963	24	0,512	TT			
		GC	0,936	13	0,411	TC	0,947	13	0,556
		CC	0,862	4	0,267	CC	0,946	26	0,191
	Transverse frontoparietal index	GG	0,920	24	0,058	TT			
		GC	0,933	13	0,377	TC	0,942	13	0,486
		CC	0,868	4	0,291	CC	0,948	26	0,212
	Prosopic index	GG	0,951	24	0,284	TT			
		GC	0,956	13	0,692	TC	0,942	13	0,480
		CC	0,899	4	0,426	CC	0,972	26	0,679
	Jugomandibular index	GG	0,973	24	0,730	TT			
		GC	0,834	13	***0,018***	TC	0,928	13	0,324
		CC	0,977	4	0,882	CC	0,968	26	0,566
	Jugofrontal index	GG	0,944	24	0,200	TT			
		GC	0,926	13	0,298	TC	0,949	13	0,578
		CC	0,890	4	0,383	CC	0,923	26	0,052
	Transverse cephalofacial index	GG	0,924	24	0,071	TT			
		GC	0,956	13	0,684	TC	0,881	13	0,074
		CC	1,000	4	1,000	CC	0,949	26	0,216
	Nasal index	GG	0,974	24	0,756	TT			
		GC	0,951	13	0,620	TC	0,961	13	0,773
		CC	0,868	4	0,292	CC	0,957	26	0,339
	Nasofacial index	GG	0,978	24	0,848	TT			
		GC	0,912	13	0,196	TC	0,871	13	0,054
		CC	0,860	4	0,262	CC	0,953	26	0,270
Mestizo (Tepango)	Cephalic index	GG	0,906	5	0,447	TT			
		GC	0,960	5	0,806	TC			
		CC				CC	0,971	10	0,899
	Transverse frontoparietal index	GG	0,842	5	0,170	TT			
		GC	0,974	5	0,899	TC			
		CC				CC	0,875	10	0,113
	Prosopic index	GG	0,869	5	0,262	TT			
		GC	0,978	5	0,922	TC			
		CC				CC	0,887	10	0,157
	Jugomandibular index	GG	0,944	5	0,691	TT			
		GC	0,988	5	0,974	TC			
		CC				CC	0,954	10	0,713
	Jugofrontal index	GG	0,954	5	0,765	TT			
		GC	0,774	5	***0,049***	TC			
		CC				CC	0,956	10	0,743
	Transverse cephalofacial index	GG	0,924	5	0,557	TT			
		GC	0,960	5	0,807	TC			
		CC				CC	0,958	10	0,766
	Nasal index	GG	0,932	5	0,608	TT			
		GC	0,902	5	0,419	TC			
		CC				CC	0,955	10	0,724
	Nasofacial index	GG	0,806	5	0,090	TT			
		GC	0,823	5	0,123	TC			
		CC				CC	0,932	10	0,465
Totonaco (Tepango)	Cephalic index	GG	0,977	39	0,605	TT			
		GC	0,984	37	0,850	TC	0,833	8	***0,064***
		CC	0,899	7	0,327	CC	0,986	75	0,552
	Transverse frontoparietal index	GG	0,964	39	0,247	TT			
		GC	0,950	37	0,093	TC	0,932	8	0,535
		CC	0,969	7	0,892	CC	0,970	75	***0,071***
	Prosopic index	GG	0,948	39	0,069	TT			
		GC	0,974	37	0,521	TC	0,830	8	0,060
		CC	0,975	7	0,932	CC	0,983	75	0,392
	Jugomandibular index	GG	0,943	39	***0,047***	TT			
		GC	0,980	37	0,725	TC	0,949	8	0,698
		CC	0,802	7	***0,043***	CC	0,959	75	***0,015***
	Jugofrontal index	GG	0,973	39	0,450	TT			
		GC	0,959	37	0,184	TC	0,983	8	0,977
		CC	0,862	7	0,156	CC	0,950	75	***0,005***
	Transverse cephalofacial index	GG	0,974	39	0,509	TT			
		GC	0,989	37	0,973	TC	0,935	8	0,560
		CC	0,970	7	0,897	CC	0,988	75	0,729
	Nasal index	GG	0,959	39	0,172	TT			
		GC	0,973	37	0,499	TC	0,922	8	0,445
		CC	0,920	7	0,466	CC	0,986	75	0,552
	Nasofacial index	GG	0,954	39	0,115	TT			
		GC	0,949	37	0,092	TC	0,923	8	0,458
		CC	0,918	7	0,455	CC	0,961	75	***0,020***

Ho: Normality trend
Significance p<0.05
Target cells no sufficient degree of freedom

Table 4. Haplotype Frequencies distribution among populations.

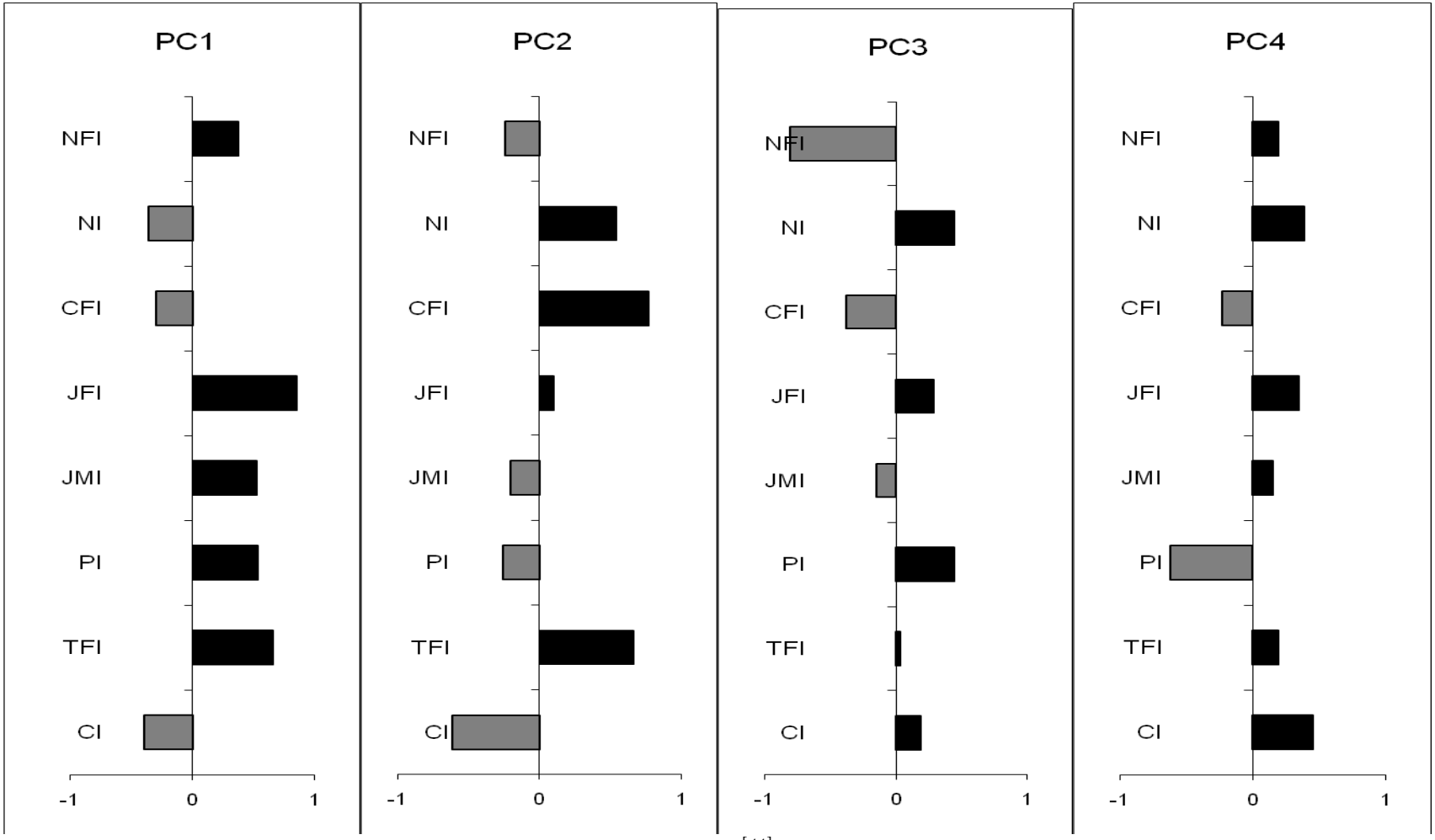
ID	Haplotype	Haplotype Frequency	Haplotype Frequency per Population										
						Native Americans						Mexican Admixed	
		N=666	African (52)	Eskimo (50)	European (78)	Kayapo (44)	Xavante (38)	Baniwa (40)	Kaingang (48)	Yanomama (44)	Totonaco (166)	Tepango (24)	Mexican (82)
h1	GCGCA	0.555556 (370)	0.673 (35)	0.66 (33)	0.538 (42)	0.659 (29)	0.816 (31)	0.675 (27)	0.646 (31)	0.568 (25)	0.47 (78)	0.458 (11)	0.341 (28)
h2	CCGCA	0.012012 (08)	0.0769 (04)	0.04 (02)	0.0256 (02)	0	0	0	0	0	0	0	0
h3	GCGTG	0.177177 (118)	0.173 (09)	0.18 (09)	0.154 (12)	0.0227 (01)	0.105 (04)	0.2 (08)	0.0833 (04)	0.182 (08)	0.181 (30)	0.0417 (01)	0.39 (32)
h4	CGACA	0.171171 (114)	0.0385 (02)	0.04 (02)	0.179 (14)	0.273 (12)	0	0.05 (02)	0.271 (13)	0.227 (10)	0.205 (34)	0.292 (07)	0.22 (18)
h5	GGGCA	0.003003 (02)	0.0385 (02)	0	0	0	0	0	0	0	0	0	0
h6	GCGTA	0.027027 (18)	0	0.04 (02)	0	0	0	0.05 (02)	0	0.0227 (01)	0.0542 (09)	0.167 (04)	0
h7	GCGCG	0.009009 (06)	0	0.04 (02)	0	0	0.0789 (03)	0.025 (01)	0	0	0	0	0
h8	GGACA	0.012012 (08)	0	0	0.0513 (04)	0.0455 (02)	0	0	0	0	0.012 (02)	0	0
h9	CCGTG	0.006006 (04)	0	0	0.0513 (04)	0	0	0	0	0	0	0	0
h10	CGGCA	0.006006 (04)	0	0	0	0	0	0	0	0	0	0	0.0488 (04)
h11	CGATA	0.010511 (07)	0	0	0	0	0	0	0	0	0.0422 (07)	0	0
h12	CGATG	0.003003 (02)	0	0	0	0	0	0	0	0	0.012 (02)	0	0
h13	CGGTG	0.003003 (02)	0	0	0	0	0	0	0	0	0.00602 (01)	0.0417 (01)	0
h14	GCACA	0.001502 (01)	0	0	0	0	0	0	0	0	0.00602 (01)	0	0
h15	CCACA	0.003003 (02)	0	0	0	0	0	0	0	0	0.012 (02)	0	0

Table 5. Spearman’s Rank Correlation Coefficients

Spearman's rank correlation matrix															
	rs4647905	rs3213849	PC1	PC2	PC3	PC4	CI	TI	PI	JMI	JFI	CFI	NI	NFI
rs4647905		0,136	-0,004	0,030	-0,064	0,041	-0,040	0,001	-0,086	-0,024	-0,008	-0,015	0,000	0,071
rs3213849	0,114		-0,085	0,046	0,072	0,078	0,113	-0,003	-0,038	0,075	-0,036	0,093	0,134	-0,128
PC1	0,960	0,327		0,039	0,028	-0,128	-0,412	0,660	0,547	0,396	0,818	-0,225	-0,368	0,246
PC2	0,727	0,594	0,653		0,037	0,012	-0,592	0,621	-0,230	-0,140	0,125	0,747	0,511	-0,245
PC3	0,459	0,406	0,748	0,667		0,029	0,199	0,048	0,374	-0,153	0,314	-0,316	0,462	-0,817
PC4	0,639	0,368	0,136	0,894	0,735		0,450	0,136	-0,636	0,102	0,258	-0,188	0,434	0,117
CI	0,646	0,189	***0,000***	**0,000**	***0,020***	***0,000***		-0,422	-0,224	-0,074	-0,150	-0,367	-0,031	-0,116
TI	0,987	0,975	***0,000***	**0,000**	0,578	0,115	***0,000***		0,104	0,081	0,751	0,320	0,036	0,033
PI	0,319	0,662	***0,000***	**0,007**	***0,000***	***0,000***	***0,009***	0,226		0,156	0,322	-0,311	-0,335	-0,151
JMI	0,784	0,384	***0,000***	0,104	0,076	0,240	0,389	0,348	0,071		0,132	-0,110	-0,086	0,158
JFI	0,929	0,676	***0,000***	0,148	***0,000***	***0,002***	0,081	***0,000***	***0,000***	0,126		-0,310	-0,063	0,026
CFI	0,867	0,282	***0,008***	**0,000**	***0,000***	***0,028***	***0,000***	***0,000***	***0,000***	0,200	***0,000***		0,194	-0,088
NI	0,998	0,121	***0,000***	**0,000**	***0,000***	***0,000***	0,721	0,679	***0,000***	0,319	0,468	***0,024***		-0,477
NFI	0,412	0,139	***0,004***	**0,004**	***0,000***	0,174	0,177	0,706	0,079	0,066	0,761	0,310	***0,000***	

Spearman's Rho values in the uper-right
P values in the lower-left

Figure 1. Eigenvectors in Principal Components



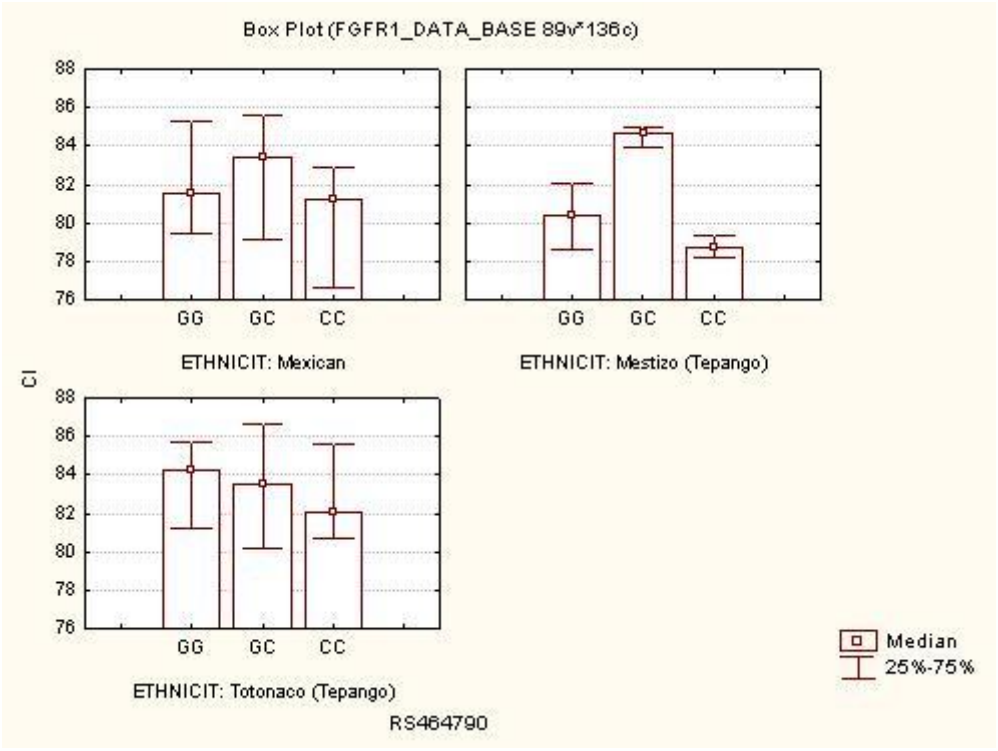


Figure 2. The distribution of mean cephalic index (CI) for rs4647905G→C genotypes within three populations. The reducing effect of the CC genotype on mean CI can be observed for each population.

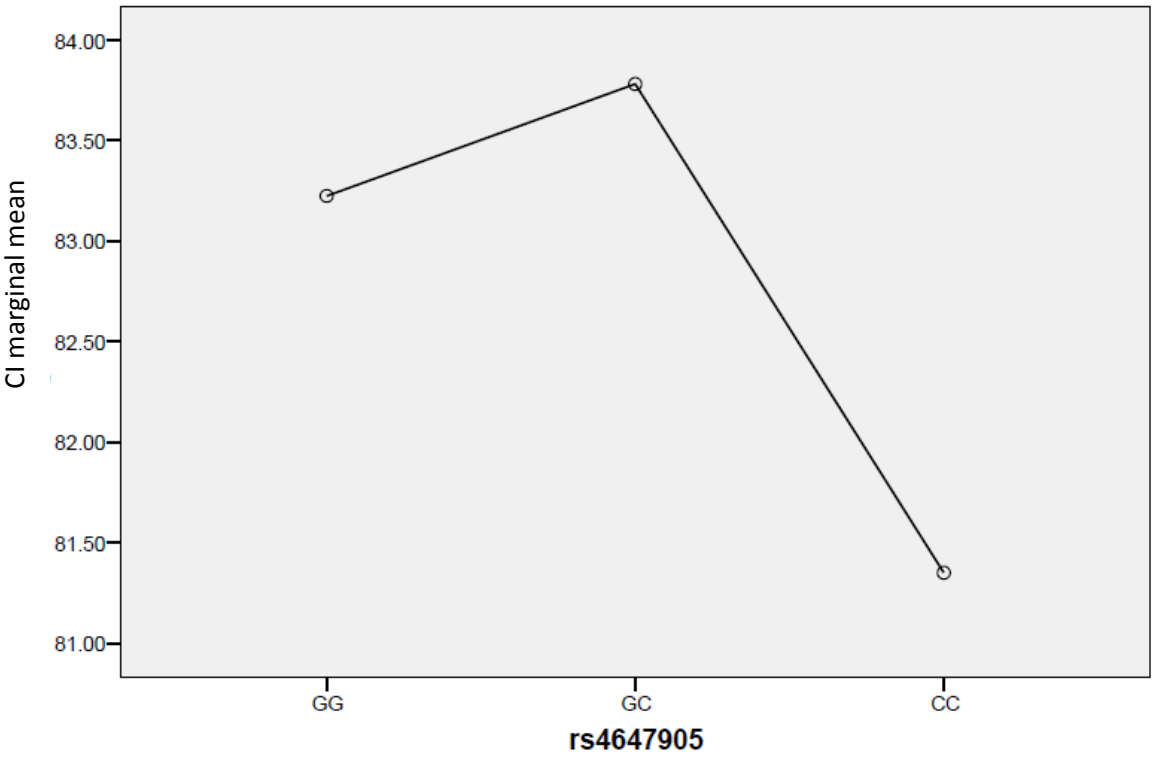


Figure 3. Cephalium Index Mean *vs* genotypes considering all individuals as one population. A reduction in CI is clear in CC genotypes.

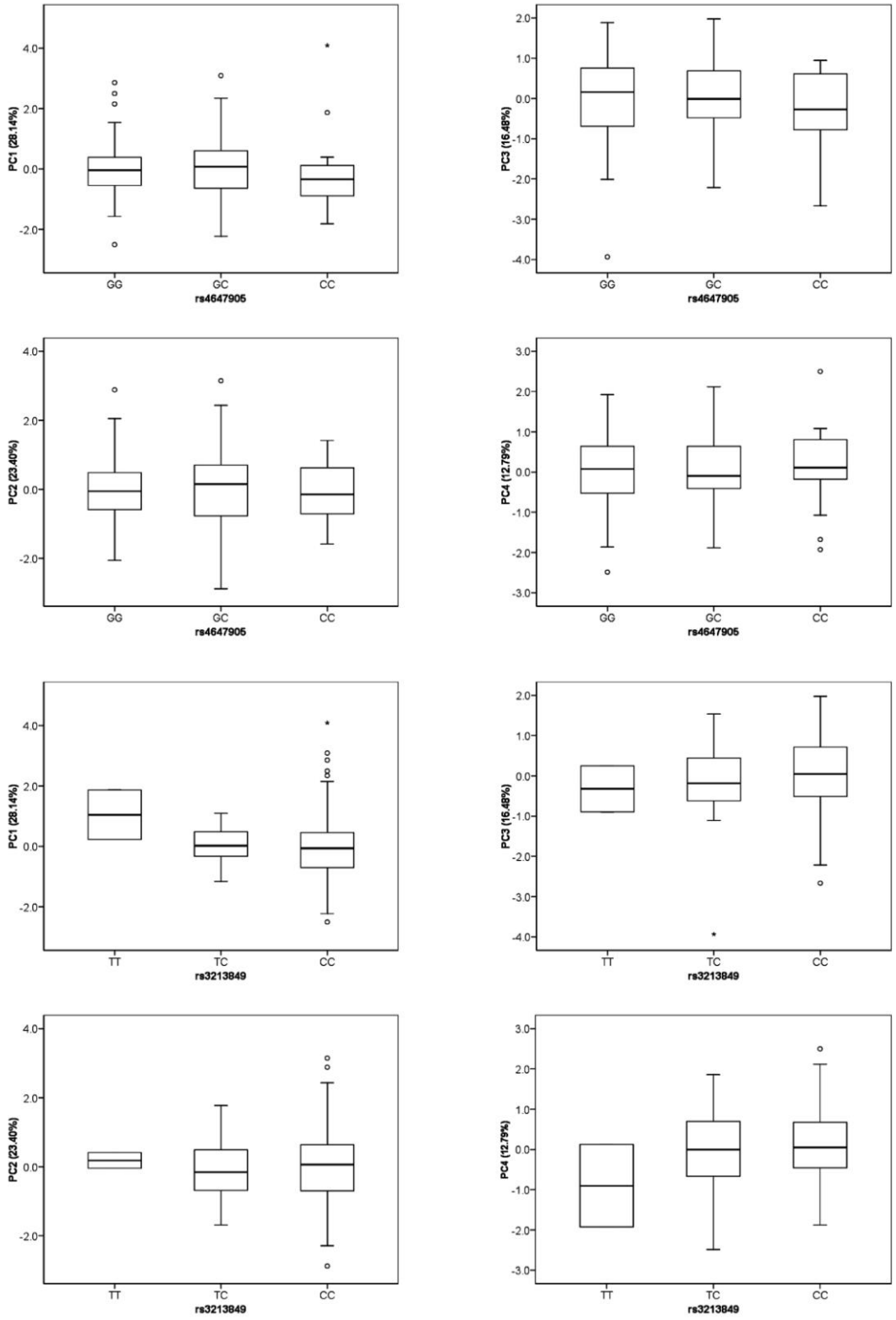


Figure 4. FGFR1 htSNPs Genotypes and Principal Components.

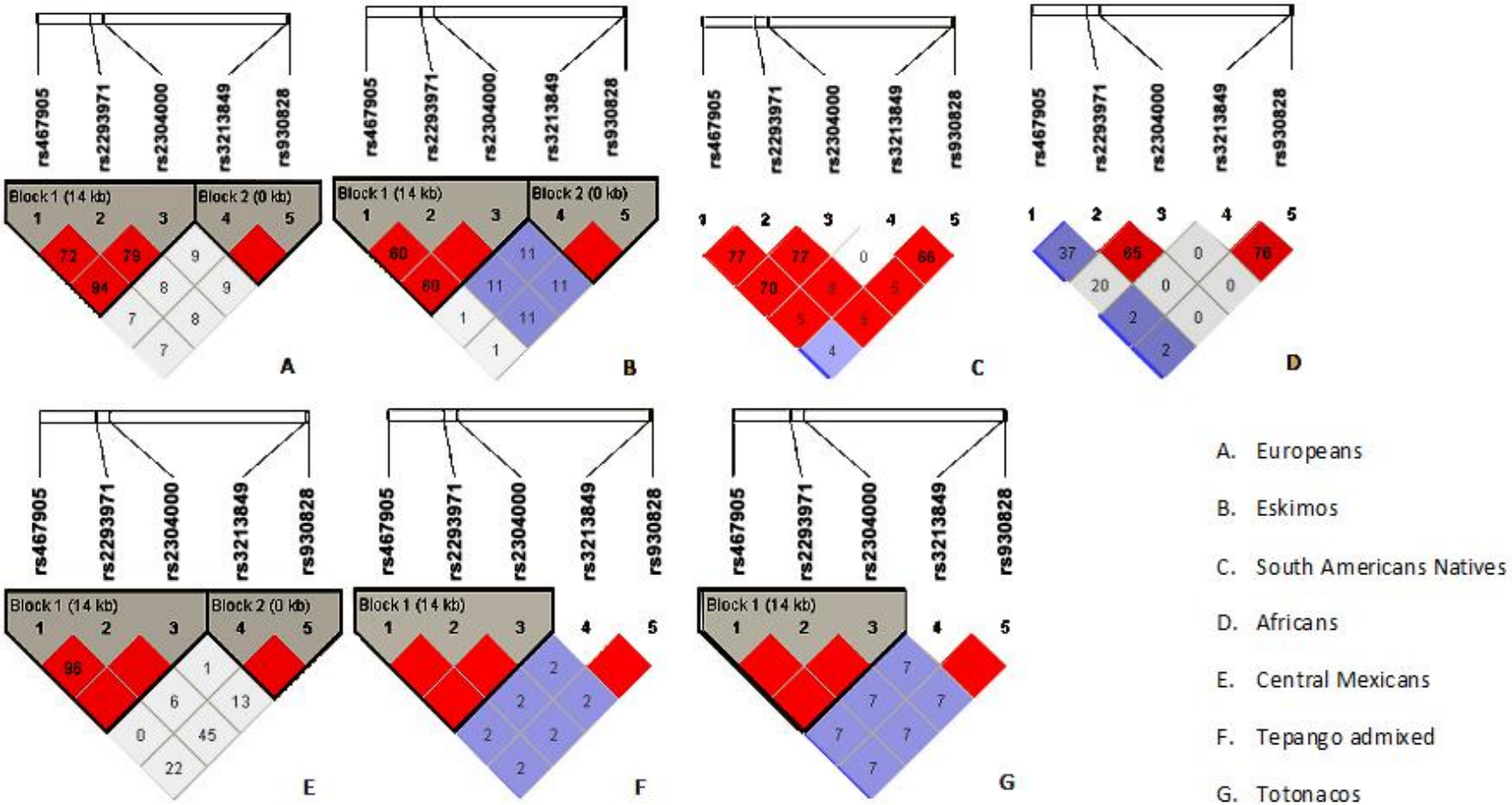


Figure 5. Pair-wise linkage disequilibrium between the five common *FGFR1* SNPs.

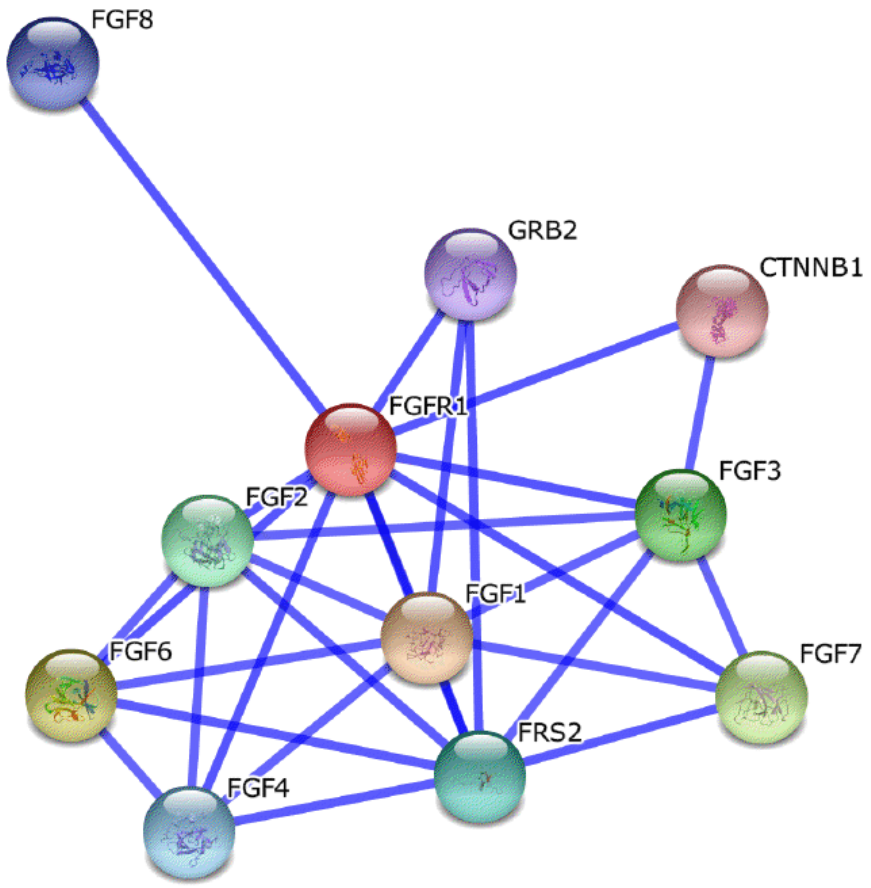


Figure 6. Eleven genes of the craniofacial development network. Image created by STRING (Search Tool for the Retrieval of Interacting Genes/P Genes/Proteins -fttp://string.embl.de) with the required highest confidence score – 0.90.

LITERATURE CITED

Acuña-Alonzo, V., Flores-Dorantes, T., Kruit, J. K., Villarreal-Molina, T., Arellano-Campos, O., Hünemeier, T., et al. (2010). A functional ABCA1 gene variant is associated with low HDL-cholesterol levels and shows evidence of positive selection in Native Americans. *Human molecular genetics*, *19*(14), 2877-2885.

Amorim CE, Wang S, Marrero AR, Salzano FM, Ruiz-Linares A, Bortolini MC. 2010 X-chromosomal Genetic Diversity and Linkage Disequilibrium Patterns in Amerindian and non-Amerindian populations. *American Journal of Physical Anthropology*, *In press.*

Barrett, J. C., Fry, B., Maller, J., & Daly, M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics (Oxford, England)*, *21*(2), 263-5.

Bamshad, M. & Wooding, S. (2003) Signatures of natural selection in the human genome. *Nature Reviews Genetics* 4:99-111

Beagle, D., Browning, B. L., & Browning, S. R. (2007). Documentation for BEAGLE 2.1. *Statistics*, 1-19.

Betti, L., Balloux, F., Hanihara, T., & Manica, A. (2010). The relative role of drift and selection in shaping the human skull. *American Journal of Physical Anthropology*, *141*(1), 76-82.

Browning, B. L., & Browning, S. R. (2009). A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *American journal of human genetics*, *84*(2), 210-23. The American Society of Human Genetics.

Comas, J. (1976). Manual de Antropología Física. Universidad Nacional Autónoma de México, México.

Coussens, A. K., & van Daal, A. (2005). Linkage disequilibrium analysis identifies an FGFR1 haplotype-tag SNP associated with normal variation in craniofacial shape. *Genomics*, *85*(5), 563-73.

Enlow, D.H. (1990) *Facial Growth*. Saunders. Philadelfia.

Helms, J. a., Cordero, D., & Tapadia, M. D. (2005). New insights into craniofacial morphogenesis. *Development (Cambridge, England)*, *132*(5), 851-61.

Itoh, N., & Ornitz, D. M. (2004). Evolution of the Fgf and Fgfr gene families. *Trends in genetics : TIG*, *20*(11), 563-9.

Lawson, H.A., Lesk, A.M., Miller, W. & Weiss, K.M. (2006). A comparative genomics approach to primate craniofacial evolution. *American Journal Physical Anthropoogy,* 129(S42).119.

Kidd, J. R., Pakstis, a. J., Zhao, H., Lu, R. B., Okonofua, F. E., Odunsi, a., et al. (2000). Haplotypes and linkage disequilibrium at the phenylalanine hydroxylase locus, PAH, in a global representation of populations. *American journal of human genetics*, *66*(6), 1882-99.

Kidd, K.K. (1987) Population Genetics if a Disease. Nature, 327, 282-283.

Kim, H. J., Rice, D. P., Kettunen, P. J., & Thesleff, I. (1998). FGF-, BMP- and Shh-mediated signalling pathways in the regulation of cranial suture morphogenesis and calvarial bone development. *Development (Cambridge, England)*, *125*(7), 1241-51.

Kimura, R., Yamaguchi, T., Takeda, M., Kondo, O., Toma, T., Haneji, K., et al. (2009). A common variation in EDAR is a genetic determinant of shovel-shaped incisors. *American journal of human genetics*, *85*(4), 528-35. The American Society of Human Genetics.

Lonjou, C., Zhang, W., Collins, A., Tapper, W. J., Elahi, E., Maniatis, N., et al. (2003). Linkage disequilibrium in human populations. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(10), 6069-74.

Nei, M. (2007). The new mutation theory of phenotypic evolution. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(30), 12235-42.

Rice, D.P., Rice, R. & Thesleff, I. (2003) Fgfr mRNA isoforms in craniofacial bone development. *Bone* 33(1): 14-27.

Richtsmeier, J.T., Weiss, K.M., Buchanan, A., Walker, A., Jablonski,N., Frazier, B.C. *et al.* (2007). Developmental genetic basis of primate craniofacial variation and human origins. *American Journal of Physical Anthropology,* 132(S44). 198.

Roseman, C. C. (2004). Detecting interregionally diversifying natural selection on modern human cranial form by using matched molecular and morphometric data. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(35), 12824-9.

Schroeder, K. B., Schurr, T. G., Long, J. C., Rosenberg, N. a., Crawford, M. H., Tarskaia, L. a., et al. (2007). A private allele ubiquitous in the Americas. *Biology letters*, *3*(2), 218-23.

Wilkie, A.O., Morriss-Kay, G.M., Jones, E.Y. & Heath, J.K. (1995) Functions of fibroblast growth factors and their receptors. *Current Biology* 5(5): 500-7.

ARTIGO III

Hünemeier T, Ruiz-Linares A, Silveira A, Paixão-Cortes V, Salzano FM, Bortolini MC. 2010. Population Data Support the Adaptive Nature of *HACNS1 Sapiens/Neandertal*-Chimpanzee Differences in a Limb Expression Domain. *American Journal of Physical Anthropology*. No prelo. DOI: 10.1002/ajpa.21378

AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 000:000–000 (2010)

Brief Communication: Population Data Support the Adaptive Nature of HACNS1 Sapiens/Neandertal-Chimpanzee Differences in a Limb Expression Domain

Tábita Hünemeier,[1] Andres Ruiz-Linares,[2] Álvaro Silveira,[1] Vanessa Rodrigues Paixão-Côrtes,[1] Francisco M. Salzano,[1] and Maria Cátira Bortolini[1]*

[1]*Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Caixa Postal 15053, 91501-970 Porto Alegre, RS, Brazil*
[2]*Department of Biology, University College London, The Galton Laboratory, London, UK*

KEY WORDS human evolution; Native Americans; morphological adaptations

ABSTRACT The 546-base pair enhancer of limb expression *HACNS1*, which is highly constrained in all terrestrial vertebrates, has accumulated 16 human-specific changes after the human-chimpanzee split. There has been discussion whether this process was driven by positive selection or biased gene conversion, without considering population data. We studied 83 South Amerindian, 11 Eskimo, 35 Europeans, 37 Bantu, and non-Bantu Sub-Saharan speakers, and 28 Brazilian mestizo samples and found no variation in this DNA region. Similar lack of variability in this region was found in four Africans, five Europeans or Euro-derived, two Asians, one Paleo-Eskimo, and one Neandertal sequence, whose whole genomes are publicly available. No difference was found. This result favors the interpretation of past positive and present conservative selection, as would expected in a region which influences *Homo*-specific traits as important as opposable thumbs, manual dexterity, and bipedal walking. Am J Phys Anthropol 000:000–000, 2010.

As humans, we have a special interest in identifying genetic modifications responsible for specific characteristics that distinguish us from the other great apes. Morphological differences can occur due to a small proportion of major-effect mutations in key structural or regulatory genes, as well as in noncoding sequences with a regulatory role (major gene effect hypothesis; Nei 1987, 2007). Prabhakar et al. (2008) have described a 546-base pair (bp) sequence that acts as an enhancer of gene expression, *HACNS1*, which is highly conserved in all terrestrial vertebrate genomes, but that has accumulated 16 human-specific changes in the ~6 million years that occurred since the human-chimpanzee split. Thirteen of these 16 mutations are found within an 81-bp functional segment of this 546-bp region. These authors showed, using a log-likelihood statistical test, that these findings are highly unexpected (P value = 9.2×10^{-12}) assuming just random events. Prabhakar et al. (2008) also demonstrated that these nucleotide substitutions promote a strong limb expression in humans when compared to the orthologous chimpanzee element, with a probable impact in our evolutionary history. Duret and Galtier (2009), on the other hand, argued that the *HACNS1* pattern of substitutions could be explained by a neutral process of biased gene conversion (BGC) associated with recombination events, which favors the fixation of AT → GC transitions, and can also drive the fixation of weakly deleterious mutations in functional elements. Prabhakar et al. (2009), after a genome-wide evaluation, demonstrated that *HACNS1* is not unusual regarding recombination events, and that in spite of an excess of AT → GC changes in this enhancer, deleterious (or neutral) modifications would not be expected to strengthen ancestral expression patterns introducing a new and robust expression domain in the human lineage. They also provide evidence that accelerated evolution in *HACNS1* is due to a common mechanism: synergy between BGC and positive selection producing a cluster of AT → GC substitutions at functional site (Prabhakar et al., 2009). Finally, Katzman et al. (2010) suggested that BGC could be an important factor to drive *HACNS1* evolution, but indicated that no data were available to distinguish whether the human-specific mutations reflect a process that was essentially like swimming upstream against an onslaught of non-selective BGC just to keep in place on the fitness landscape, or whether the mutation stress pushed these elements into a configuration that enabled some positive selection for higher fitness in humans. Up to now, no population data were considered in this discussion.

Native Americans have a well-known pattern of molecular variation, with striking differences from the other major human geographic groups, when neutral genetic systems are considered (Wang et al., 2007). Since

Grant sponsors: Conselho Nacional de Desenvolvimento Científico e Tecnológico; Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul.

*Correspondence to: Maria Cátira Bortolini, Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Caixa Postal 15053, 91501-970 Porto Alegre, RS, Brazil.
E-mail: maria.bortolini@ufrgs.br

Received 19 February 2010; accepted 9 June 2010

DOI 10.1002/ajpa.21378
Published online in Wiley Online Library (wileyonlinelibrary.com).

analyses with human populations have revealed that patterns of variation under natural selection are different from those expected under neutrality (Meyer et al., 2006) this study was designed to evaluate the two current hypotheses about more recent *HACNS1* molecular evolution. Data are presented on the 710-bp sequence which includes *HACNS1* and adjacent regions, in widely spread South Amerindians with distinct demographic histories. To broaden the ethnic and continental coverage, Siberian Eskimos, Europeans, Bantu, and non-Bantu Sub-Saharan speakers, and admixed Brazilian subjects were also investigated. Additionally a search was performed in published whole *sapiens* and Neandertal genome sequences.

SUBJECTS AND METHODS

A total of 194 samples were analyzed, including 83 Native Americans from 12 different populations (Apalai, Arara, Galibi, Kuben Kran Keng, Mundurucu, Karitiana, Xavante, Ache, Guarani, Ticuna, Wayuu and Zenu), widely spread all over the continent. Eleven Siberian Eskimos, 37 Sub-Saharan Africans, 35 Europeans, and 28 Brazilian mestizos were also included as independent non-Amerindian samples. The 37 African samples were obtained from Bantu-speaking and non Bantu-speaking subjects living in the Democratic Republic of Congo, Cameroon and Ivory Coast, while the Europeans are Spaniards. Additional information about these African and European populations can be found in Silva et al. (2006) and Bortolini et al. (2004), respectively. The Eskimo speak a Yupic language and live in a community in Siberia's extreme east. Twenty-eight Gaucho from southern Brazil were also sampled. They can be characterized as mestizo with Amerindian, African, and European ancestries (Marrero et al., 2007).

Ethical approval for this study was provided by the Brazilian National Ethics Commission (CONEP Resolutions nos. 123/1998 and 1333/2002), as well as by ethics committees in the countries where the non-Brazilian samples were collected.

The *HACNS1* 710-bp fragment was amplified with primers (F-5′TCTCGTCGGCATTACTCATCGTCA3′; R-5′CA AATGGAGGCTTTTCTGCA3′), specifically designed for this study. PCR reactions involved 20 to 50 ng of genomic DNA, 5 μL of Hot Start Master Mix Kit (Qiagen, Hilden, Germany), 4 μL of RNAse free water, and 0.5 μL (10 pmol/μL) of each primer pair, submitted to 94°C for 15 min, 33 cycles at 94°C for 30 s, 57°C for 40 s, 72°C for 30 s, followed by 10 min at 72°C. The PCR products were detected by 1.0% agarose gel electrophoresis and ethidium bromide staining, after purification with Microclean homemade kit. The DNA fragment was sequenced in an ABI 3730xl Sequencer, and analyzed with BioEdit v7.0.9 and Sequence Scanner v1.0.

The EvoNC program (Wong and Nielsen, 2004) was used to test if the variation in *HACNS1* could be explained by a neutral model or not. Sequences from six nonhuman primate species (*Otolemur garnettii, Callithrix jacchus, Macaca mulatta, Pongo pygmaeus, Gorilla gorilla, Pan troglodytes*), as well as from *Homo sapiens* and *Homo neanderthalensis* were assembled by searching Ensembl (http://www.ensembl.org) and the USCS genome browser (http://genome.ucsc.edu). The analyzed sequence included *HACNS1* (659 bp) plus exon1 of the GTPase activating protein *AGP1/CENTG2* gene (163 bp) located downstream. These alignments and a primate phylogenetic tree were then used to identify the *HACNS1* evolutionary pattern. The EvoNC program compares the rate of substitution in noncoding regions relative to the rate of synonymous substitution in coding regions. The parameter ζ, zeta, is then calculated. Three models are implemented: A null model is compared with two alternative models with two and three categories of ξ, respectively. The neutral model assumes two sets of sites with $\zeta 0 < 1$ and $\zeta 1 = 1$; the two-category model also assumes two sets with $\zeta 0 \leq 1$ and $\zeta 1 \geq 1$; and the three-category model assumes three sets with: $\zeta 0 < 1$, $\zeta 1 = 1$, and $\zeta 2 > 1$. Basically, $\zeta = 1$ when the sites in the noncoding region are evolving neutrally, $\zeta > 1$ when there is positive selection, and $\zeta < 1$ when negative selection is present. We also performed a posterior probability analysis to classify each individual site as belonging to one of the above-indicated ζ classes (Wong and Nielsen, 2004).

RESULTS AND DISCUSSION

The 13 substitutions clustered in the 81-bp region identified by Prabhakar et al. (2008), as well as the three others which distinguish us from chimpanzees, were observed in all individuals investigated, suggesting that they are fixed in our species. An additional mutation (position at chromosome 2: 236774267; Fig. 1) not reported by Prabhakar et al. (2008) since they sequenced just 546 bp (236773657 to 236774202) while we sequenced 710 bp (236773577 to 236774286), was also found distinguishing the two species (see Fig. 1). No intra or interpopulation variation was found, suggesting a conserved structure. It is worth mentioning that a 4,000-year-old permafrost-preserved Paleo-Eskimo sample, as well as one Khoisan, one Bantu, two Yoruba, one Han-Chinese, one Korean, and five European or European-derived individuals, who had their complete genomes sequenced, also present all these specific human substitutions (Rasmussen et al., 2010; Schuster et al., 2010; http://www.ensembl.org/; http://www.galaxy.org/). The comparison with the recently published Neandertal genome (Green et al., 2010) shows that 8 of the 13 human specific *HACNS1* substitutions present between 236774003 and 236774084 are also present in this hominid. Unfortunately, the Neandertal genome does not provide information for the complete *HACNS1* region, but other *sapiens*-specific mutations located outside this 81 bp region, including those that we describe here (236774267), are also present in the Neandertal genome (see Fig. 1).

Prabhakar et al. (2008) demonstrated using a log-likelihood statistic test that *HACNS1* is under functional constrain in humans, since its rapid divergence is highly unexpected given it strong conservation in others species (P value $= 9.2 \times 10^{-12}$). Here we performed an additional test to check if the *sapiens*/Neandertal variation can be explained by a neutral model. Table 1 shows the results. Comparison between the models showed a significant evidence for positive selection when the *sapiens*/Neandertal sequence is compared with those of other primates, including the chimpanzee (*Pan troglodytes*). The likelihood-ratio tests indicated that about 18% of the sites in this region are under positive selection. Interestingly, all the 13 sites recognized as exclusive of the *Homo* lineage showed signs of positive selection with a posterior probability greater than 99%. Our analysis thus provides additional evidence that a simple neutral model does not explain the variation that occurred after the *Homo-Pan* split.

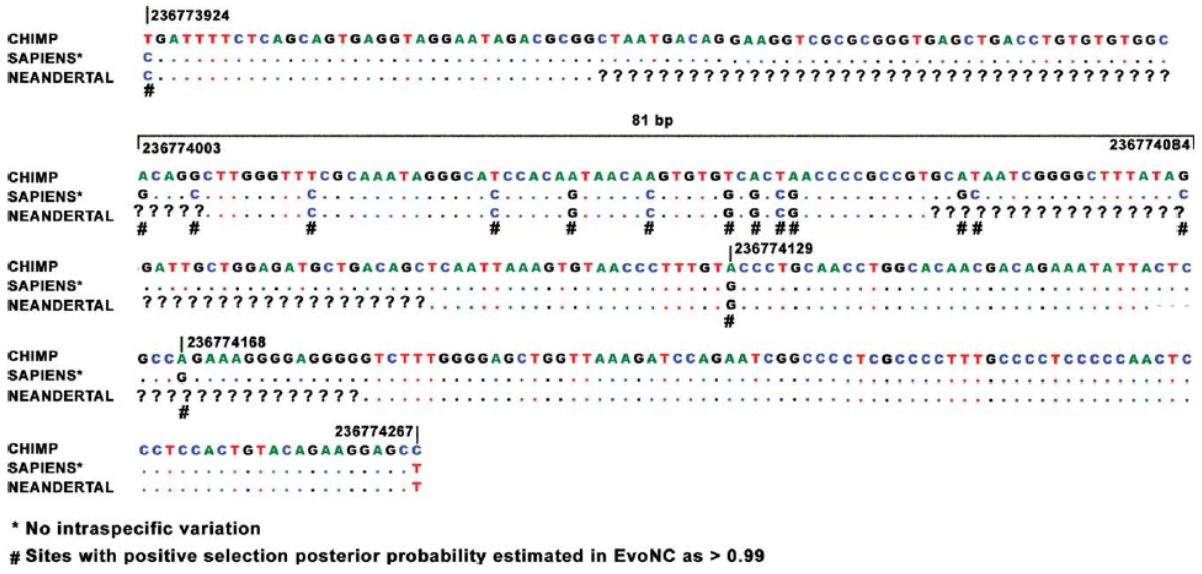


Fig. 1. Identification of the *sapiens* and Neandertal specific substitutions in *HACNS1* found within the 710 bp sequenced. Sites under positive selection are indicated (posterior probability >99%). The missing parts in the Neandertal sequence were assumed to be identical to *sapiens* in the neutral/selective tests.

TABLE 1. Parameters and likelihood scores under diferent models considering variable ζ among sets of nucleotides

	Model	Estimated parameters	ℓ	*P-value*
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	

χ^2 df = 2; Likelihood Ratio Test: $2\Delta\ell = 2(\ell 1 - \ell 0)$.

Native Americans present lower genetic diversity (as measured by π and other heterozygosity indices) and higher levels of population structure (as determined by F_{ST} or other similar statistics) than those seen in populations/groups from other continents (Cavalli-Sforza et al., 1994; Wang et al., 2007). The opposite is generally found when Sub-Saharan populations are investigated (Rosenberg et al., 2002). These diversity/divergence patterns are mainly due to demographic processes (successive founder effects with later expansions) related to the dispersal of modern *Homo sapiens* from Africa to other continents (Alonso and Armour, 2001; Ramachandran et al., 2005; Fagundes et al., 2007; Santos-Lopes et al., 2007; Wang et al., 2007). Since demographic effects affect the whole genome, deviations from this classical diversity/divergence model, as observed in this study, are expected when portions of the genome are under pressure by natural selection (Bamshad and Wooding, 2003; Barreiro et al., 2008).

Our results at the population level support the idea that these 13 substitutions confer some specific advantage to the *sapiens* lineage, probably to the *Homo* lineage, and that in some moment of the hominid evolutionary history they were fixed due to positive selection. After fixation, strong purifying selection has kept the intraspecific/intragenus conservation of the 81-bp cluster.

The real impact of this finding remains to be explored with more populational and functional studies, but the site of expression of these mutations suggests an influence in such *sapiens* or *Homo*-specific traits as opposable thumbs, manual dexterity, and ankle or foot adaptations for bipedal walking.

ACKNOWLEDGMENTS

The authors are very grateful to the individuals who donated the samples analyzed here and to the Fundação Nacional do Índio for logistic support in the Amerindian collections made in Brazil. They also thank Sandro L. Bonatto, Maria Luiza Petzl-Erler, Kim Hill, Ana Magdalena Hurtado, and Wilson A. da Silva Júnior for the Eskimo, Guarani, Ache and Sub-Saharan African samples, respectively.

LITERATURE CITED

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.

Bortolini MC, Thomas MG, Chikhi L, Aguilar JA, Castro-de-Guerra D, Salzano FM, Ruiz-Linares AR. 2004. Ribeiro's typology, genomes, and Spanish colonialism, as viewed from Gran Canaria and Colombia. Genet Mol Biol 27:1–8.

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.

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH, Hansen NF, Durand EY, Malaspinas AS, Jensen JD, Marques-Bonet T, Alkan C, Prüfer K, Meyer M, Burbano HA, Good JM, Schultz R, Aximu-Petri A, Butthof A, Höber B, Höffner B, Siegemund M, Weihmann A, Nusbaum C, Lander ES, Russ C, Novod N, Affourtit J, Egholm M, Verna C, Rudan P, Brajkovic D, Kucan Z, Gusic I, Doronichev VB, Golovanova LV, Lalueza-Fox C, de la Rasilla M, Fortea J, Rosas A, Schmitz RW, Johnson PL, Eichler EE, Falush D, Birney E, Mullikin JC, Slatkin M, Nielsen R, Kelso J, Lachmann M, Reich D, Pääbo S. 2010. A draft sequence of the Neandertal genome. Science 328:710–722.

Katzman S, Kern AD, Pollard KS, Salama SR, Haussler D. 2010. GC-biased evolution near human accelerated regions. PLoS Genet 6:e1000960.

Marrero AR, Bravi C, Stuart S, Long JC, Leite FPN, Kommers T, Carvalho CMB, Pena SDJ, Ruiz-Linares A, Salzano FM, Bortolini MC. 2007. Pre- and post-Columbian gene and cultural continuity: the case of the Gaucho from southern Brazil. Hum Hered 64:160–171.

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, Noonan 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.

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., A worldwide phylogeography for the human X chromosome. PLoS One 2:e557, 2007.

Schuster SC, Miller W, Ratan A, Tomsho LP, Giardine B, Kasson LR, Harris RS, Petersen DC, Zhao F, Qi J, Alkan C, Kidd JM, Sun Y, Drautz DI, Bouffard P, Muzny DM, Reid JG, Nazareth LV, Wang Q, Burhans R, Riemer C, Wittekindt NE, Moorjani P, Tindall EA, Danko CG, Teo WS, Buboltz AM, Zhang Z, Ma Q, Oosthuysen A, Steenkamp AW, Oostuisen H, Venter P, Gajewski J, Zhang Y, Pugh BF, Makova KD, Nekrutenko A, Mardis ER, Patterson N, Pringle TH, Chiaromonte F, Mullikin JC, Eichler EE, Hardison RC, Gibbs RA, Harkins TT, Hayes VM. 2010. Complete Khoisan and Bantu genomes from southern Africa. Nature 463:943–947.

Silva WA Jr, Bortolini MC, Schneider MPC, Marrero AR, Elion J, Krishnamoorty R, Zago MA. 2006. mtDNA haplogroup analysis of Black Brazilian and sub-Saharan populations: implications for the Atlantic slave trade. Hum Biol 78:29–41.

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.

Wong WS, Nielsen R. 2004. Detecting selection in non-coding regions of nucleotide sequences. Genetics 167:949–958.

ARTIGO IV

Gene-culture Dynamics: An Example Involving Native Americans

Tábita Hünemeier,[1] Víctor Acuña-Alonzo,[2,3] Veronica Contini, [1] Francisco Rothhammer, [4,5] Jean-Michel Dugoujon, [6] Stephane Mazíeres, [6,7] María Teresa Villarreal-Molina, [8] Vanessa Rodrigues Paixão-Côrtes,[1] Pau Corral–Montañes, [9] Francisco M. Salzano,[1] Samuel Canizales-Quinteros, [8, 10] Andres Ruiz-Linares,[9] and Maria Cátira Bortolini[1]

[1]*Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Caixa Postal 15053, 91501-970 Porto Alegre, RS, Brazil*
[2]*Unidad de Genómica de Poblaciones Aplicada a la Salud, Instituto Nacional de Medicina Genómica, Mexico City, Mexico*
[3]*Molecular Genetics Laboratory, Escuela Nacional de Antropología e Historia, Mexico City, Mexico*
[4] *Programa de Genética Humana, Instituto de Ciencias Biomédicas, Facultad de Medicina, Universidad de Chile, Santiago, Chile*
[5] *Instituto de Alta Investigación, Universidad de Tarapaca´, Arica, Chile*
[6] *Laboratoire d'Anthropologie Moléculaire et Imagerie de Synthèse, FRE 2960 CNRS and Université de Toulouse 3 Paul Sabatier, Toulouse, France*
[7]*Anthropologie Bioculturelle, Faculté de Médecine, CS80011, Marseille, France*
[8]*Unit of Molecular Biology and Genomic Medicine, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán,*
[9] *Universitat Pompeu Fabra, 08003 Barcelona, Spain*
[10]*Departamento de Biología, Facultad de Química,Universidad Nacional Autónoma de México, Mexico City, Mexico*
[11]*The Galton Laboratory, Department of Biology, University College London, London, United Kingdom*

Number of pages: 21

Abbreviated title: gene-culture coevolution

KEY WORDS: Native Americans, gene-culture coevolution, *ABCA1*

Contract grant sponsors: Conselho Nacional de Desenvolvimento Científico e Tecnológico; Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul; Programme Interdisciplinaire CNRS: Amazonie- Analyse, modélisation et ingénierie des systèmes amazoniens.

Correspondence to: Maria Cátira Bortolini, Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Caixa Postal 15053, 91501-970 Porto Alegre, RS, Brazil. E-mail: maria.bortolini@ufrgs.br

Manuscrito em preparação (a ser submetido para a revista *American Journal of Physical Anthropology*)

ABSTRACT

New data on the ATP-binding cassette transporter A1 (*ABCA1*) *Arg230Cys* polymorphism for 106 individuals affiliated to 12 South Amerindian populations and interpreted with those from 38 other native American groups are presented. The variant distribution was considered in detail, and the *230Cys* age estimated (8,268 ± 5,916 BP) is compatible with its origin occurring in the America continent. Correlation of its frequencies with the *Zea* pollen archeological information in Meso America/Central America suggests that maize domestication was driving force in the variant´s increase in frequencies in that region. These results may represent the first example of gene-culture coevolution involving an autochthonous Native American allele.

Modern humans differ from each other in many ways. It is known that a significant portion of the differences in physical appearance, susceptibility to diseases, and other traits found within and between human populations can be attributable to genetic variants (Bamshad and Wooding, 2003). Even considering that many of these variants are eliminated or fixed as a consequence of drift, it is expected that at least part of these differences resulted from selective pressures in different geographical regions, each with its own climate, pathogens and sources of food (Young et al., 2005; Harris and Meyer, 2006; Williamson et al., 2007; Barreiro et al., 2008; Hancock et al., 2008, 2010). Additionally, human cultural practices have drastically modified environmental conditions and behaviors, promoting fast and strong changes at the genome level, many times associated to positive selection and adaptation (gene-culture dynamics; Laland et al., 2010; Richerson et al., 2010). A particularly important event in the *Homo sapiens sapiens* history, which triggered a new and striking gene-culture-coevolution cycle, was the development of agriculture and animal domestication during the Neolithic period (~10,000 years ago). Two well-studied cases of recent and fast selection due to Neolithic cultural pressures are the high copy number variation of the amylase gene and the spread of alleles for adult lactose tolerance in agricultural and pastoral societies, respectively (Holden and Mace, 1997; Bersaglieri et al., 2004; Beja-Pereira et al., 2003, ; Burger et al., 2007; Tishkoff et al., 2007). Another interesting example was identified in West Africa, among the Kwa-speaking agriculturalists. They cut and clear the forest to grow yams, increasing the amount of standing water after raining, thus providing better breeding grounds for malaria-carrying mosquitoes (Laland et al., 2010). This intensified the action of natural selection for the *HbS* allele, since in heterozygous condition, it confers protection against

malaria (Livingstone, 1958). Laland et al. (2010) in their recent review about gene-culture coevolution listed more then one hundred genes which were identified as having being subjected to recent and rapid selection due to cultural pressures, thus providing an initial idea about the extension of the phenomenon in our evolutionary history.

Genetic studies with Native American populations, using blood groups and other classical polymorphisms, started in a systematic and regular way more then 50 years ago with some pioneer investigations (Salzano, 1957; Neel and Salzano, 1967). Today, knowledge about America´s peopling, population dispersion, structure, admixture dynamics and phylogenetic relationships have been obtained using molecular markers (Pena et al., 1995; Santos et al., 1996; Bonato and Salzano, 1997a and 1997b; Tarazona-Santos and Santos, 2002; Hunley and Long, 2005; Bortolini et al., 2002, 2003; Salzano, 2006; Wang et al., 2007; Fagundes et al., 2008; Lewis and Long, 2008; Mazières et al., 2008, 2009). But well-documented examples for gene-culture dynamics involving Native Americans are rare, and most of them are related to linguistic and genetic coevolution (Hunley et al., 2007).

Recently, however, Tovo-Rodrigues et al. (2010) investigated the D4 dopamine receptor gene (*DRD4*) distribution in several Amerindian groups and found a significant frequencies in allele distribution between hunter-gatherers and agriculturalists, with an increase of the *7R* allele among hunter-gatherers. In another recent publication (Acuã-Alonzo et al., 2010) it was demonstrated that the *230Cy*s allele (*Arg230Cys*, rs9282541) of the ATP-binding cassette transporter A1 (*ABCA1*) gene, which was previously associated with low HDL-cholesterol levels and obesity-related comorbidities, is exclusively present in Native American and mestizo individuals (Acuña-Alonzo et al., 2010). The latter

authors verified that the cells expressing the *ABCA1*230Cys* allele showed a 27% cholesterol efflux reduction, confirming that this variant has a functional effect *in vitro*. Additionally they demonstrated that *230Cys* resides on a haplotype which is the target of an ongoing directional selective sweep, suggesting an American origin and selective advantage for *230Cys* in the past during food deprivation periods. On the other hand, under current modern lifestyle *230Cys* may have become a major susceptibility allele for low HDL levels and correlated metabolic diseases (Acuña-Alonzo et al., 2010). These results would provide another example of the "thrifty" genotype hypothesis, which suggests that variants that increase the efficiency of energy use and storage during periods of famine would be positively selected during specific periods of human evolution history (Neel, 1962).

Here we present additional data about the *Arg230Cys* polymorphism in Native Americans and then used all available information to date the origin of the variant allele, as well as to test the hypothesis that it could represent the first example of gene-culture coevolution involving an autochthonous Native American allele.

SUBJECTS AND METHODS

Populations

New data for the *Arg230Cys* polymorphism were generated for twelve Amerindian populations from Bolivia, Chile, Colombia and French Guiana (N= 106). Table 1 shows the linguistic affiliation, as well as geographical location of these and 38 other populations previously studied for this polymorphism (Acuña-Alonzo et al., 2010). Additional

information about the tribes investigated for the present study can be found in Bortolini et al. (2002, 2003), Mazières et al. (2008, 2009) and Wang et al (2007).

Ethical approvals were provided by ethics committees of the countries where the samples were collected.

SNP genotyping

The 718 bp of *ABCA1* Exon 7 were amplified with primers (F - 5'GCCAAATCATGTGTCCCACTC 3' ; R – 5'TGGCTTAAACTCAGCCACCC3') designed for this study. PCR reactions involved 20 to 50 ng of genomic DNA, 5 uL of Hot Start Master Mix Kit (Qiagen, Hilden, Germany), 4 uL of RNAse free water, and 0.5 uL (10 pmol/lL) of each primer pair, submitted to 94°C for 15 min, 33 cycles at 94°C for 30 s, 59°C for 40 s, 72°C for 30 s, followed by 10 min at 72°C. The PCR products were detected by 1.0% agarose gel electrophoresis and ethidium bromide staining, after purification with Microclean home-made kit. The DNA fragment was sequenced in an ABI 3730xl Sequencer, and analyzed with BioEdit v7.0.9 and Sequence Scanner v1.0.

Allele frequencies vs maize domestication

Pollen samples taken from sediments in lakes, swamps, and archeological deposits provide an independent view of the presence or absence of *Zea* (maize and teosina) in the Americas and have been used to estimate the age of maize domestication and dispersion (Blake, 2006). The latter author summarized the *Zea* pollen dates from several archeological sites in Americas. However, to test the connection between maize culture and the *ABCA1*230Cys* we only used archeological sites located in Mesoamerica/Central America near places where allele frequencies were available. The dispersion, adjusted by

regression analysis, was conducted with the use of SPSS15.0, while the statistical significance was accessed by an ANOVA test.

*ABCA1*230C*ys allele age

An estimate of the age of the variant was generated with a set of SNPs located at medium distance from the core site using the moment estimator method, in which the age of an allele can be estimated, based on intra-allelic variation following the exponential decay of linkage disequilibrium due to recombination and mutation rates (Slatking and Rannala, 2000). The SNPs used for this analysis represent a subset of an unpublished major pannel obtained for several Native American populations, including those investigated in present study (A. Ruiz-Linares et al., personal communication).

Maps

An American map of *ABCA1*230C*ys frequencies, as well as a *Zea* pollen relics map were obtained using IDRISI 15.0 software (Eastman, 2006).

RESULTS

Table 2 presents the genotype and allele frequencies for 106 individuals analyzed. The variant was detected in Chilean Aymara, Hulliche, Ingano, Palukir, Quechua and Wayuu, but not in the Aymara from Bolivia, or in groups from Colombia (Embera, Kogi, Ticuna, Zenu) and Chilote from Chile the allele was not found. These results, however, should be considered with caution since sample sizes are small form some populations investigated here.

Figure 1A illustrates the *ABCA1*230Cys* distribution considering all data shown in Table 2. The variant reaches high frequencies in Mesoamerica/Central America, especially in Mexican west coast. In South America, particularly in Amazonia and Central Brazil, some high frequency spots can be observed. Table 3 presents the proportion of populations within each main geographic region with absence, low (0.01-0.05), medium (0.06-0.15), or high (> 0.15) *230Cys* frequencies. Interestingly, 83% of the Mesoamerican/Central American populations have *230Cys* with frequencies higher than 0.15. On the other hand, the absence or low frequencies of the allele are observed in the three North American groups tested, while high heterogeneity is observed in South America, especially considering the lowlands (Amazonia/Brazilian Plateau and Chaco) area. No populations located in the Andes present *230Cys* frequencies higher than 0.15.

A significant correlation of the *ABCA1*230Cys* allele frequencies with the distribution of the *Zea* pollen relics (Figure 1B) in Mesoamerica/Central America was observed (Figure 2 and Table S1; $r = 0.97$, $p < 0.001$), implying that maize domestication probably is the force driving up the frequency and expansion of *ABCA1*230Cys* at least in Mesoamerica/Central America.

Finally, we estimated the *ABCA1*230Cys* age based on phased SNP haplotypes from an unpublished Native American dataset (A. Ruiz-Linares et al., personal communication). Ten 10 located upstream and at medium distance from the 230 site were selected, since the method based on the moments estimator (Slatkin and Rannala, 2000) is not suitable for a region of low average recombination rates. The analysis leads to a preliminary estimate that the mutation occurred $8{,}268 \pm 5{,}916$ years before present (Table S1), which is compatible with an American origin.

DISCUSSION

We can now examine our preliminary hypothesis in the light of the results obtained. Maize (*Zea mays* ssp. *mays*), the most important crop of the Americas, was domesticated probably once from a wild grass-teosinte (*Zea mays* ssp. *parviglumis*) in Mexico (Central Balsas region) about 6,300-10,000 calendar years before present (Matsuoka et al., 2002; Dull, 2006; Jaenicke-Després and Smith, 2006; Lesure, 2008; Ranere et al., 2009). Several lines of evidence indicated that the Mesoamerican village lifestyle began with maize domestication (Raymond and Deboer, 2006; Chisholm and Blake, 2006). From its origin region, maize journeyed south, traveling hand-in-hand with pottery and bringing sedentary life to the Andes. Other crops were also present in pre-Columbian Mesoamerican civilizations (squash and beans, with more recent origin; Brown, 2010), but maize was the base of diet for most part of them. For example, Benedict and Steggerda (1936) showed that 75% of the calories consumed by the Mayas were derived from maize. Additionally, Mesoamerica was the world´s only region were an ancient civilization lacked a domestic herbivore. Protein from domesticated animal sources would have been therefore scarce in Pre-Hispanic Mesoamerica/Central America by comparison with other parts of the ancient urbanized world, including the Andes (Parsons, 2010). As a whole, these studies demonstrated the importance of maize for the diet of the first Mesoamerican/Central American sedentary communities. It is worth mention, however, that early farmers commonly display signs of growth disruption questioning the common assumption that farming and sedentary lifestyle brought increased dietary stability and health homeostasis (Cohen, 2008). Several studies reveal that homeostasis should have declined with sedentary farming. Bioarcheologists and paleopathologists have detected a deterioration in Mesoamerican health indices from ~8,000 to ~500 years before present (Kennett et al.,

2006 and references therein). Domestic crops are more vulnerable than wild ones, crowding promotes crop diseases and storage systems often fail (estimates suggest that as much as 30% of stored food is lost even in a modern sophisticated system; Cohen, 2008). Clearly, diet based on one or few crops were deleterious to health in the pre-Columbian era (Steckel et al., 2002). In this context, it is reasonable to suppose that the *ABCA1*230Cys* allele could have had a selective advantage during periods of food scarcity experimented by Mesoamericans/Centro Americans during the implementation of the sedentary life style based on maize culture. The strong correlation between maize culture propagation and *230Cys* frequencies in this region reinforces this suggestion, even considering that the allele´s advantage may have been lost after technological innovations were implemented and the agricultural production was stabilized. Peng et al. (2010) presented evidences of a similar case of gene-culture coevolution. They suggested positive selection for the *ADH1B*47His* allele caused by the emergence and expansion of rice domestication in East Asia.

Other environmental factors may also be involved in the *ABCA1* allele distribution since cholesterol plays an important role in various infectious processes, such as the entry and replication of Dengue virus type 2 and flaviviral infection (Lee et al., 2008). Additionally, the ABCA1 transporter participates in infectious and/or thrombotic disorders involving vesiculation, since homozygous *ABCA1* gene deletions confer complete resistance against cerebral malaria in mice (Simons et al., 2000; Combes et al., 2005). These findings can be showing an additional causal factor to *ABCA1*230Cys* selective sweep associated to agricultural development. Sedentary village life with corresponding growth in local population density can promote an increase in the mortality rate, particularly in children under 5 years of age. Archeological and paleoecological evidences

in Europe, for example, showed that during the Neolithic demographic transition (communities of hunters/gatherers starting to cultivate and domesticate wild plants or wild animals) the causes of increased infant mortality would include lack of drinking water supplies, contamination by faeces, emergence of highly virulent zoonoses, as well as increase in the prevalences of other germs such as *Rotavirus* and *Coronavirus* (diarrhea, one of the main killers of children under 5 years of age), *Streptococcus, Staphylococcus*, *Plasmodium* (*P. falciparum*, and *P. vivax*, which is believed to have emerged more recently) or *Herpesvirus* (Bocquet-Appel, 2008). The real impact of the *ABCA1*230Cys* variant in these infectious processes, however, deserve additional functional studies.

Finally, the gene-culture coevolution scenario suggested by us for Mesoamerica/Central America cannot be extrapolated to other American regions. Both North and South America present much more diversity in relation to habitats, people and culture. For instance, maize arrived in South America about 4,000 calendar years before present, but apparently the level of consumption seen in Mesoamerica/Central America was rarely seen there. Archaeological data indicate that only during the implantation and expansion of the Inca Empire (800-500 YBP) the level of maize consumption was important, nevertheless not similar with that seen in Mesoamerica/Central America (Schwarcz, 2006; Tykot et al.,2006). Additionally, lowland South Amerindians present lower intrapopulation genetic variation and higher levels of population structure when compared to those seen in Andean populations (Tarazona-Santos et al., 2001; Wang et al., 2007). These results indicate low levels of gene flow between villages/populations and low effective population sizes, favoring the role of genetic drift. Conversely, the Andean groups show opposing characteristics. This correlates well with distinguishing patterns of gene flow and historical effective sizes of these indigenous populations, as well as cultural

differences, paleoclimatic and environment changes of their habitats (Tarazona-Santos et al., 2001). The significant role of randomic processes makes difficult to define in Lowlands a particular selective sweep, since favorable alleles can be easily lost and signals of natural selection obscured due to drift.

CONCLUSION

Our analyses provide evidences that *230Cys*, a common, functional and private Native American allele, with an estimated origin of 8,268 ± 5,916 years before present, was the target in Mesoamerica/Central America of an ongoing directional selective sweep consistent with the origin and spread of maize culture. But this scenario is restricted to that region. In Amerindians from South America, genetic drift and/or cultural changes could have disturbed or obscured the postulated *ABCA1*230Cys* allele adaptative advantage.

ACKNOWLEDGMENTS

We are very grateful to the volunteers that contributed in the sampling. We thank M. Villena and R. Vasquez for assistance in the collection of samples in Bolivia. We also thank R. Bisso-Machado for his help with tables editing.

LITERATURE CITED

Acuña-Alonzo V, Flores-Dorantes T, Kruit JK, Villarreal-Molina T, Arellano-Campos O, Hünemeier T, Moreno-Estrada A, Ortiz-López MG, Villamil-Ramírez H, León-Mimila P, Villalobos-Comparan M, Jacobo-Albavera L, Ramírez-Jiménez S, Sikora M, Zhang LH, Pape TD, Granados-Silvestre Mde A, Montufar-Robles I, Tito-Alvarez AM, Zurita-Salinas C, Bustos-Arriaga J, Cedillo-Barrón L, Gómez-Trejo C, Barquera-Lozano R, Vieira-Filho JP, Granados J, Romero-Hidalgo S, Huertas-Vázquez A, González-Martín A, Gorostiza A, Bonatto SL, Rodríguez-Cruz M, Wang L, Tusié-Luna T, Aguilar-Salinas CA, Lisker R, Moises RS, Menjivar M, Salzano FM, Knowler WC, Bortolini MC, Hayden MR, Baier LJ, Canizales-Quinteros S. 2010. A functional ABCA1 gene variant is associated with low HDL-cholesterol levels and shows evidence of positive selection in Native Americans. Hum Mol Genet. 19:2877-2885.

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.

Beja-Pereira A, Luikart G, England PR, Bradley DG, Jann OC, Bertorelle G, Chamberlain AT, Nunes TP, Metodiev S, Ferrand N, Erhardt G. 2003. Gene-culture coevolution between cattle milk protein genes and human lactase genes. Nat Genet 35:311-313. Erratum in: Nat Genet 2004, 36:106.

Benedict F G and Steggerda M .1936. The Food of Present-day Maya Indians of Yucatan. Washington, DC:Carnegie Institute of Washington, Publication 456, Contribution 18.

Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, Drake JA, Rhodes M, Reich DE, Hirschhorn JN 2004. Genetic signatures of strong recent positive selection at the lactase gene. Am J Hum Genet 74:1111-1120.

Blake M. 2006. Dating the initial spread of *Zea mays*. In: Staller J E, Tykot R H and Benz B F, editors. Histories of maize. San Diego: Academic Press. p.55-68.

Bocquet-Appel JP. 2008. Explaining the Neolithic demographic transition. In: Bocquet-Appel J P and Bar-Yosef O, editors, The Neolithic demographic transition and its consequence. New York: Springer. p35-56.

Bonatto SL and Salzano FM. 1997a. A single and early migration for the peopling of the Americas supported by mitochondrial DNA sequence data. Proc Natl Acad Sci USA 94: 1866-1871.

Bonatto SL and Salzano FM. 1997b. Diversity and age of the four major mtDNA haplogroups, and their implications for the peopling of the New World. Am J Hum Genet 61:1413-1423.

Bortolini MC, Salzano FM, Bau CHD, Layrisse Z, Petzl-Erler ML, Tsuneto LT, Hill K, Hurtado AM, Castro-de-Guerra D, Bedoya G, Ruiz-Linares A. 2002. Y-chromosome biallelic polymorphisms and Native American population structure. Ann Hum Genet 66: 255-259.

Bortolini MC, Salzano FM, Thomas MG, Stuart S, Nasanen SPK, Bau CHD, Hutz MH, Layrisse Z, Petzl-Erler ML, Tsuneto LT, Hill K, Hurtado AM, Castro-de-Guerra D, Torres MM, Groot H, Michalski R, Nymadawa P, Bedoya G, Bradman N, Labuda D, Ruiz-Linares A. 2003. Y-chromosome evidence for differing ancient demographic histories in the Americas. Am J Hum Genet 73: 524-539.

Brown C H. 2010. The Pastoral Niche in Pre-Hispanic Mesoamerica. In: Staller J; Carrasco M, editors, Pre-Columbian foodways: interdisciplinary approaches to food, culture, and markets in ancient Mesoamerica. New York: Springer. p273-291.

Burger J, Kirchner M, Bramanti B, Haak W, Thomas MG. 2007. Absence of the lactase-persistence-associated allele in early Neolithic Europeans. Proc Natl Acad Sci U S A 104: 3736-3741.

Campbell L. 1997. American Indian languages: the historical linguistics of Native America. New York: Oxford University Press.

Chisholm B And Blake M.2006. Diet in prehistoric Soconusco. In: Staller J E, Tykot R H and Benz B F, editors. Histories of maize. San Diego: Academic Press. p.161-167.

Cohen N M. 2008. Implications of the NDT for worldwide health and mortality in prehistory. In: Bocquet-Appel J P and Bar-Yosef O, editors, The Neolithic demographic transition and its consequence. New York: Springer. p481-500.

Combes V, Coltel N, Alibert M, van Eck M, Raymond C,Juhan-Vague I, Grau GE, Chimini G. 2005. ABCA1 gene deletion protects against cerebral malaria: potential pathogenic role of microparticles in neuropathology Am J Pathol 166: 295–302.

Dull R. 2006. The maize revolution: a view from El Salvador. In: Staller J E, Tykot R H and Benz B F, editors. Histories of maize. San Diego: Academic Press.p 357 -363.

Eastman JR. 2006. IDRISI 15.0: The Andes edition. Clark University, Worcester.

Fagundes NJR, Kanitz R, Eckert R, Valls ACS, Bogo MR, Salzano FM, Smith DG, Silva Jr WA, Zago MA, Ribeiro-dos-Santos AK, Santos SEB, Petz-Erler ML, Bonatto SL. 2008. Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. Am J Hum Genet 82: 583-592.

Hancock AM, Witonsky DB, Gordon AS, Eshel G, Pritchard JK, Coop G, Di Rienzo A. 2008. Adaptations to climate in candidate genes for common metabolic disorders. PLoS Genet 4:e32.

Hancock AM, Witonsky DB, Ehler E, Alkorta-Aranburu G, Beall C, Gebremedhin A, Sukernik R, Utermann G, Pritchard J, Coop G, Di Rienzo A. 2010. Human adaptations to diet, subsistence, and ecoregion are due to subtle shifts in allele frequency. Proc Natl Acad Sci U S A 107 Suppl 2:8924-8930.

Harris EE, Meyer D (2006) The molecular signature of selection underlying human adaptations. Yearb Phys Anthropol 49: 89-130.

Holden C, Mace R. 1997. Phylogenetic analysis of the evolution of lactose digestion in adults. Hum Biol 81:597-619.

Hunley K, Long JC. 2005. Gene flow across linguistic boundaries in Native North American populations. Proc Natl Acad Sci U S A 102: 1312-1317.

Hunley KL, Cabana GS, Merriwether DA, Long JC. 2007. A formal test of linguistic and genetic coevolution in native Central and South America. Am J Phys Anthropol 132:622-631.

Hunley KL, Spence JE, Merriwether DA. 2008. The impact of group fissions on genetic structure in Native South America and implications for human evolution. Am J Phys Anthropol 135:195-205.

Jaenicke-Després V R and. Smith B D (2006), Ancient DNA and the integration of archaeological and genetic approaches to the study of maize domestication. In: Staller J E, Tykot R H and Benz B F, editors. Histories of maize. San Diego: Academic Press. p 83-92.

Kennett D J, Voorhies B and Martorana D. 2006. An ecological model for the origins of maize-based food production on the Pacific Coast of Southern Mexico. In: Kennett D J and Winterhalder B (editors), Behavioral ecology and the transition to agriculture, Berkeley: University of California Press. p103-136.

Laland KN, Odling-Smee J, Myles S. 2010. How culture shaped the human genome: bringing genetics and the human sciences together. Nat Rev Genet 11:137-148.

Lee CJ, Lin HR, Liao CL, Lin YL (2008) Cholesterol effectively blocks entry of flavivirus. J Virol 82: 6470–6480.

Lesure R G. 2008. The Neolithic demographic transition in Mesoamerica? Larger implications of the strategy of relative chronology. In: Bocquet-Appel J P and Bar-Yosef O, editors, The Neolithic demographic transition and its consequence. New York Dordrecht Heidelberg London: Springer. p107-138.

Lewis MP (ed.) 2009. Ethnologue: languages of the world. 16th edition. SIL International, Dallas. Online version: http://www.ethnologue.com/.

Lewis CM Jr, Long JC. 2008. Native South American genetic structure and prehistory inferred from hierarchical modeling of mtDNA. Mol Biol Evol 25: 478-486.

Livingstone F B. 1958. Anthropological implications of sickle-cell distribution in West Africa. Am. Anthropol. 60: 533–562.

Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez JG, Buckler E, Doebley J. 2002. A single domestication for maize shown by multilocus microsatellite genotyping Proc Natl Acad USA 99: 6080-6084.

Mazières S, Guitard E, Crubézy E, Dugoujon JM, Bortolini MC, Bonatto SL, Hutz MH, Bois E, Tiouka F, Larrouy G, Salzano FM. 2008. Uniparental (mtDNA, Y-chromosome) polymorphisms in French Guiana and two related populations--implications for the region's colonization. Ann Hum Genet 72:145-156.

Mazières S, Sevin A, Callegari-Jacques SM, Crubézy E, Larrouy G, Dugoujon JM, Salzano FM. 2009. Population genetic dynamics in the French Guiana region. Am J Hum Biol 21:113-117.

Neel, JV. 1962 . Diabetes mellitus: a ‘thrifty’ genotype rendered detrimental by ‘progress’? Am J Hum Genet 14: 353–362.

Neel JV and Salzano FM. 1967. Further studies on the Xavante Indians. X. Some hypotheses-generalizations resulting from these studies. Am J Hum Genet 19:554–574.

Parsons JR. 2010. The pastoral niche in pre-hispanic Mesoamerica. In: Staller J; Carrasco M, editors, Pre-Columbian foodways: Interdisciplinary approaches to food, culture, and markets in ancient Mesoamerica. New York Dordrecht Heidelberg London: Springer. p109-136.

Pena SD, Santos FR, Bianchi NO, Bravi CM, Carnese FR, Rothhammer F, Gerelsaikhan T, Munkhtuja B, Oyunsuren T. 1995. A major founder Y-chromosome haplotype in Amerindians. Nat Genet 11: 15-16.

Peng Y, Shi H, Qi XB, Xiao CJ, Zhong H, Ma RL, Su B. 2010. The ADH1B Arg47His polymorphism in east Asian populations and expansion of rice domestication in history. BMC Evol Biol. 20: 10-15.

Ranere AJ, Piperno DR, Holst I, Dickau R, Iriarte J. 2009. The cultural and chronological context of early Holocene maize and squash domestication in the Central Balsas River Valley, Mexico. Proc Natl Acad USA 106: 5014-5018.

Raymond J S and. Deboer W R (2006) Maize on the move. In: Staller J E, Tykot R H and Benz B F, editors. Histories of maize. San Diego: Academic Press. p 337 -341.

Richerson PJ, Boyd R, Henrich J. 2010. Gene-culture coevolution in the age of genomics. Proc Natl Acad Sci U S A 107: 8985-8992.

Salzano FM, 1957. The blood groups of South American Indians. Am J Phys Anthropol. 15: 555-579.

Salzano FM. 2006. The prehistoric colonization of the America. In: Anthropological genetics: theory, methods and applications. Crawford M. (editor). Cambridge University Press, Cambridge, UK.

Salzano FM and Callegari-Jacques SM. 1988. South American Indians: a case study in evolution. Oxford Science Publications. Oxford University Press, New York USA. 259p.

Santos FR, Pandya A, Tyler-Smith C, Pena SD, Schanfield M, Santos FR, Rodriguez-Delfin L, Pena SD, Moore J, Weiss KM. 1996. North and South Amerindians may have the same major founder Y chromosome haplotype.Am J Hum Genet 58:1369-1370.

Schwarcz HP. 2006. Stable carbon isotope analysis and human diet: A synthesis.. In: Staller J E, Tykot R H and Benz B F, editors. Histories of maize. San Diego: Academic Press. p.315-324.

Slatkin M, Rannala B. 2000. Estimating allele age. Annu Rev Genom Hum Genet 1: 225-249.

Simons K, Toomre D. 2000 Lipid rafts and signal transduction. Nat Rev Mol Cell Biol 1: 31–39.

Steckel R H, Rose J C., Larsen C S and Walker P L .2002. Skeletal health in the western hemisphere from 4000 B.C. to the present. Evolut Anthropol 11:142-155.

Tarazona-Santos E, Carvalho-Silva DR, Pettener D, Luiselli D, De Stefano GF, Labarga CM, Rickards O, Tyler-Smith C, Pena SD, Santos FR. 2001. Genetic differentiation in South Amerindians is related to environmental and cultural diversity: evidence from the Y chromosome. Am J Hum Genet 68:1485-1496.

Tarazona-Santos E, Santos FR. 2002. The peopling of the Americas: a second major migration? Am J Hum Genet. 70:1377-1380.

Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, Silverman JS, Powell K, Mortensen HM, Hirbo JB, Osman M, Ibrahim M, Omar SA, Lema G, Nyambo TB, Ghori J, Bumpstead S, Pritchard JK, Wray GA, Deloukas P. 2007. Convergent adaptation of human lactase persistence in Africa and Europe. Nat Genet 39: 31-40.

Tovo-Rodrigues L, Callegari-Jacques SM, Petzl-Erler ML, Tsuneto L, Salzano FM, Hutz MH. 2010. Dopamine receptor D4 allele distribution in Amerindians: A reflection of past behavior differences? Am J Phys Anthropol [Epub ahead of print].

Tykot RH, Burger RL, Van der Merwe N. 2006. The importance of maize in initial period and early horizon Peru. In: Staller J E, Tykot R H and Benz B F, editors. Histories of maize. San Diego: Academic Press. p.187-233.

Wang J, Burnett JR, Near S, Young K, Zinman B, Hanley AJ, Connelly PW, Harris SB, Hegele RA. 2000. Common and rare ABCA1 variants affecting plasma HDL cholesterol. Arterioscler. Thromb. Vasc. Biol., 20, 1983–1989.

Wang S, Lewis Jr 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.

Williamson SH, Hubisz MJ, Clark AG, Payseur BA, Bustamante CD, Nielsen R. 2007. Localizing recent adaptive evolution in the human genome. PLoS Genet 3: e90.

Young JH, Chang YP, Kim JD, Chretien JP, Klag MJ, Levine MA, Ruff CB, Wang NY, Chakravarti A. 2005. Differential susceptibility to hypertension is due to selection during the out-of-Africa expansion.PLoS Genet. 1: e82.

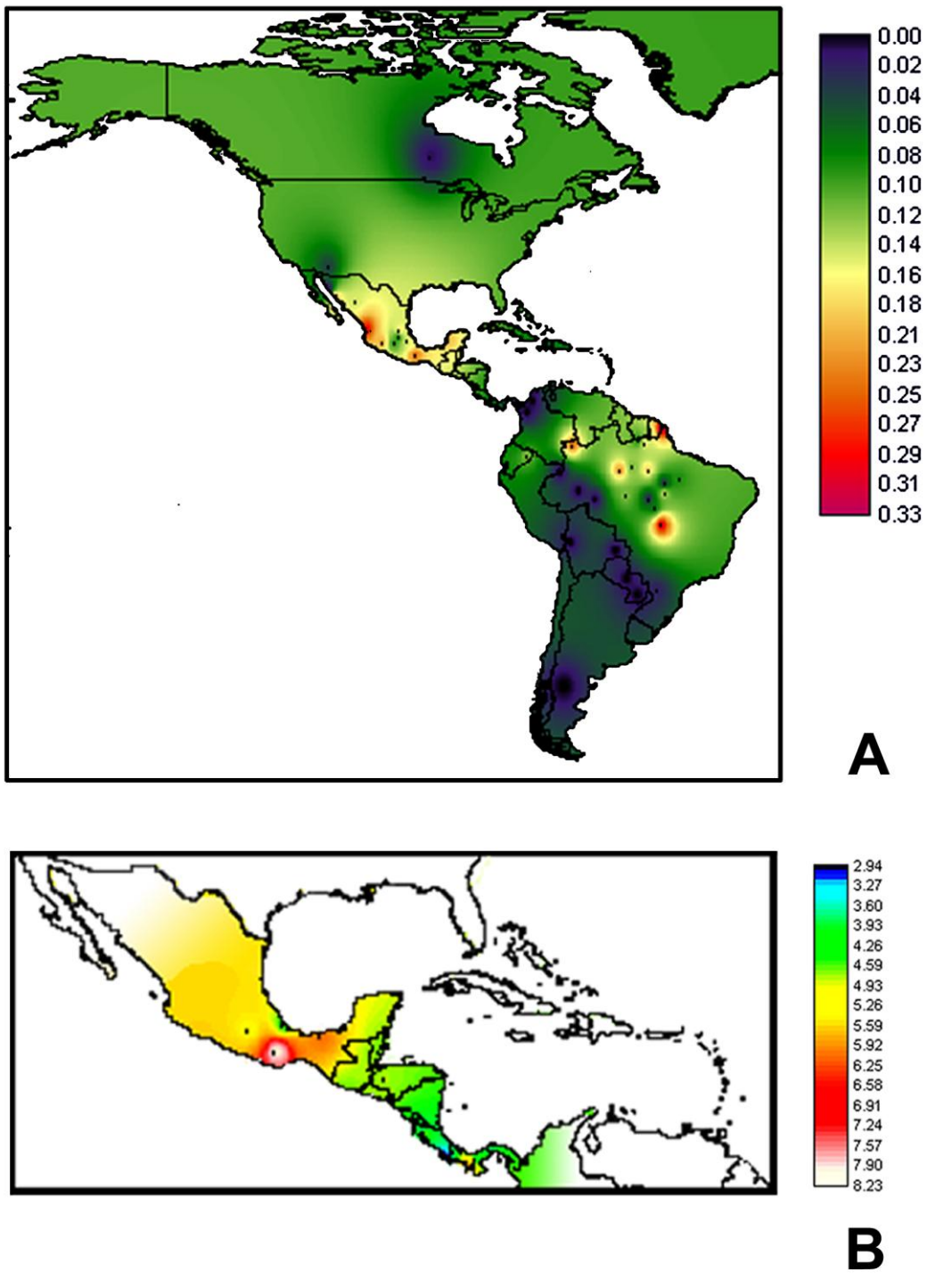


Fig. 1 (A). *ABCA1*230Cys* distribution in the Americas. **(B)** *Zea* pollen relics based on Blake (2006).

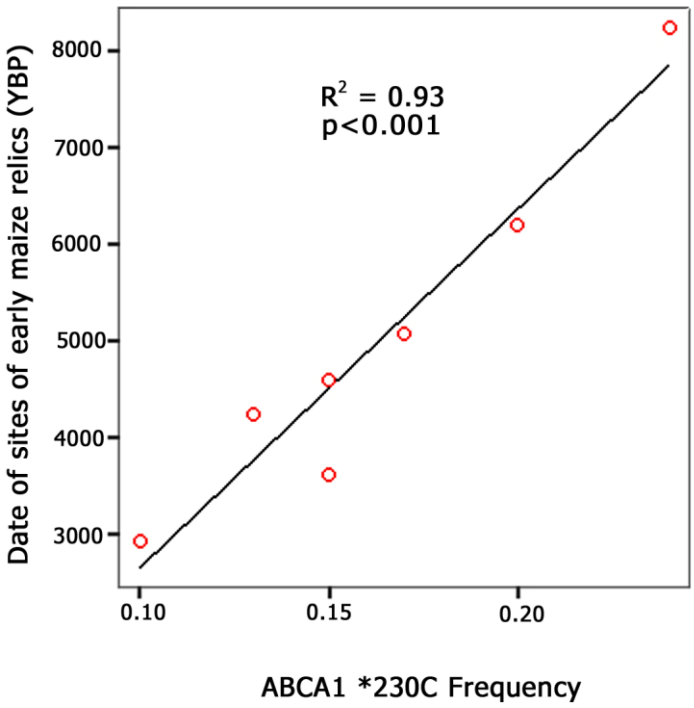


Fig. 2. *ABCA1**230Cys frequencies *versus* radiocarbon ages of maize domestication (*Zea* pollen relics; Blake, 2006). The dispersion was adjusted with linear regression analysis conducted with the use of SPSS15.0, and the statistical significance was accessed by the ANOVA test.

TABLE 1. Geographic location and linguistic classification of the Native American populations with Arg230Cys data available

Population	Country	Geographic location	Linguistic classification	Reference for information about *Arg230Cys* polymorphism
Native American				
North America and Mesoamerica				
Cora	Mexico	22° 3' N 104° 55' W	Uto-Aztecan/Southern Uto-Aztecan/Corachol-Aztecan/Cora-Huichol/Cora[1]	Acuña-Alonzo et al., 2010
Eskimo	Canada		Eskimo-Aleut/Eskimo/Inuit-Inupiaq[1]	Acuña-Alonzo et al., 2010
Mayan	Mexico	20°13' N 90°28' W	Mayan/Yucatecan[2]	Acuña-Alonzo et al., 2010
Mazahua	Mexico	19° 26' N 100° 00' W	Otomanguean/Western Otomanguean/Oto-Pame-Chinantecan/Oto-Pamean/Mazahua[1]	Acuña-Alonzo et al., 2010
Nahua	Mexico	19° 58' N 97° 37' W	Uto-Aztecan/Southern Uto-Aztecan/Corachol-Aztecan/Nahuan/Core Nahua/Nahuatl[1]	Acuña-Alonzo et al., 2010
Oji-Cree	Canada	53° 04' N 93° 19' W	Algic/Algonquian/Ojibwa-Potawatomi and Algic/Algonquian/Cree-Montagnais	Wang et al., 2000
Otomí	Mexico	20° 28' N 99° 13' W	Otomanguean/Western Otomanguean/Oto-Pame-Chinantecan/Oto-Pamean/Otomi[1]	Acuña-Alonzo et al., 2010
Pima	USA	33° 10' N 111° 52' W	Uto-Aztecan/Southern Uto-Aztecan/Pimic/Pima-Papago[1]	Acuña-Alonzo et al., 2010
Purepecha	Mexico	19° 36' N 102° 14' W	Tarascan/Purepecha[2]	Acuña-Alonzo et al., 2010
Seri	Mexico	29°00' N 112°09' W	Isolate[1]	Acuña-Alonzo et al., 2010
Tarahumara	Mexico	26° 49' N 107° 04' W	Uto-Aztecan/Southern Uto-Aztecan/Taracahitic/Tarahumaran/Tarahumara[1]	Acuña-Alonzo et al., 2010
Teenek	Mexico	21° 36' N 98° 58' W	Mayan/Huastecan/Huastec, San Luís Potosí[2]	Acuña-Alonzo et al., 2010
Totonac	Mexico	19° 57' N 97° 44' W	Totonacan/Totonac[2]	Acuña-Alonzo et al., 2010
Yaqui	Mexico	27° 29' N 110° 40' W	Uto-Aztecan/Southern Uto-Aztecan/Taracahitic/Cahitan[1]	Acuña-Alonzo et al., 2010
Zapotec	Mexico	17°14' N 96°14' W	Otomanguean/Eastern Otomanguean/Popolocan-Zapotecan/Zapotecan[1]	Acuña-Alonzo et al., 2010
South America				
Aché	Paraguay	23° S 58° W	Tupian stock/Tupi-Guaraní family/Guaraní group/ Guajaki[1]	Acuña-Alonzo et al., 2010
Apalaí	Brazil	01°20' N 54°40' W	Cariban/Central branch/Apalai[1]	Acuña-Alonzo et al., 2010
Arara	Brazil	03° 30' S 54°10' W	Cariban/South Amazonian branch/Arara group/Arara-Pariri[1]	Acuña-Alonzo et al., 2010
Aymara	Bolivia	16°30' S 68°9' W	Aymaran/Aymara[1]	Present study
Aymara	Chile	22° S 70° W	Aymaran/Aymara[1]	Present study
Ayoreo	Paraguay	16-22° S 58-63° W	Zamucoan/Ayoreo[1]	Acuña-Alonzo et al., 2010
Baniwa (Içana-River)	Brazil	01° N 67° 50' W	Arawakan/Maipuran/Northern Maipuran/Inland[2]	Acuña-Alonzo et al., 2010
Chilote	Chile	42°30' S 73°55' W	They have lost their native language	Present study
Emberá	Colombia	7° N 76° W	Chocoan/Emberá Group[1]	Present study

Cont.

TABLE 1. Cont.

Population	Country	Geographic location	Linguistic classification	Reference for information about *Arg230Cys* polymorphism
Gorotire	Brazil	07° 44' S 51° 10' W	Jean/Northern branch/Kayapó[1]	Acuña-Alonzo et al., 2010
Guaraní	Brazil	25° 20' S 52° 30' W	Tupian stock/Tupi-Guaraní family/Guaraní group/Guaraní language[1]	Acuña-Alonzo et al., 2010
Huilliche	Chile	41° S 73° W	Araucanean/Huilliche[2]	Present study
Ingano	Colombia	1° N 77° W	Quechuan/Quechua II/B/Inga, Jungle[2]	Present study
Jamamadí	Brazil	07° 15' S 66° 41' W	Arauan/Jamamadí language area/Jamamadí[1]	Acuña-Alonzo et al., 2010
Karitiana	Brazil	08° 45' S 63° 51' W	Tupian stock/Tupí-Guaraní family/Arikem branch/Karitiana[1]	Acuña-Alonzo et al., 2010
Kichwa	Ecuador	0°53' S 76°05' W	Quechuan/Quechua II/B/Quichua, Tena Lowland[2]	Acuña-Alonzo et al., 2010
Kogi	Colombia	11° N 74° W	Chibchan/Chibchan B/Eastern Chibchan/Colombian subgroup/Arhuacan group/Cágaba/Kogi[1]	Present study
Kuben-Kran-Kegn	Brazil	08°10' S 58°8' W	Jean/Northern branch/Kayapó[1]	Acuña-Alonzo et al., 2010
Lengua	Paraguay	23° S 56° W	Mascoyan/Lengua[1]	Acuña-Alonzo et al., 2010
Mapuche	Chile	40° 30' S 69° 20' W	Araucanian/Mapudungun[2]	Acuña-Alonzo et al., 2010
Mekranoti	Brazil	08° 40' S 54° W	Je stock/Jean/Northern branch/Kayapó[1]	Acuña-Alonzo et al., 2010
Mura	Brazil	03°34' S 59° 12' W	Muran/Pirahã[1]	Acuña-Alonzo et al., 2010
Pakaás-Novos	Brazil	11° 08' S 65° 05' W	Chapacuran/Wari/Orowari[1]	Acuña-Alonzo et al., 2010
Palikur	French Guiana	4° N 51° 45' W	Maipurean/Northern Division/Eastern branch/Palikur language area/Palikur[1]	Present study
Parakatejê (Gavião)	Brazil	05º 03' S 48º 36' W	Macro-Ge/Ge-Kaingang/Ge/Northwest/Timbira[2]	Acuña-Alonzo et al., 2010
Quechua	Bolivia	14°30' S 69° W	Quechuan/Quechua II/C/Quechua, South Bolivian[2]	Present study
Sateré-Mawé	Brazil	03° S 57° W	Tupian Stock/Mawé-Sateré[1]	Acuña-Alonzo et al., 2010
Ticuna	Colombia	3° 53'° S 70°W	Language isolate[2]	Present study
Tirió	Brazil	01° 57' N 55°49 W	Cariban/Tiriyó group/Tiriyó subgroup/Tirió[1]	Acuña-Alonzo et al., 2010
Txukahamãe	Brazil	10° 20' S 53° 5' W	Jean/Northern branch/Kayapó[1]	Acuña-Alonzo et al., 2010
Wayuu	Colombia	11° N 73° W	Arawakan/Maipuran/Northern Maipuran/Caribbean[2]	Present study
Xavante	Brazil	13° 20' S 51° 40' W	Jean/Central branch/Xavante[1]	Acuña-Alonzo et al., 2010
Xikrin	Brazil	05°55' S 51°11' W	Jean/Northern branch/Kayapó[1]	Acuña-Alonzo et al., 2010
Yanomámi	Brazil	02°30' – 04°30' N 64° W	Yanomam/Yanomámi[2]	Acuña-Alonzo et al., 2010
Zenu	Colombia	9° N 75° W	Choco/Embera/Northern/Emberá-Catío[2]	Present study

[1]According to Campbell (1997); [2]According to Lewis (2009).

TABLE 2. *Genotype and allele frequencies for the Ar230Cys ABCA1 polymorphism in Native Americans*

Population	*N*	Genotype frequency			Allele frequency		Reference
		Arg230Arg	*Arg230Cys*	*Cys230Cys*	*Arg230*	*230 Cys*	
Native American							
North America and Mesoamerica							
Cora	123	62	51	10	0.71	0.29	Acuña-Alonzo et al., 2010
Eskimo	30	30	0	0	1.00	0.00	Acuña-Alonzo et al., 2010
Mayan	110	68	39	3	0.80	0.20	Acuña-Alonzo et al., 2010
Mazahua	83	68	15	0	0.91	0.09	Acuña-Alonzo et al., 2010
Nahua	267	185	73	9	0.83	0.17	Acuña-Alonzo et al., 2010
Oji-Cree	80	78	2	0	0.99	0.01	Wang et al., 2010
Otomí	42	35	7	0	0.92	0.08	Acuña-Alonzo et al., 2010
Pima	2563	2364	191	8	0.96	0.04	Acuña-Alonzo et al., 2010
Purepecha	35	22	11	2	0.79	0.21	Acuña-Alonzo et al., 2010
Seri	87	87	0	0	1.00	0.00	Acuña-Alonzo et al., 2010
Tarahumara	109	81	23	5	0.85	0.15	Acuña-Alonzo et al., 2010
Teenek	67	45	20	2	0.82	0.18	Acuña-Alonzo et al., 2010
Totonac	113	86	24	3	0.87	0.13	Acuña-Alonzo et al., 2010
Yaqui	45	30	11	4	0.79	0.21	Acuña-Alonzo et al., 2010
Zapotec	125	71	50	4	0.76	0.24	Acuña-Alonzo et al., 2010
South America							
Aché	23	23	0	0	1.00	0.00	Acuña-Alonzo et al., 2010
Apalaí	22	15	7	0	0.84	0.16	Acuña-Alonzo et al., 2010
Arara	24	15	9	0	0.81	0.19	Acuña-Alonzo et al., 2010
Aymara (Bolivia)	16	16	0	0	1.00	0.00	Present study
Aymara (Chile)	22	20	2	0	0.95	0.05	Present study
Ayoreo	30	30	0	0	1.00	0.00	Acuña-Alonzo et al., 2010
Baniwa	19	13	3	3	0.76	0.24	Acuña-Alonzo et al., 2010
Chilote	2	2	0	0	1.00	0.00	Present study
Emberá	3	3	0	0	1.00	0.00	Present study
Gorotire	7	6	0	1	0.86	0.14	Acuña-Alonzo et al., 2010
Guaraní	30	29	1	0	0.98	0.02	Acuña-Alonzo et al., 2010
Huilliche	13	10	3	0	0.89	0.11	Present study
Ingano	6	5	1	0	0.92	0.08	Present study
Jamamadí	26	26	0	0	1.00	0.00	Acuña-Alonzo et al., 2010
Karitiana	20	20	0	0	1.00	0.00	Acuña-Alonzo et al., 2010
Kichwa	79	60	18	1	0.88	0.12	Acuña-Alonzo et al., 2010
Kogi	03	3	0	0	1.00	0.00	Present study
Kuben-Kran-Kegn	17	13	4	0	0.88	0.12	Acuña-Alonzo et al., 2010
Lengua	29	29	0	0	1.00	0.00	Acuña-Alonzo et al., 2010
Mapuche	40	40	0	0	1.00	0.00	Acuña-Alonzo et al., 2010
Mekranoti	25	24	1	0	0.98	0.02	Acuña-Alonzo et al., 2010
Mura	18	11	6	1	0.78	0.22	Acuña-Alonzo et al., 2010
Pakaás-Novos	25	23	2	0	0.96	0.04	Acuña-Alonzo et al., 2010
Palikur	3	1	2	0	0.67	0.33	Present study
Parakatejê (Gavião)	78	65	12	1	0.91	0.09	Acuña-Alonzo et al., 2010
Quechua	16	15	1	0	0.97	0.03	Present study
Sateré-Mawé	25	20	4	1	0.88	0.12	Acuña-Alonzo et al., 2010

Cont.

TABLE 2. Cont.

Population	N	Genotype frequency			Allele frequency		Reference
		Arg230Arg	*Arg230Cys*	*Cys230Cys*	*Arg230*	*230 Cys*	
Ticuna	1	1	0	0	1.00	0.00	Present study
Tirió	25	21	4	0	0.92	0.08	Acuña-Alonzo et al., 2010
Txukahamãe	30	26	4	0	0.93	0.07	Acuña-Alonzo et al., 2010
Wayuu	17	15	2	0	0.94	0.06	Present study
Xavante	21	10	9	2	0.69	0.31	Acuña-Alonzo et al., 2010
Xikrin	17	16	1	0	0.97	0.03	Acuña-Alonzo et al., 2010
Yanomámi	25	20	4	1	0.88	0.12	Acuña-Alonzo et al., 2010
Zenu	4	4	0	0	1.00	0.00	Present study

TABLE 3. *Proportion of populations and their respective 230Cys*ABCA1 allele frequencies*

Geographical region	No. of populations	N	Proportion of populations and their respective *230Cys*ABCA1* allele frequencies (%)			
			0	0.01 -0.05	0.06 -0.15	> 0.15
North America (Canada and USA)	3	2673	33	66	0	0
Mesoamerica/Central America	12	1206	8.3	0	8.3	83.4
South America (Highlands: Andes and surroundings)	7	115	42.8	28.6	28.6	0
South America (Lowlands: Amazon/ Brazilian Central Plateau/Chaco)	28	646	32.1	14.3	32.1	21.5
Total	50	4640				

North America: Eskimo, Oji-Cree and Pima; Mesoamerica/Central America: Cora, Maya, Mazahua, Nahua, Otomí, Purepecha, Seri, Tarahumara, Teenek, Totonac, Yaqui, Zapotec; South America (Highlands): Mapuche, Aymara, Quechua, Chilote, Hulliche and Ingano; South America (Lowlands): Ache, Apalaí, Arara, Ayoreo, Embera, Gorotire, Guarani, Baniwa (Içana-River), Jamamadi, Karitiana, Kichwas, Kogi, Kuben-Kran-Keng, Lengua, Mekranoti, Mura, Pacaás-Novos, Palicur, Parkatejê (Gavião), Sateré-Mawé, Ticuna, Tiryió, Txukahamãe, Wayuu, Xavante, Xicrin, Yanomama, Zenu.

TABLE S1. *Zea* pollen relics, populations, age and *230Cys*ABCA1* frequencies used for the regression analysis

Zea pollen relics	Population[1]	Radiocarbon years BP	Calendar years BP[2]	*230Cys*ABCA1*
Oaxaca	Zapotec	8240	9212	0.24
Tabasco	Maya	6208	7122	0.20
Mexico state	Nahuatl	5090	5835	0.17
Guatemala	Kaqchikel-Quiche[3]	4600	5318	0.15
Veracruz	Totonac	4250	4818	0.13
Costa Rica/Panamá	Guaymi[3]	3630	3943	0.15
Costa Rica	Cabecar[3]	2940	3096	0.10

[1]Located near of the archeological sites of *Zea* pollen relics; [2]Conversion according http://www.radiocarbon.ldeo.columbia.edu/research/radcarbcal.htm; [3]Unpublished data (A. Ruiz-Linares et al., personal communication).

DISCUSSÃO GERAL

Como as discussões específicas relativas a cada um dos temas desta tese já foram abordadas nos artigos apresentados, este item será reservado somente para a apresentação de um enfoque mais geral sobre os assuntos estudados.

O primeiro passo para que um gene seja indicado como um bom candidato para estudos que visam desvendar a relação deste com alguma características fenotípica, é avaliar se há alguma evidência funcional de que o mesmo está envolvido com o característica em consideração. Desta forma, é possível especular que genes envolvidos em situações patológicas, como é o caso dos *FGFRs e TCOF1,* por exemplo, sejam bons candidatos para estudos que visam identificar as causas moleculares que levam as variações normais encontradas dentro e entre espécies. Outra especulação possível e bastante utilizada é a extensão da homologia de uma espécie para outra, ou seja, assumir em um primeiro momento que uma alteração numa seqüência de DNA de uma espécie que é responsável por determinado fenótipo, poderia ser responsável por um fenótipo similar em outra espécie.

Neste contexto de homologia de genes em diferentes espécies, o *TCOF1* poderia ser um forte candidato para revelar alterações craniofaciais em primatas, e até mesmo dentro da espécie humana, já que Haworth et al. (2005) demonstraram que uma única substituição no éxon 4 (*C396T, Pro117Ser*) deste gene resultaria em alterações intra-específicas em cães. O alelo T foi significativamente associado a raças que apresentam braquicefalia (e.g. Bulldog Francês e Boxer). O artigo de Haworth e colegas tem sido desde sua publicação citado como um ótimo exemplo de associação de uma variante

funcional com uma característica morfológica definida (ver como exemplo a revisão de Wayne e Ostrander, 2007). Deste modo, era possível especular que entre primatas o mesmo poderia estar envolvido na retração dos ossos da face, e entre populações humanas no fenótipo braquifacial encontrados em diversas populações nativo-americanas.

A partir destas premissas iniciamos o seqüenciamento do éxon 4 em diferentes espécies de primatas, nativo-americanos, europeus, africanos e esquimós. Concomitante a isso, replicamos o estudo de Harworth et al. (2005) com cães para assegurar que a associação encontrada pelos autores não era espúria, pois o número amostral usado por eles era pequeno e seus achados não haviam ainda sido replicados.

Os resultados referentes aos cães se encontram detalhados em Hünemeier et al. (2009) mostraram claramente que embora o polimorfismo possa estar envolvido nos mecanismos de variação morfológica da face, sendo encontrado com mais freqüência em cães braquicéfalos do que meso e dolicocéfalos, esta associação não é direta como descrita por Harworth et al. (2005). Foram encontrados em nosso estudo cães dolicocéfalos homozigotos para o alelo *T/Ser,* bem como cães braquicéfalos homozigotos para o alelo *C/Pro.* Considerando a intricada rede de regulação e expressão deste gene, a mutação por nós estudada pode ser co-dependente de outras variantes na mesma rede de interação gênica (Figura 4). Este resultado salienta o que já foi comentado anteriormente, que características morfológicas de natureza complexa devem sempre que possível ser estudadas num contexto genético de rede para que a relação genótipo-fenótipo seja melhor compreendida.

Dentro da ordem *Primates,* uma característica morfológica interessante é a evidente diferenciação facial entre as subordens *Platyrrihni* (Macacos do Novo Mundo) e

Catarrihni (Macacos do Velho Mundo). Enquanto os macacos do Novo Mundo apresentam um nariz mais achatado com as fossas nasais em posição lateral, os macacos do velho mundo apresentam um nariz mais longo com as fossas nasais na extremidade frontal do mesmo, situadas lado a lado (Hershkovitz, 1977). Com o objetivo de avaliar se a variante *T/117Ser* encontrada em cães estava presente em primatas, buscamos nos bancos de dados disponíveis seqüências do éxon 4 do gene *TCOF1* nas espécies *Pan throglodytes*, *Pongo pigmeus*, *Macaca mulatta* e *Callithrix jacchus*, bem como em *Homo neandertalensis*. Adicionalmente seqüenciamos um espécime de *Gorilla gorilla*, além de quinze nativo-americanos, cinco europeus, e seis esquimós (Hunemeier et al., dados não publicados). Nenhuma variação no exon 4 foi encontrada nas diferentes populações humanas, bem como na seqüência no *Homo neandertalensis*. No entanto, a espécie de macaco do novo mundo *Callitrix jacchus* apresentava três substituições quando comparada aos demais, todas elas promovendo alterações de aminoácidos. *Pongo pigmeus* e *Macaca mulatta* apresentavam, por sua vez, uma alteração de aminoácido cada em sítios diferentes, mas nenhum dos primatas testados aqui apresentava alteração no sítio descrito por Haworth et al. (2005) em cães (éxon 4, *C396T, Pro117Ser*).

Para avaliar a possível presença de seleção sobre os sítios variáveis, testamos as sete espécies (*Homo sapiens sapiens*, *Homo neandertalensis*, *Pan throglodytes*, *Pongo pigmeus*, *Macaca mulatta*, *Callithrix jacchus* e *Gorilla gorilla*) com o programa *CODEML*, parte do pacote *PAML_4*, que estimou-se as taxas de substituições sinônimas e não-sinônimas (*dN/dS* ou ω; Yang and Nielsen, 2002). Dentro da ordem *Primate* não encontramos nenhum indício de seleção (ω=0,56), no entanto, quando a mesma análise é replicada incluindo outras espécies de mamíferos (*Equus caballus*, *Canis Familiaris*, *Mus musculos*, *Cavia porcellus* e *Lexodonta africana*) são detectados sítios com sinais de

possível seleção positiva, ou relaxamento de seleção, ω=1.03697. Estas análises preliminares parecem integrar-se as nossas inferências em cães, evidenciando algum papel deste gene na morfogênese do plano facial em mamíferos (Hünemeier et al., dados não publicados). Embora algumas alterações possam ser neutras, em alguns ramos filogenéticos (*Homo*, por exemplo) nenhuma alteração foi detectada (evidenciando seleção purificadora) enquanto em outros as alterações não podem ser facilmente explicadas pelo modelo neutro.

Com o *enhancer HACNS1,* por outro lado, foi possível demonstrar que a ação da seleção positiva na linhagem *Homo* foi responsável pela fixação das mutações específicas encontradas nesta região quando comparada com o gênero *Pan*. A ausência de variação intra-específica no *Homo sapiens* e no *Neandertal* reforça o papel funcional do *enhancer*, mantendo inalterado posteriormente pela ação extrema de seleção purificadora. Como discutido no artigo, é possível que o mesmo esteja envolvido na regulação de genes que caracterizam morfologicamente nosso gênero, tais como postura ereta e destreza das mãos (Hünemeier et al. 2010). Recentemente através de comunicação pessoal Shyam Prabhakar manifestou sua satisfação com nossos achados que demonstraram que a constrição funcional do *enhancer HACNS1* não se limita a nossa espécie mas também a de outro hominídeo.

Nossos estudos com o gene *FGFR1,* por sua vez, demonstram mais uma vez como um gene pode contribuir de maneira sutil no desenvolvimento de fenótipos complexos, e como este deve ser avaliado num contexto de rede de genes do desenvolvimento. Coussens e Daals (2005), descreveram a associação entre uma mutação no *FGFR1* (alelo C do polimorfismo rs4647905) com dolicocefalia em humanos. Nossos achados concordam

parcialmente com estes achados. Encontramos uma visível correlação entre a redução do índice cefálico em indivíduos homozigotos para o alelo C nas três populações estudadas (Totonacos, miscigenados de Tepango e miscigenados da Cidade do México), sendo que a única associação estatisticamente significativa foi na população miscigenada de Tepango. Além disto, encontramos uma associação entre o índice jugo-mandibular e o polimorfismo rs3212849 nas populações miscigenadas. A correlação entre índice jugo-mandibular e o alelo T deste polimorfismo é evidente ainda quando não há associação entre as variáveis.

Um ponto importante visto em nossas análises é a forte correlação entre etnicidade/origem populacional e os índices estudados, sendo a mesma completamente ajustada quando se avalia o polimorfismo rs3213849, e com ajuste intermediário entre índice cefálico e o polimorfismo rs4647905. Há uma integração morfológica entre dois *Tag-haplotype-SNPs* em populações com algum componente europeu, sendo que os dois se associam a características craniofaciais diferentes e não relacionadas de acordo com nossa análise multivariada. Já em nativos americanos, parece não haver dois diferentes blocos em desequilíbrio de ligação dentro do gene *FGFR1*. Neste grupo populacional foi encontrado um único e maior bloco em desequilíbrio de ligação abrangendo todos os SNPs em um único haplótipo. Tendo-se isto em vista, ao se avaliar estes mesmos *Tag-haplotype-SNPs* do gene *FGFR1* em populações sem caracterização morfológica, se torna mais evidente a diferenciação do padrão de distribuição dos haplótipos e ligação entre os polimorfismos. De modo geral, parece que o *background* genético de cada população atua como fator crucial na determinação de características craniofaciais, mas como se tratam de *SNPs* sinalizadores herdados conjuntamente com outros *SNPs* em outras regiões gênicas, não se pode inferir diretamente uma associação genótipo-fenótipo, e sim uma associação entre

largas regiões que provavelmente carregam algum alelo funcional e uma variação de fenótipo.

Diferentes populações apresentam histórias demográficas diferentes, tendo sido sujeitas a diferentes processos e fatores evolutivos que levaram a diferenciação de seu *background* genético. Estas diferenças podem ser resultados tanto de processos evolutivos casuais (deriva) quanto direcionais (seleção natural). Deste modo, mudanças benéficas podem surgir em algum contexto ecológico ou cultural específico levando ao aumento desta característica na população ao longo das gerações.

Os resultados preliminares também envolvendo estudos com genótipo-fenótipo apresentados no apêndice II com os indígenas Xavantes e outros, dentro deste contexto, são um exemplo ilustrativo sobre como hábitos culturais (brigas e alianças entre grupos promovendo fissões e fusão, respectivamente) podem catalisar processos micro-evolutivos. Neel e Salzano (1967) foram os primeiros a descrever o fenômeno, chamado de modelo de fusão-fissão. No caso especifico envolvendo os Xavante, o resultado pode ter sido uma rápida e marcante modificação morfológica. A separação entre os Gê e outros grupos é estimado com dados lingüísticos em cerca de 3.000 anos antes no presente, sendo que o último *split*, entre Xavante e Kayapó não teria mais do que 1.500 anos. Nossas análises preliminares (Apêndice II; Hünemeier et al., dados não publicados) demonstram que tanto com dados moleculares (mtDNA e marcadores não-neutros) quanto com medidas craniométricas, o *background* biológico dos Xavante em menos de dois mil anos se alterou de maneira singular considerando o padrão nativo-americano devido à fenômenos microevolutivos, em especial a deriva genética, catalisados por processos culturais, como aqueles que promovem fissão e fusão na propagação das aldeias. Com base nisto se

estabelece um argumento favorável ao modelo de povoamento da América proposto por González-José et al (2008), pois tensões sociais que levariam a rupturas com posterior isolamento de alguns grupos humanos seriam o suficiente para gerar uma amplitude de diversidade genotípica e fenotípica em poucas gerações. Até este momento, este seria o primeiro exemplo bem documentado de diferenciação morfológica rápida associada aos nativos americanos. Como conseqüência, temos várias implicações, como por exemplo, não se faz necessário qualquer postulação sobre migrações diversas com indivíduos com distintos *background* biológicos. Uma mesma origem, a partir de um grupo ancestral relativamente heterogêneo geneticamente, e com posterior ação de fatores micro-evolutivos e culturais autóctones podem parcimoniosamente explicar as muitas vezes surpreendentes variações morfológicas encontradas entre os diferentes povos nativos americanos.

Outro importante fator cultural catalisador de fenômenos micro-evolutivos foi a adoção da agricultura e domesticação de animais durante o Neolítico. Como visto no artigo "*Gene-culture dynamics: an example involving Native Americans*" a implementação da agricultura pode ser um processo dispendioso, que pode levar a déficits nutricionais pois ocorre um relativamente rápido aumento da população, dependentes muitas vezes de uma ou poucas plantas cultivadas. Isso poderia ter levado a ciclos de fartura e fome, favoráveis a seleção de alelos "econômicos" como é o caso do *ABCA1*230Cys*. Tal panorama de co-evolução ilustra como a cultura sob diferentes cenários é agente ativo na evolução humana.

Vale destacar ainda, que a variante *ABCA1*230Cys* além de ser a primeira variante genética exclusiva nativo-america funcional descrita, constitui-se no primeiro exemplo bem documentado de co-evolução gene-cultura dentro da América.

Finalmente, considerando tudo o que foi exposto, torna-se evidente que as relações fenótipo-genótipo são influenciadas por diversos fatores biológicos/culturais/ambientais. O mais importante é assumir que a variação existente dentro da espécie humana, é resultado de uma intricada rede de interações genéticas, que atuam em todos os níveis do desenvolvimento. Adicionalmente, as interações desta rede no *Homo sapiens sapie*ns se distingue daquelas de outros mamíferos, em especial pela ação catalisadora de processos derivados justamente da característica que torna nossa espécie singular, a cultura.

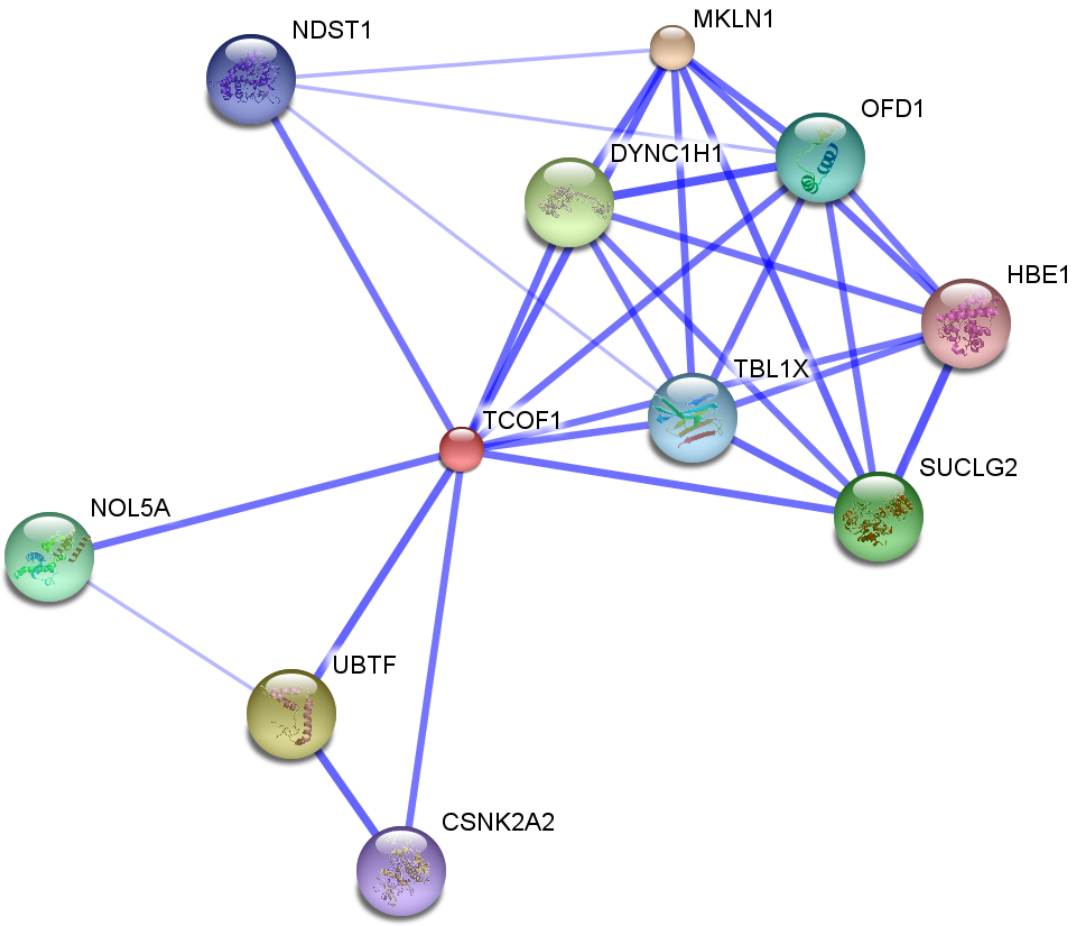


Figura 4. Onze genes da rede do desenvolvimento craniofacial . Imagem criada usando o programa STRING (Search Tool for the Retrieval of Interacting Genes/P Genes/Proteins - fttp://string.embl.de) considerando o maior nível de confiança (0,90). Diferentes espessuras dos ramos representam o nível de confiança.

REFERÊNCIAS BIBLIOGRÁFICAS

Acuña-Alonzo, V., Flores-Dorantes, T., Kruit, J. K., Villarreal-Molina, T., Arellano-Campos, O., Hünemeier, T., et al. (2010). A functional ABCA1 gene variant is associated with low HDL-cholesterol levels and shows evidence of positive selection in Native Americans. *Human Molecular Genetics*, *19*(14), 2877-2885.

Albuisson, J., Pêcheux, C., Carel, J., Lacombe, D., Leheup, B., Lapuzina, P., et al. (2005). Kallmann syndrome: 14 novel mutations in KAL1 and FGFR1 (KAL2). *Human mutation*, *25*(1), 98-9.

Betti, L., Balloux, F., Hanihara, T., & Manica, A. (2010). The relative role of drift and selection in shaping the human skull. *American journal of physical anthropology*, *141*(1), 76-82.

Blake, M. (2006). Dating the initial spread of *Zea mays*. In: Staller J E, Tykot R H and Benz B F, editors. Histories of maize. San Diego: Academic Press. p.55-68.

Bocquet-Appel, J.P., Bar-Yosef, O. (eds). (2008). The Neolithic demographic transition and its consequence. New York: Springer.

Coussens, A. K., & van Daal, A. (2005). Linkage disequilibrium analysis identifies an FGFR1 haplotype-tag SNP associated with normal variation in craniofacial shape. *Genomics*, *85*(5), 563-73.

Dixon, J., Jones, N. C., Sandell, L. L., Jayasinghe, S. M., Crane, J., Rey, J., et al. (2006). Tcof1/Treacle is required for neural crest cell formation and proliferation deficiencies that cause craniofacial abnormalities. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(36), 13403-8.

Enard, W., Przeworski, M., Fisher, S. E., Lai, C. S., Wiebe, V., Kitano, T., et al. (2002). Molecular evolution of FOXP2, a gene involved in speech and language. *Nature*, *418*(6900), 869-72.

Enlow, D.H. (1990) *Facial Growth*. Saunders. Philadelfia.

Francis-West, P.H., Robson, L. & Evans, D.J. (2003) Craniofacial development: the tissue and molecular interactions that control development of the head. *Advances in Anatomy, Embryololy and Cell Biolology*, 169: III-VI, 1-138.

González-José, R., Bortolini, M. C., Santos, F. R., & Bonatto, S. L. (2008). The peopling of America: craniofacial shape variation on a continental scale and its interpretation from an interdisciplinary view. *American Journal of Physical Anthropology*, *137*(2), 175-87.

Hancock, A.M., Witonsky, D.B., Ehler, E., Alkorta-Aranburu, G., Beall, C., Gebremedhin, A., Sukernik, R., Utermann, G., Pritchard, J., Coop, G. & Di Rienzo A. (2010). Human adaptations to diet, subsistence, and ecoregion are due to subtle shifts in allele frequency. *Proceedings of the National Academy of Sciences of the United States of America 107* Suppl 2:8924-8930.

Harvati, K., & Weaver, T. D. (2006). Human cranial anatomy and the differential preservation of population history and climate signatures. *The anatomical record. Part A, Discoveries in molecular, cellular, and evolutionary biology*, *288*(12), 1225-33.

Haworth, K. E., Islam, I., Breen, M., Putt, W., Makrinou, E., Binns, M., et al. (2001). Canine TCOF1; cloning, chromosome assignment and genetic analysis in dogs with different head types. *Mammalian Genome*, *12*(8), 622-629.

Hershkovitz, P. (1977). Living New World monkeys (Platyrrhini): with an introduction to primates. Vol. 1. The Univiversity of Chicago Press, Chicago.

Howard, T.D., Paznekas, W.A., Green, E.D., Chiang, L.C., Ma, N., Ortiz de Luna, R.I., Garcia Delgado, C., Gonzalez-Ramos, M., Kline, A.D. & Jabs, E.W. (1997) Mutations in TWIST, a basic helix-loop-helix transcription factor, in Saethre-Chotzen syndrome**.** *Nature Genetics* 15(1): 36-41.

Hünemeier, T., Ruiz-Linares, A., Silveira, A., Paixão-Côrtes, V. R., Salzano, F. M., Bortolini, M. C., et al. (2010). Brief communication: Population data support the adaptive

nature of HACNS1 sapiens/neandertal-chimpanzee differences in a limb expression domain. *American Journal of Physical Anthropology*, *000*(June), 1-4.

Hünemeier, T., Salzano, F. M., & Bortolini, M. C. (2009). TCOF1 T/Ser variant and brachycephaly in dogs. *Animal Genetics*, *40*(3), 357-8.

Johnson, A. W. & Earle, T., 1987, The evolution of human societies: from foraging group to agrarian state, Stanford University Press, Stanford.

Junqueira LC, Carneiro J (2004) Histologia Básica. Guanabara Koogan, Rio de Janeiro.

Kennett, D. J., Voorhies, B. & Martorana, D. (2006). An ecological model for the origins of maize-based food production on the Pacific Coast of Southern Mexico. In: Kennett D J and Winterhalder B (editors), Behavioral ecology and the transition to agriculture, Berkeley: University of California Press. p103-136.

Kim, H. J., Rice, D. P., Kettunen, P. J., & Thesleff, I. (1998). FGF-, BMP- and Shh-mediated signalling pathways in the regulation of cranial suture morphogenesis and calvarial bone development. *Development (Cambridge, England)*, *125*(7), 1241-51.

Laland, K. N., Odling-Smee, J., & Myles, S. (2010). How culture shaped the human genome: bringing genetics and the human sciences together. *Nature Reviews Genetics*, *11*(2), 137-48. Nature Publishing Group.

Lawson, H.A., Lesk, A.M., Miller, W. & Weiss, K.M. (2006). A comparative genomics approach to primate craniofacial evolution. *American Journal Physical Anthropoogy*, 129(S42).119.

Masotti, C., Armelin-Correa, L. M., Splendore, A., Lin, C. J., Barbosa, A., Sogayar, M. C., et al. (2005). A functional SNP in the promoter region of TCOF1 is associated with reduced gene expression and YY1 DNA-protein interaction. *Gene*, *359*, 44-52.

Neel, JV. (1962) . Diabetes mellitus: a ‘thrifty’ genotype rendered detrimental by ‘progress’? *American Journal of Human Genetics* 14: 353–362.

Neel, J.V. & Salzano, F.M. (1967). Further studies on the Xavante Indians. X. Some hypotheses-generalizations resulting from these studies. *American Journal of Human Genetics* 19:554–574.

Nei, M. (2007). The new mutation theory of phenotypic evolution. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(30), 12235-12242.

Passos-Bueno, M.R., Wilcox, W.R., Jabs, E.W., Sertie, A.L., Alonso, L.G. & Kitoh, H. (1999) Clinical spectrum of fibroblast growth factor receptor mutations. *Human Mutation,* 14(2): 115-25.

Paixão-Côrtes VR, Braga TS, Salzano FM, Munsdstock KS, Mundstock CA, Bortolini MC (2010a) *PAX9* and *MSX1* transcription genes in non-syndromic dental agenesis. Archives of Oral Biology (submitted).

Paixão-Côrtes VR, Meyer D, Pereira T, Maziéres S, Krishnamoorthy R, Elion J, Zago MA, Silva Jr WA, Salzano FM, Bortolini MC. (2010b) Genetic variation among major human geographic groups supports a peculiar evolutionary trend in PAX9 and its role in third molar agenesis. PLOS One (submitted).

Peng, Y., Shi, H., Qi, X.B., Xiao, C.J., Zhong, H., Ma, R.L. & Su B. (2010). The ADH1B Arg47His polymorphism in east Asian populations and expansion of rice domestication in history. *BMC Evolutionary Biolology*. 20: 10-15.

Pereira, T. V., Salzano, F. M., Mostowska, A., Trzeciak, W. H., Ruiz-Linares, A., Chies, J. a., et al. (2006). Natural selection and molecular evolution in primate PAX9 gene, a major determinant of tooth development. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(15), 5676-81.

Prabhakar, S., Visel, A., Akiyama, J. a., Shoukry, M., Lewis, K. D., Holt, A., et al. (2008). Human-specific gain of function in a developmental enhancer. *Science*, *321*(5894), 1346-50. doi: 10.1126/science.1159974.

Relethford, J. H. (2010). Population-specific deviations of global human craniometric variation from a neutral model. *American Journal of Physical Anthropology*, *142*(1), 105-11.

Rice, D.P., Aberg, T., Chan, Y., Tang, Z., Kettunen, P.J., Pakarinen, L., Maxson, R.E. & Thesleff, I. (2000) Integration of *FGF* and *TWIST* in calvarial bone and suture development. *Development* 127: 1845.

Rice, D.P., Rice, R. & Thesleff, I. (2003) Fgfr mRNA isoforms in craniofacial bone development. *Bone* 33(1): 14-27.

Richerson, P. J., Boyd, R., & Henrich, J. (2010). Colloquium paper: gene-culture coevolution in the age of genomics. *Proceedings of the National Academy of Sciences of the United States of America, 107 Suppl* , 8985-92.

Richtsmeier, J.T., Weiss, K.M., Buchanan, A., Walker, A., Jablonski, N., Frazier, B.C., Willmore, K.E., Lawson, H., Yan, P., Hill, C.A., Roseman, C.C., Cheverud, J.M., Falk, D., Hildebolt, C.F. & Rogers J. (2007). Developmental genetic basis of primate craniofacial variation and human origins. *American journal of Physical Anthropology*, 132(S44):198.

Rockman, M. V. (2008). Reverse engineering the genotype-phenotype map with natural genetic variation. *Nature*, *456*(7223), 738-44.

Roseman, C. C. (2004). Detecting interregionally diversifying natural selection on modern human cranial form by using matched molecular and morphometric data. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(35), 12824-9.

Shows, K. H., Ward, C., Summers, L., Li, L., Ziegler, G. R., Hendrickx, A. G., et al. (2006). Reduced TCOF1 mRNA level in a rhesus macaque with Treacher Collins-like syndrome: further evidence for haploinsufficiency of treacle as the cause of disease. *Mammalian Genome, 17*(2), 168-77.

Splendore, A., Fanganiello, R. D., Masotti, C., Morganti, L. S., & Passos-Bueno, M. R. (2005). TCOF1 mutation database: novel mutation in the alternatively spliced exon 6A and update in mutation nomenclature. *Human mutation*, *25*(5), 429-34.

Steckel, R. H., Rose, J. C., Spencer Larsen, C., & Walker, P. L. (2002). Skeletal health in the Western Hemisphere from 4000 B.C. to the present. *Evolutionary Anthropology: Issues, News, and Reviews*, *11*(4), 142-155.

Sutter, N. B., Bustamante, C. D., Chase, K., Gray, M. M., Zhao, K., Zhu, L., et al. (2007). A single IGF1 allele is a major determinant of small size in dogs. *Science*, *316*(5821), 112-5.

Wang, S., Lewis, C. M., Jakobsson, M., Ramachandran, S., Ray, N., Bedoya, G., et al. (2007). Genetic variation and population structure in Native Americans. *PLoS Genetics*, *3*(11), e185.

Wilkie, A.O., Morriss-Kay, G.M., Jones, E.Y. & Heath, J.K. (1995) Functions of fibroblast growth factors and their receptors. *Current Biology* 5(5): 500-7.

Yang Z, Nielsen R. (2002) Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Molecular Biology Evolution* 19(6):908-917.

APÊNDICE I

Human Molecular Genetics, 2010, Vol. 19, No. 14 **2877–2885**
doi:10.1093/hmg/ddq173
Advance Access published on April 23, 2010

A functional *ABCA1* gene variant is associated with low HDL-cholesterol levels and shows evidence of positive selection in Native Americans

Víctor Acuña-Alonzo[1,5,†], Teresa Flores-Dorantes[1,†], Janine K. Kruit[6], Teresa Villarreal-Molina[1], Olimpia Arellano-Campos[2], Tábita Hünemeier[7], Andrés Moreno-Estrada[8], Ma Guadalupe Ortiz-López[9,10], Hugo Villamil-Ramírez[1], Paola León-Mimila[1], Marisela Villalobos-Comparan[1], Leonor Jacobo-Albavera[1], Salvador Ramírez-Jiménez[1], Martin Sikora[11], Lin-Hua Zhang[6], Terry D. Pape[6], Ma de Ángeles Granados-Silvestre[9,10], Isela Montufar-Robles[9,10], Ana M. Tito-Alvarez[12], Camilo Zurita-Salinas[12], José Bustos-Arriaga[13], Leticia Cedillo-Barrón[13], Celta Gómez-Trejo[5], Rodrigo Barquera-Lozano[5], Joao P. Vieira-Filho[14], Julio Granados[3], Sandra Romero-Hidalgo[15], Adriana Huertas-Vázquez[16], Antonio González-Martín[17], Amaya Gorostiza[5,17], Sandro L. Bonatto[18], Maricela Rodríguez-Cruz[19], Li Wang[20], Teresa Tusié-Luna[1], Carlos A. Aguilar-Salinas[2], Ruben Lisker[4], Regina S. Moises[14], Marta Menjivar[9], Francisco M. Salzano[7], William C. Knowler[21], M. Cátira Bortolini[7], Michael R. Hayden[6], Leslie J. Baier[21] and Samuel Canizales-Quinteros[1,*]

[1]Unit of Molecular Biology and Genomic Medicine, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán (INCMNSZ), Universidad Nacional Autónoma de México (UNAM), Mexico City 14000, Mexico, [2]Department of Endocrinology and Metabolism, [3]Department of Immunology and Rheumatology and [4]Direction of Research, INCMNSZ, Mexico City 14000, Mexico, [5]Molecular Genetics Laboratory, Escuela Nacional de Antropología e Historia (ENAH), Mexico City 14030, Mexico, [6]Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada V6T1Z4, [7]Departamento de Genetica, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS 91501-970, Brazil, [8]Department of Genetics, Stanford University, Stanford, CA 94305, USA, [9]Department of Biology, Faculty of Chemistry, Universidad Nacional Autónoma de México (UNAM), Mexico City 04510, Mexico, [10]Laboratory of Molecular Endocrinology, Hospital Juárez de México, Mexico City, Mexico, [11]Institute of Evolutionary Biology (UPF-CSIC), CEXS-UPF-PRBB, Barcelona 37-49 08003, Spain, [12]Laboratorios Zurita & Zurita, Quito N12-120, Ecuador, [13]Department of Molecular Biomedicine, Cinvestav, IPN, Mexico City 07360, Mexico, [14]Division of Endocrinology, Department of Medicine, Federal University of São Paulo, São Paulo 04034-970, Brazil, [15]Department of Computational Genomics, Instituto Nacional de Medicina Genómica, Mexico City 01900, México, [16]Cedars-Sinai Heart Institute, Cedars-Sinai Medical Center, Los Angeles, CA 90048-5909, USA, [17]Department of Zoology and Physical Anthropology, Faculty of Biology, Complutense University of Madrid, Madrid 28040, Spain, [18]Faculty of Biological Sciences, Pontifical Catholic University of Rio Grande do Sul, Porto Alegre 90619-900 RS, Brazil, [19]Unit of Medical Research in Nutrition, Pediatrics Hospital, National Medical Center XXI Century, Mexican Institute of Social Security, Mexico City 06725, Mexico, [20]Key Laboratory of Molecular and Developmental Biology, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China, [21]National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Phoenix, AZ 85014, USA

Received January 7, 2010; Revised March 19, 2010; Accepted April 21, 2010

*To whom correspondence should be addressed at: Unidad de Biología Molecular y Medicina Genómica, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Vasco de Quiroga, No. 15 Colonia Sección 16, Tlalpan 14000, México D.F., Mexico. Tel/Fax: +52 5556550011; Email: cani@servidor.unam.mx
†These authors contributed equally to this work.

2878 Human Molecular Genetics, 2010, Vol. 19, No. 14

It has been suggested that the higher susceptibility of Hispanics to metabolic disease is related to their Native American heritage. A frequent cholesterol transporter *ABCA1* (ATP-binding cassette transporter A1) gene variant (*R230C*, rs9282541) apparently exclusive to Native American individuals was associated with low high-density lipoprotein cholesterol (HDL-C) levels, obesity and type 2 diabetes in Mexican Mestizos. We performed a more extensive analysis of this variant in 4405 Native Americans and 863 individuals from other ethnic groups to investigate genetic evidence of positive selection, to assess its functional effect *in vitro* and to explore associations with HDL-C levels and other metabolic traits. The *C230* allele was found in 29 of 36 Native American groups, but not in European, Asian or African individuals. *C230* was observed on a single haplotype, and *C230*-bearing chromosomes showed longer relative haplotype extension compared with other haplotypes in the Americas. Additionally, single-nucleotide polymorphism data from the Human Genome Diversity Panel Native American populations were enriched in significant integrated haplotype score values in the region upstream of the *ABCA1* gene. Cells expressing the *C230* allele showed a 27% cholesterol efflux reduction ($P < 0.001$), confirming this variant has a functional effect *in vitro*. Moreover, the *C230* allele was associated with lower HDL-C levels ($P = 1.77 \times 10^{-11}$) and with higher body mass index ($P = 0.0001$) in the combined analysis of Native American populations. This is the first report of a common functional variant exclusive to Native American and descent populations, which is a major determinant of HDL-C levels and may have contributed to the adaptive evolution of Native American populations.

INTRODUCTION

It has been suggested that genetic susceptibility of Hispanics to type 2 diabetes (T2D), obesity and dyslipidemia is related to their Native American heritage (1–3). We recently found a frequent non-synonymous variant (*R230C*, rs9282541) within the ATP-binding cassette transporter A1 gene (*ABCA1*) associated with low high-density lipoprotein cholesterol (HDL-C) levels (the most common dyslipidemia in populations with Native American ancestry), obesity and T2D in Mexican Mestizos (4,5). ABCA1 plays a key role in cholesterol efflux and transfer from peripheral cells to lipid-poor apolipoprotein A1 (ApoA1), the first step in HDL particle formation (6,7).

The *R230C* variant was initially described in the Oji-Cree population (8). To date, it has been found only in Native American and Mexican-Mestizo populations (4). We performed a large-scale analysis including individuals from 36 Native North and South American groups and assessed the effect of this variant on anthropometric and metabolic traits. Because it was previously suggested that *R230C* may have conferred selective advantage as a thrifty gene and/or resistance against certain infectious diseases (4), we performed a more thorough analysis seeking evidence of positive selection.

RESULTS

The *C230* allele was present in the majority of the Native American populations at an average frequency of 12% (range 0–31%) (Fig. 1; Supplementary Material, Table S1), but was absent from 863 additional individuals belonging to different European (Spaniard and Dutch) and Asian groups (Han Chinese, Manchu, Mongolian, Siberians and Eskimos) (Fig. 1). The distribution of this allele was not structured according to language or geographic groups (North versus South America), as evidenced by analysis of molecular variance (AMOVA) ($P = 0.978$ and 0.895, respectively). However, the *C230* allele frequency increased at tropical latitudes (between the tropics of Cancer and Capricorn) and gradually decreased at higher latitudes both to the North and South (Fig. 1), showing a significant correlation ($r^2 = 0.328$; $P = 0.02$).

The C230 allele is located on a single genetic haplotype

To perform a phylogenetic reconstruction of the evolutionary relationships, 15 additional single-nucleotide polymorphisms (SNPs) within a 50 kb block were analyzed in 20 Native American and 25 Mexican-Mestizo trios to define haplotype blocks within the region. Together with data from HapMap populations, a total of 58 haplotypes were identified (Supplementary Material, Table S2). Seven haplotypes were found in Native Americans, and the *C230* allele was clearly found in only one genetic block (haplotype 32) in all Mexican-Mestizo, North and South American native individuals analyzed. Maximum parsimony (MP)-based network analysis is shown in Figure 2. The phylogenetic reconstruction uncovered two major lineages (haplogroups A and B) defined by the non-synonymous polymorphism *R219K* within the *ABCA1* gene. The *C230* allele occurred in haplogroup B characterized by the ancestral *R219* allele.

Positive selection testing

The long-range haplotype (LRH) test showed that the extension of linkage disequilibrium (LD) was much longer in *C230* than in non-*C230* chromosomes [relative extended haplotype homozygosity (REHH) = 9.8 and 5.9 at 365 and −434 kb from the core, respectively; Fig. 3]. The former value remained significant compared with REHH values from additional regions elsewhere in the genome, genotyped in a similar set of Native American populations ($P = 0.036$ at 300 kb from the core and $P = 0.021$ at a marker H of 0.04). REHH values were significant in both Kichwa ($P = 0.018$ at a distance of 300 kb and $P = 0.007$ at a marker H of 0.04) and Nahua populations ($P = 0.043$ at a distance of 300 kb and $P = 0.021$ at a marker H of 0.04). The only other two significant core haplotypes were located upstream

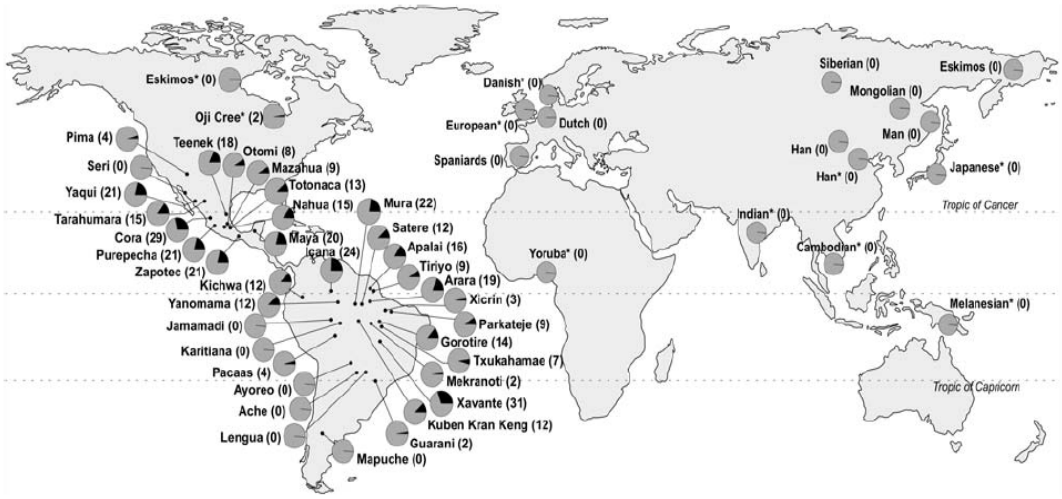


Figure 1. Frequency distribution of the *C230* allele [(%) black-shaded area] in Native American, European, Asian and African populations. *C230* frequency data from populations with an asterisk were obtained from previous reports (8–11 and HapMap and SNP500CANCER databases).

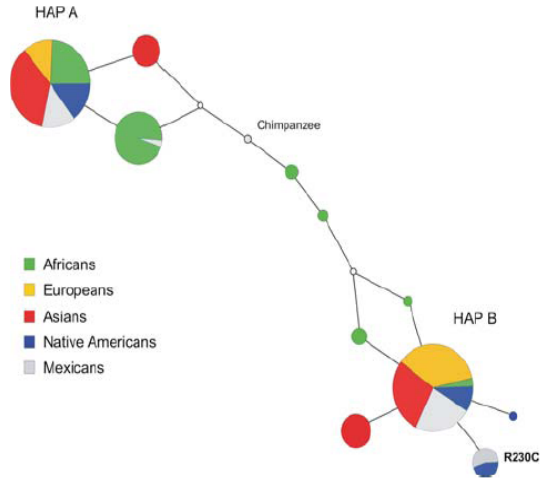


Figure 2. MP-based network describing the evolutionary relationships of 11 distinct haplotypes. Native American and Mexican-Mestizo haplotypes were established in trios; the remainder were inferred from HapMap groups. Each haplotype is represented by a circle whose area reflects the overall number of copies observed and whose color-coding indicates the frequency of the haplotype in the HapMap groups and Native Americans and Mexican Mestizos. Line length is proportional to the number of differences between haplotypes. Non-filled circles represent non-sampled haplotypes reconstructed by the MP algorithm as evolutionary intermediaries between observed haplotypes. The phylogenetic reconstruction uncovered two major lineages (haplogroups A and B) defined by the non-synonymous polymorphism *R219K*. The *R230C* variant occurred on haplogroup B characterized by the ancestral *R219* allele, which is frequent in Europe, Asia and America but infrequent in African populations. The *C230* allele was found in only one genetic block in all Mexican-Mestizo and Native American individuals analyzed.

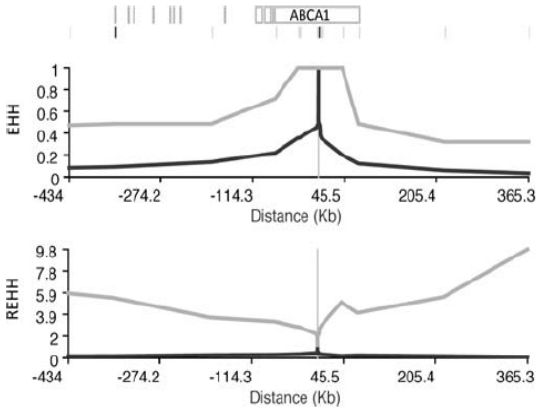


Figure 3. EHH and REHH of *ABCA1/R230C* and 23 additional SNPs × physical distance in Native American individuals. *C230*-bearing chromosomes (in gray) appear to have greatly extended LD compared with non-*C230*-bearing chromosomes (in black). Gray vertical lines in the upper track represent genotyped SNPs, and boxes indicate annotated genes.

the *ABCA1* gene (Supplementary Material, Table S3). Furthermore, in Native Americans from the Human Genome Diversity Panel (HGDP) (*R230C* genotypes not available), the *ABCA1* 5′ region (~75 kb upstream *R230C*) was clearly enriched for outliers of the integrated haplotype score (iHS) statistic genome-wide distribution (iHS > 2.5) (Fig. 4). In agreement with the REHH analysis, the highest iHS values are clustered upstream *ABCA1*.

Association of R230C with HDL-C levels and other metabolic traits

Overall, the prevalence of hypoalphalipoproteinemia (HA) was the most common dyslipidemia (65% in Mexican and South American natives; Supplementary Material, Table S4). Table 1 shows the effect of *R230C* on HDL-C and total cholesterol levels and body mass index (BMI). The *R230C*/*C230C* genotypes were significantly associated with low HDL-C levels in Pimas ($P = 6.4 \times 10^{-5}$) and in the combined analysis of eight Mexican native groups ($P = 5.3 \times 10^{-8}$). In South

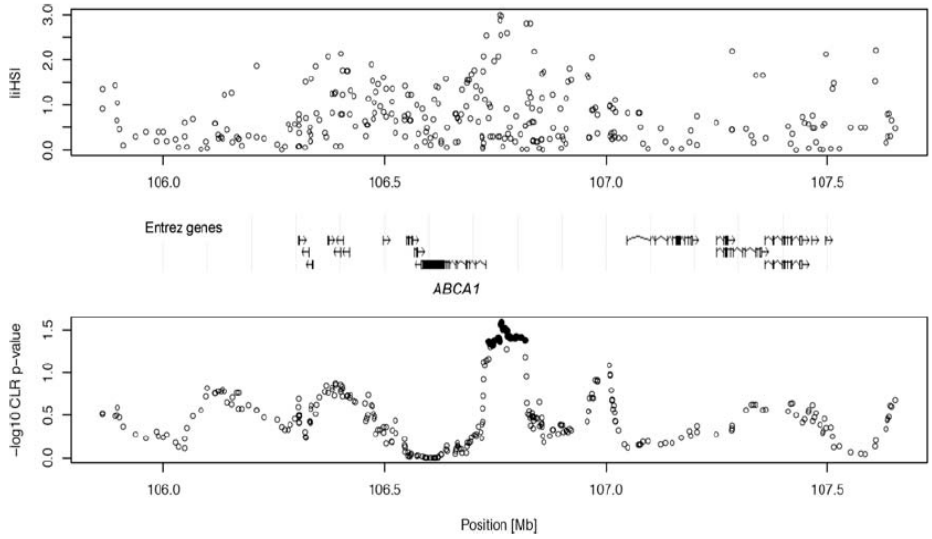


Figure 4. iHS values for individual SNPs flanking the *ABCA1* region (2 Mb) × physical distance (top panel) and *P*-values for the composite likelihood ratio (CLR) test based on a 31-SNP sliding window analysis to detect local regions enriched for high iHS values (bottom panel) in the combined Native American sample. Filled circles indicate *P*-values <0.05 genome-wide significance level.

Table 1. Association of *R230C* with lipid levels and BMI in Native American populations

Native American population (*n*)	HDL-C levels Effect (SE)	*P*-value	Total cholesterol Effect (SE)	*P*-value	BMI Effect (SE)	*P*-value
North America						
USA						
Pima (2563)[a]	−0.075 (0.019)	6.4×10^{-5}	−0.071 (0.014)	1.8×10^{-7}	0.008 (0.015)	0.586
Mexico						
Yaquis (45)	—	—	—	—	0.044 (0.018)	0.012
Teenek (67)	−0.051 (0.026)	0.057	−0.010 (0.024)	0.671	0.001 (0.016)	0.978
Coras (123)	−0.033 (0.014)	0.021	−0.006 (0.013)	0.681	0.032 (0.011)	0.006
Purepechas (15)	−0.040 (0.074)	0.603	−0.048 (0.046)	0.333	0.034 (0.019)	0.097
Mazahuas (83)	−0.039 (0.036)	0.281	−0.031 (0.041)	0.444	0.006 (0.019)	0.758
Nahuas (267)	−0.040 (0.014)	0.014	−0.014 (0.020)	0.470	−0.004 (0.008)	0.617
Totonacas (113)	−0.028 (0.021)	0.180	−0.031 (0.015)	0.042	0.022 (0.013)	0.085
Zapotecs (106)	−0.047 (0.022)	0.038	0.007 (0.019)	0.723	0.007 (0.013)	0.605
Mayans (110)	−0.040 (0.017)	0.023	−0.043 (0.015)	0.004	−0.007 (0.013)	0.554
Mexican natives combined	−0.038 (0.007)	5.3×10^{-8}	−0.019 (0.007)	0.027	0.010 (0.004)	0.012
South America						
Kichwas (79)	−0.043 (0.030)	0.153	−0.005 (0.020)	0.791	0.024 (0.012)	0.050
Parkatejé (78)	−0.029 (0.026)	0.270	−0.002 (0.030)	0.945	0.046 (0.012)	0.0003
All Native Americans combined	−0.042 (0.006)	1.77×10^{-11}	−0.021 (0.006)	7.15×10^{-5}	0.011 (0.003)	0.0001

Effect values are presented as effect size per C230 allele copy, standard error (SE). Linear regression was performed on the basis of log-transformed values for HDL-C levels (mg/dl), total cholesterol levels (mg/dl) and BMI (kg/m^2), adjusting for age, gender and diabetes status. HDL-C and total cholesterol levels were also adjusted for BMI.

[a]*P*-value adjusted by age, gender, birth year, diabetes status and family membership.

American native groups, biochemical data were available from two populations (Parkatejés and Kichwas), and although HDL-C levels were lower in *C230* carriers, the differences did not reach statistical significance. Altogether, the combined results of all Native American groups showed a highly significant effect of the *C230* allele (−4.2% per copy, $P = 1.77 \times 10^{-11}$). Interestingly, differences in the effect of *R230C* on lipid profiles were observed in some Native American populations. *R230C* was strongly associated with low total cholesterol and triglyceride levels in Pimas ($P = 1.8 \times 10^{-7}$ and $P = 7.0 \times 10^{-4}$, respectively) and Mayans ($P = 0.004$ and 0.010, respectively). Although the combined analysis in Native American groups also showed a significant association with low total cholesterol levels ($P = 7.15 \times 10^{-5}$), it was clearly not as significant as the association with low HDL-C levels ($P = 1.77 \times 10^{-11}$). The *C230* allele was associated with higher BMI in the combined analysis of Mexican native groups ($P = 0.012$), in Native South American populations ($P = 0.0003$ and 0.050 for Parkatejé and Kichwas, respectively), but not in Pima Indians. Altogether, combined

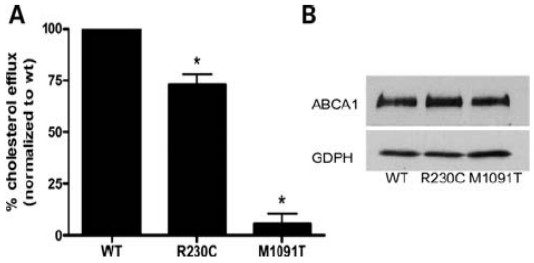


Figure 5. Functional characterization of the ABCA1/*R230C* variant by lipid efflux assay. (**A**) Polyclonal stable cell lines expressing the ABCA1 wild-type (WT), variant *R230C* and mutant *M1091T* (known defective in lipid efflux) were generated, and efflux activity for cholesterol was performed as described in Materials and Methods. Data represent mean ± SD of two to five experiments as a percent of the ApoA1-dependent efflux induced by wild-type ABCA1. Each assay was performed in triplicate. $^{*}P < 0.001$. (**B**) The expression of WT, variant *R230C* and mutant *M1091T* ABCA1 protein in Flp-In cells was assessed by western immunoblotting.

results showed that the *C230* allele was associated with higher BMI ($P = 0.0001$), and this association was evidently more significant in male than in female individuals ($P = 4.05 \times 10^{-6}$ versus $P = 0.02$).

Sequencing and *in vitro* functional analysis

To rule out the presence of another possible causal variant in LD with *C230*, all 50 exons and the promoter region of *ABCA1* were sequenced in a limited number of individuals (2 of each genotype); however, no promoter or coding variant in LD with *R230C* was found. Cholesterol efflux from Flp-In cell lines expressing the *ABCA1 C230* allele was significantly lower (27%) than that of cells expressing the wild-type *R230* allele ($P < 0.001$) (Fig. 5A), confirming that this variant has a functional effect *in vitro*. In contrast, the *C230* and *R230* cell lines showed no differences in phospholipid efflux. The *C230* allele expressed ABCA1 protein at levels comparable with that of *T1091* and wild-type alleles (Fig. 5B).

DISCUSSION

R230C, a private allele to the Americas

The *R230C* allele first identified in Oji-Crees and Mexican Mestizos was found in most Amerindian groups throughout the Americas, but not in any ethnic group from other continents (4,8). This is in agreement with previous studies that have not found this allele in 7717 Caucasian, Asian, African and South-Pacific Rim individuals (HapMap and SNP500CANCER databases) (8–11). This absence strongly suggests that *C230* is a private allele (exclusive to Native American and Native American-derived populations), although it may be present in some non-Amerindian populations not included in this analysis. Its presence on the same haplotype in both North and South Americans suggests that it may have arisen among Native American founders in Beringia or North-East Asia. This is in agreement with recent studies suggesting that founder populations stayed in Beringia long enough to give rise to exclusive genetic variants (12–14). Although private alleles in the HLA system and other genes have been previously reported in some Native American populations (15,16), there is only one previous report of a common private autosomal allele (microsatellite *D9S1120 9RA*) ubiquitous in the Americas (17). It is noteworthy that *D9S1120* and *R230C* (*ABCA1*) are both located on chromosome 9q, although separated by a 19 Mb distance. We genotyped *DS9S1120* in 16 *C230C* Native American and Mestizo homozygotes and found no allele in LD, indicating that the two ancestry informative markers are independent.

The *C230* allele distribution varied among different Amerindian populations (0–0.31). The complex demographic processes that these populations have gone through must have played a crucial role in this distribution. Initially, the moderate bottleneck in the out of Beringia process led to a relatively small effective population size, so genetic drift could have been one of the main causes of fluctuation (18,19). In groups that later expanded demographically to constitute societies formed by thousands of individuals such as Mesoamericans, genetic drift was less likely to cause differences in the distribution of *C230* frequencies.

Evidence suggesting R230C underwent positive selection

Understanding the impact of natural selection acting on particular genes in human populations can provide insights into the genetic etiology of human disease. Interestingly, *ABCA1* has been recognized as one of the genes most likely to have been subject to positive selection in humans since the divergence from the common ancestor of our lineage and that of chimpanzees (20,21). The results of the REHH and iHS analyses for the *ABCA1* gene region in Native Americans are not compatible with a simple neutral evolutionary model, but are consistent with the hypothesis that the *R230C* variant resides on a haplotype which is the target of an ongoing directional selective sweep. It must be acknowledged, however, that with the currently available genotyping data, it is not possible to define whether the *R230C* haplotype is also responsible for the signal resulting from the iHS test.

The geographical distribution of the *C230* allele clearly differs from the North-to-South gradient described for genome-wide neutral markers (22), suggesting the possibility of a climate-related adaptive process, as has been previously described for other genes involved in energy metabolism (23). In the context of Neel's hypothesis (24), *R230C* carriers could have had a selective advantage. Because the C230 ABCA1 protein shows decreased cholesterol efflux, the presence of this variant could favor intracellular cholesterol and energy storage. Specifically, adipose tissue benefits various biological functions including the ability to accommodate fluctuations in energy supply such as severe famine, the regulation of reproductive function and providing energy for the immune system now known to have a significant energy cost (25,26). However, under current westernized lifestyle changes, this allele may have become a major susceptibility allele for low HDL-C levels and other metabolic traits, which is consistent with the association of the *R230C* variant with higher BMI in Native American populations, and with obesity, T2D and metabolic syndrome in Mexican Mestizos (4,5). However, other environmental factors may also be involved in *C230* allele frequency distribution. For

instance, cholesterol plays an important role in various infectious processes such as the entry and replication of Dengue virus type 2 and flaviviral infection (27). The ABCA1 transporter is known to participate in infectious and/or thrombotic disorders involving vesiculation, since homozygous *ABCA1* gene deletions confer complete resistance against cerebral malaria in mice (28,29). Interestingly, areas with higher *C230* allele frequencies correspond to dengue, yellow fever and malaria distributions in the Americas (30). Altogether, these different lines of evidence suggest that the *ABCA1 C230* allele may have been important for survival throughout the colonization of the Americas.

Association of R230C with HDL-C levels and other metabolic traits

Overall, the prevalence of low HDL-C levels was not only higher in Native Americans than in European, Asian and African individuals (3), but also the most common dyslipidemia (65% in Mexican and South American natives). The Pima population is known to have much lower HDL-C and total cholesterol levels than US Caucasians (31). *R230C*/*C230C* genotypes were strongly associated with low HDL-C levels in Native American rural populations, Pimas and urban Mexican Mestizos (4,9). In fact, the sole presence of the *C230* allele explains ~4% of the HDL-C level variation in these populations, which is higher than the variation explained by any other SNP associated with HDL-C levels identified through genome-wide scans in Europeans and Indian Asians (9). This is consistent with both *in silico* (PANTHER subPSEC score −4.27) and *in vitro* evidence confirming that the *R230C* variant is functional (27% decrease in cholesterol efflux) (22). The functional effect is significant, but mild compared with the *T1091* allele previously identified in Tangier patients (32).

Environmental factors and further genetic variation (within *ABCA1* or other genes) may play a relevant role in the association of the *C230* allele with other metabolic traits. Lower total cholesterol and triglyceride levels were found in *C230* carriers only in Pimas and Mayans. In addition, the *C230* allele was associated with higher BMI in Mexican native groups, but not in the Pima Indians. Moreover, a gender effect was observed, as the association of *R230C* with higher BMI was more significant in males. Interestingly, in a previous study, a transcription factor 7-like 2 (*TCF7L2*) gene haplotype (HapA) with evidence of positive selection was also associated with higher BMI only in male individuals (33). Further studies are required to confirm the role of *R230C* in these metabolic and other fat storage-related traits such as non-alcoholic fatty liver disease, which is highly prevalent in Hispanic populations (34,35).

The *C230* allele has also been associated with T2D in the Mexican-Mestizo population (5). The overall frequency of T2D in most Mexican native groups was also high (11.3%); however, the study design was not appropriate for a case–control association in these groups. Interestingly, the *R230C* was only marginally associated with T2D in Pimas ($P = 0.06$) despite the large sample size and the previous finding that HDL-C concentrations in non-diabetic Pima Indian women were negatively associated with the development of T2D (36). Impaired ABCA1 function causes cholesterol accumulation in beta cells in animal models, suggesting that beneficial reductions in plasma lipids may limit the extent of beta cell damage and could partially mask glucose homeostasis disturbances (37). The highly significant association of *R230C* with reduced total cholesterol and triglyceride serum levels observed in Pimas may be one of the factors explaining this marginal association. The role of *R230C* as a risk allele for T2D in Mexican native groups and its interaction with environmental factors requires further analysis.

In conclusion, to the best of our knowledge, this is the first report of a common functional variant exclusive to Native American and descent populations associated with low HDL-C levels and other metabolic traits. We present several lines of evidence in favor of positive selection for the *R230C* allele possibly contributing to the adaptive evolution of Native American populations and providing insight into the genetic etiology of currently prevalent metabolic disease.

MATERIALS AND METHODS

Subjects

The study included a total of 4405 adult individuals from 36 different Native American groups and 863 Europeans and Asians. All Mexican and South American natives and their ancestors (two generations) were born in the same community and spoke their own native language. Field research was conducted by multidisciplinary teams.

Ethics statement

This study was conducted according to the principles expressed in the Declaration of Helsinki and was approved by the Ethics Committees of all participant institutions. Participants provided written informed consent. Local authorities gave their approval to participate in the study, and a translator was used as needed.

Anthropometric and biochemical analyses

Anthropometric and metabolic parameters were available for 2563 Pimas, 1016 Mexican and 157 South American native individuals (Supplementary Material, Table S4). The Pima subjects included are part of an ongoing longitudinal study of the etiology of T2D in the Gila River Indian community in Central Arizona (38). All biochemical measurements in 1050 Mexican natives and Kichwas (from Ecuador) were performed by the INCMNSZ with commercially available standardized methods as described by Villarreal-Molina *et al.* (4). Biochemical parameters of Parkatejé individuals have been previously described (39). T2D and HA were defined according to the American Diabetes Association and National Cholesterol Education Program (NCEP) criteria, respectively (40,41).

DNA sequencing of the *ABCA1* gene

Genomic DNA was extracted from peripheral blood leukocytes. The 50 exons and proximal promoter region of the

ABCA1 gene were amplified in samples from six individuals (two *R230R*, two *R230C* and two *C230C*) as described previously (8). Amplicons were sequenced using ABI PRISM BigDye Terminators version 3.1 on an ABI 3100 automated sequencer according to the manufacturer's protocol (Applied Biosystems, Foster City, CA, USA).

SNP genotyping

The *R230C* variant and 23 SNPs spanning an 800 kb region were genotyped using TaqMan assays (ABI Prism 7900HT Sequence Detection System; Applied Biosystems). The 23 SNPs were selected from the HGDP for being informative in Native American populations and were genotyped in 10 Kichwa, 7 Nahua and 3 Zapotec trios (mother–father–offspring). The names and chromosomal position of all SNPs analyzed are given in Supplementary Material, Table S5. Genotyping call rate exceeded 95% per SNP, and no discordant genotypes were observed in 40 duplicate samples. Deviation from Hardy–Weinberg equilibrium was not observed for any SNP.

Generation of R230C variant constructs and cell lines

Polyclonal stable cell lines expressing the *ABCA1 R230C* variant were generated using the Flp-In system (Invitrogen, Carlsbad, CA, USA) as described previously (42). The generation and detailed biochemical characterization of many of these cell lines are described elsewhere (43). Briefly, the *R230C* variant was generated by PCR-based site-directed mutagenesis using the primers 230F,5′-GAGCGAGTACTT TGTTCCAACATG and 230R,5′-CATGTTGGAACAAAGT ACTCGCTC and cloned into pcDNA5/FRT (Invitrogen). The plasmid was completely sequenced prior to transfection into 293 Flp-In cells. The *M1091T ABCA1* mutation previously identified in Tangier patients was used as control (32).

Cholesterol and phospholipid efflux assay

Efflux experiments were performed as described previously (42). Briefly, cells were loaded with 1 μCi of [^{3}H] cholesterol or 2 μCi of [^{3}H]-choline (Amersham Biosciences, Little Chalfont, Buckingham, UK) for 24 h. The following day, the medium was removed and replaced with serum-free medium containing 5 mg/ml delipidated bovine serum albumin (Sigma, St Louis, MO, USA). After 1 h of incubation, 20 μg/ml human ApoA1 (Athens Research and Technology, Athens, GA, USA) was added for 4 h. For cholesterol efflux, the medium was collected and cells were lysed in 0.1 N NaOH/0.1% SDS. For phospholipid efflux, the [^{3}H]-choline/phospholipids/ApoA1 in the medium was collected by immunoprecipitation with ApoA1 antibody and cells were digested for protein assay. The radioactivity in the samples was quantified by scintillation counting. Cholesterol efflux is expressed as a percent of counts in medium over total (medium + cells). Phospholipid efflux is expressed as counts of the immunocollected [^{3}H]-choline/phospholipid/ApoA1 normalized to cell protein. Data are expressed as percent of the ApoA1-dependent efflux induced by wild-type *ABCA1*. Significance was calculated using a one-way ANOVA test with a Newman–Keuls post-test using GraphPad Prism 4 software (San Diego, CA, USA). ABCA1 expression was determined by western blotting as described previously (44), using anti-ABCA1 or anti-glyceraldehyde 3-phosphate dehydrogenase (GAPDH) antibodies (Chemicon, Temecula, CA, USA).

Statistical analyses

Population genetics. Allele and genotype frequencies, Hardy–Weinberg equilibrium and AMOVA were calculated using Arlequin 3.11 software (45). The linguistic classification of Native American languages was adopted from Campbell (46). Network analyses of haplotypes from HapMap, Mexican-Mestizo and Native American population data were performed using a median-joining and maximum-parsimony method (Network 4.510 software) (47).

Positive selection tests. The LRH test was applied to examine the decay of LD (Sweep software) (48) within an 800 kb region flanking *R230C* using data obtained from the 20 Native American trios described earlier. LD decay was then compared with LRH data generated elsewhere covering a total of 24 Mb of the genome in Native American populations (49). To further explore the presence of positive selection signatures in the *ABCA1* region, the iHS was estimated as described previously (50,51), using publicly available genotype data for ~650 000 SNPs genome-wide distributed in five Native American groups from the HGDP (52). With an approach similar to that described by Nielsen *et al*. (53) to detect regions with aberrant allele frequency spectra (test 1), we applied a composite likelihood test to detect regions with aberrant 'iHS spectra'. We first categorized |iHS| in bins of size 0.1 and then estimated the probability of observing an SNP in each bin, both in the whole-genome data set (background distribution) and in each 31-SNP sliding window over the whole genome. Two composite likelihoods were estimated for each window, multiplying the probability of observing each SNP in the window, by either the probability of observing the SNP in the genome-wide background or that estimated from the window. A log-likelihood ratio was then estimated comparing both likelihoods, where extreme values indicate unusual iHS patterns compared with the rest of the genome.

Associations with HDL-C and metabolic traits. Associations of *R230C* genotypes with HDL-C and other metabolic traits were tested using linear regression models (assuming an additive model) adjusting for covariates including age, sex and BMI (SPSS, version 15.0, statistical package; Chicago, IL, USA). All variables tested were log-transformed for the analysis. Combined association tests were conducted using a Mantel–Haenszel-like model (54). The combined estimated effect was computed as a weighted average of the individual estimated effects using weights proportional to the inverse of the standard errors squared (33).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We thank the volunteers from all native populations of North and South America for their participation; Luz E. Guillén-Pineda for her technical assistance; Gastón Macín for his participation in sample collection in Nahua and Totonaca communities; and the staff of the Diabetes Epidemiology and Clinical Research Section, National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), NIH, for conducting the examinations of the Pima Indian participants.

Conflict of Interest statement: None declared.

FUNDING

This research was supported by grant 69856 from the Consejo Nacional de Ciencia y Tecnología (CONACYT) México, and partly supported by grant 660 from the Fundación Mexicana para la Salud-Silanes; by the Intramural Research Program of the National Institute of Digestive and Kidney Diseases (NIDDK), NIH; and by a grant of the Canadian Institutes of Health Research (CIHR). V.A.-A., T.F.-D., M.V.-C. and L.J.-A. were supported by fellowships from CONACYT, México. J.K.K. was supported by fellowships of the Michael Smith Foundation for Health Research (MSFHR) and the CIHR. M.R.H. holds a Canada Research Chair in Human Genetics.

WEB RESOURCES

The URLs for data presented herein are as follows:

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/

HapMap database, http://www.hapmap.org/

SNP500CANCER, http://snp500cancer.nci.nih.gov

HGDP-CEPH database, http://hgdp.uchicago.edu/cgi-bin/gbrowse/HGDP

Arlequin 3.11 software, http://cmpg.unibe.ch/software/arlequin3/

Network 4.510 software, http://www.fluxus-engineering.com

Sweep software, http://www.broad.mit.edu/mpg/sweep/index.html

REFERENCES

1. Lorenzo, C., Serrano-Rios, M., Martinez-Larrad, M.T., Gabriel, R., Williams, K., Gonzalez-Villalpando, C., Stern, M.P., Hazuda, H.P. and Haffner, S.M. (2001) Was the historic contribution of Spain to the Mexican gene pool partially responsible for the higher prevalence of type 2 diabetes in Mexican-origin populations? The Spanish Insulin Resistance Study Group, the San Antonio Heart Study, and the Mexico City Diabetes Study. *Diabetes Care*, **24**, 2059–2064.
2. Cossrow, N. and Falkner, B. (2004) Race/ethnic issues in obesity and obesity-related comorbidities. *J. Clin. Endocrinol. Metab.*, **89**, 2590–2594.
3. Aguilar-Salinas, C.A., Canizales-Quinteros, S., Rojas-Martínez, R., Mehta, R., Villarreal-Molina, M.T., Arellano-Campos, O., Riba, L., Gómez-Pérez, F.J. and Tusié-Luna, M.T. (2009) Hypoalphalipoproteinemia in the populations with Native American origin: an opportunity to assess the interaction of genes and environment. *Curr. Opin. Lipidol.*, **20**, 92–97.
4. Villarreal-Molina, M.T., Aguilar-Salinas, C.A., Rodriguez-Cruz, M., Riaño, D., Villalobos-Comparan, M., Coral-Vazquez, R., Menjivar, M., Yescas-Gomez, P., Königsoerg-Fainstein, M., Romero-Hidalgo, S. *et al.* (2007) The ATP-binding cassette transporter A1 R230C variant affects HDL cholesterol levels and BMI in the Mexican population: association with obesity and obesity-related comorbidities. *Diabetes*, **56**, 1881–1887.
5. Villarreal-Molina, M.T., Flores-Dorantes, M.T., Arellano-Campos, O., Villalobos-Comparan, M., Rodríguez-Cruz, M., Miliar-García, A., Huertas-Vazquez, A., Menjivar, M., Romero-Hidalgo, S., Wacher, N.H. *et al.* (2008) Association of the ATP-binding cassette transporter A1 R230C variant with early-onset type 2 diabetes in a Mexican population. *Diabetes*, **57**, 509–513.
6. Attie, A.D., Kastelein, J.P. and Hayden, M.R. (2001) Pivotal role of ABCA1 in reverse cholesterol transport influencing HDL levels and susceptibility to atherosclerosis. *J. Lipid Res.*, **42**, 1717–1726.
7. Oram, J.F. and Vaughan, A.M. (2006) ATP-binding cassette cholesterol transporters and cardiovascular disease. *Circ. Res.*, **99**, 1031–1043.
8. Wang, J., Burnett, J.R., Near, S., Young, K., Zinman, B., Hanley, A.J., Connelly, P.W., Harris, S.B. and Hegele, R.A. (2000) Common and rare ABCA1 variants affecting plasma HDL cholesterol. *Arterioscler. Thromb. Vasc. Biol.*, **20**, 1983–1989.
9. Kooner, J.S., Chambers, J.C., Aguilar-Salinas, C.A., Hinds, D.A., Hyde, C.L., Warnes, G.R., Gómez-Pérez, F.J., Frazer, K.A., Elliott, P., Scott, J. *et al.* (2008) Genome-wide scan identifies variation in MLXIPL associated with plasma triglycerides. *Nat. Genet.*, **4**, 149–151.
10. Hu, S., Zhong, Y., Hao, Y., Luo, M., Zhou, Y., Guo, H., Liao, W., Wan, D., Wei, H., Gao, Y. *et al.* (2009) Novel rare alleles of ABCA1 are exclusively associated with extreme high-density lipoprotein-cholesterol levels among the Han Chinese. *Clin. Chem. Lab. Med.*, **47**, 1239–1245.
11. Frikke-Schmidt, R., Nordestgaard, B.G., Jensen, G.B. and Tybjaerg-Hansen, A. (2004) Genetic variation in ABC transporter A1 contributes to HDL cholesterol in the general population. *J. Clin. Invest.*, **114**, 1343–1353.
12. González-José, R., Bortolini, M.C., Santos, F.R. and Bonatto, S.L. (2008) The peopling of America: craniofacial shape variation on a continental scale and its interpretation from an interdisciplinary view. *Am. J. Phys. Anthropol.*, **137**, 175–187.
13. Kitchen, A., Miyamoto, M.M. and Mulligan, C.J. (2008) A three-stage colonization model for the peopling of the Americas. *PLoS ONE*, **3**, e1596.
14. Fagundes, N.J.R., Kanitz, R. and Bonatto, S.L. (2008) A reevaluation of the Native American mtDNA genome diversity and its bearing on the models of early colonization of Beringia. *PLoS ONE*, **3**, e3157.
15. Williams, R.C. and McAuley, J.E. (1992) HLA class I variation controlled for genetic admixture in the Gila River Indian community of Arizona: a model for the Paleo-Indians. *Hum. Immunol.*, **33**, 39–46.
16. Salzano, F.M. (2002) Molecular variability in Amerindians: widespread but uneven information. *An. Acad. Bras. Cienc.*, **74**, 223–263.
17. Schroeder, K.B., Jakobsson, M., Crawford, M.H., Schurr, T.G., Boca, S.M., Conrad, D.F., Tito, R.Y., Osipova, L.P., Tarskaia, L.A., Zhadanov, S.I. *et al.* (2009) Haplotypic background of a private allele at high frequency in the Americas. *Mol. Biol. Evol.*, **26**, 995–1016.
18. Fagundes, N.J.R., Ray, N., Beaumont, M., Neuenschwander, S., Salzano, F.M., Bonatto, S.L. and Excoffier, L. (2007) Statistical evaluation of alternative models of human evolution. *Proc. Natl Acad. Sci. USA*, **104**, 17614–17619.
19. Fagundes, N.J.R., Kanitz, R., Eckert, R., Valls, A.C., Bogo, M.R., Salzano, F.M., Smith, D.G., Silva, W.A. Jr, Zago, M.A., Ribeiro-dos-Santos, A.K. *et al.* (2008) Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. *Am. J. Hum. Genet.*, **8**, 583–592.
20. Brunham, L.R., Singaraja, R.R., Pape, T.D., Kejariwal, A., Thomas, P.D. and Hayden, M.R. (2005) Accurate prediction of the functional significance of single nucleotide polymorphisms and mutations in the ABCA1 gene. *PLoS Genet.*, **1**, e83.
21. Clark, A.G., Glanowski, S., Nielsen, R., Thomas, P.D., Kejariwal, A., Todd, M.A., Tanenbaum, D.M., Civello, D., Lu, F., Murphy, B. *et al.* (2003) Inferring nonneutral evolution from human–chimp–mouse orthologous gene trios. *Science*, **302**, 1960–1963.
22. Wang, S., Lewis, C.M., Jakobsson, M., Ramachandran, S., Ray, N., Bedoya, G., Rojas, W., Parra, M.V., Molina, J.A., Gallo, C. *et al.* (2007) Genetic variation and population structure in native Americans. *PLoS Genet.*, **3**, e185.
23. Hancock, A.M., Witonsky, D.B., Gordon, A.S., Eshel, G., Pritchard, J.K., Coop, G. and Di Rienzo, A. (2008) Adaptations to climate in candidate genes for common metabolic disorders. *PLoS Genet.*, **4**, e32.

24. Neel, J.V. (1962) Diabetes mellitus: a 'thrifty' genotype rendered detrimental by 'progress'? *Am. J. Hum. Genet.*, **14**, 353–362.
25. Prentice, A.M., Rayco-Solon, P. and Moore, S.E. (2005) Insights from the developing world: thrifty genotypes and thrifty phenotypes. *Proc. Nutr. Soc.*, **64**, 153–161.
26. Wells, J.C. (2009) Ethnic variability in adiposity and cardiovascular risk: the variable disease selection hypothesis. *Int. J. Epidemiol.*, **38**, 63–71.
27. Lee, C.J., Lin, H.R., Liao, C.L. and Lin, Y.L. (2008) Cholesterol effectively blocks entry of flavivirus. *J. Virol.*, **82**, 6470–6480.
28. Simons, K. and Toomre, D. (2000) Lipid rafts and signal transduction. *Nat. Rev. Mol. Cell. Biol.*, **1**, 31–39.
29. Combes, V., Coltel, N., Alibert, M., van Eck, M., Raymond, C., Juhan-Vague, I., Grau, G.E. and Chimini, G. (2005) ABCA1 gene deletion protects against cerebral malaria: potential pathogenic role of microparticles in neuropathology. *Am. J. Pathol.*, **166**, 295–302.
30. Rogers, D.J. and Randolph, S.E. (2000) The global spread of malaria in a future, warmer world. *Science.*, **289**, 1763–1766.
31. Howard, B.V., Davis, M.P., Pettitt, D.J., Knowler, W.C. and Bennett, P.H. (1983) Plasma and lipoprotein cholesterol and triglyceride concentrations in the Pima Indians: distributions differing from those of Caucasians. *Circulation.*, **68**, 714–724.
32. Clee, S.M., Kastelein, J.J., van Dam, M., Marcil, M., Roomp, K., Zwarts, K.Y., Collins, J.A., Roelants, R., Tamasawa, N., Stulc, T. *et al.* (2000) Age and residual cholesterol efflux affect HDL cholesterol levels and coronary artery disease in ABCA1 heterozygotes. *J. Clin. Invest.*, **106**, 1263–1270.
33. Helgason, A., Pálsson, S., Thorleifsson, G., Grant, S.F., Emilsson, V., Gunnarsdottir, S., Adeyemo, A., Chen, Y., Chen, G., Reynisdottir, I. *et al.* (2007) Refining the impact of TCF7L2 gene variants on type 2 diabetes and adaptive evolution. *Nat. Genet.*, **39**, 218–225.
34. Browning, J.D., Szczepaniak, L.S., Dobbins, R., Nuremberg, P., Horton, J.D., Cohen, J.C., Grundy, S.M. and Hobbs, H.H. (2004) Prevalence of hepatic steatosis in an urban population in the United States: impact of ethnicity. *Hepatology*, **40**, 1387–1395.
35. Guerrero, R., Vega, G.L., Grundy, S.M. and Browning, J.D. (2009) Ethic differences in hepatic steatosis: an insulin resistance paradox? *Hepatology*, **49**, 791–801.
36. Fagot-Campagna, A., Narayan, K.M., Hanson, R.L., Imperatore, G., Howard, B.V., Nelson, R.G., Pettitt, D.J. and Knowler, W.C. (1997) Plasma lipoproteins and incidence of non-insulin-dependent diabetes mellitus in Pima Indians: protective effect of HDL cholesterol in women. *Atherosclerosis*, **128**, 113–119.
37. Brunham, L.R., Kruit, J.K., Verchere, C.B. and Hayden, M.R. (2008) Cholesterol in islet dysfunction and type 2 diabetes. *J. Clin. Invest.*, **118**, 403–408.
38. Knowler, W.C., Bennett, P.H., Hamman, R.F. and Miller, M. (1978) Diabetes incidence and prevalence in Pima Indians: a 19-fold greater incidence than in Rochester, Minnesota. *Am. J. Epidemiol.*, **108**, 497–505.
39. Vieira-Filho, J.P., Reis, A.F., Kasamatsu, T.S., Tavares, E.F., Franco, L.J., Matioli, S.R. and Moisés, R.S. (2004) Influence of the polymorphisms Tpr64Arg in the beta 3-adrenergic receptor gene and Pro12Ala in the PPAR gamma 2 gene on metabolic syndrome-related phenotypes in an indigenous population of the Brazilian Amazon. *Diabetes Care*, **27**, 621–622.
40. The Expert Committee on the Diagnosis Classification of Diabetes Mellitus (2003) Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care*, **26**, S5–S20.
41. Expert Panel on Detection, Evaluation and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) (2001) Executive summary of the Third Report of the National Cholesterol Education Program (NCEP). *JAMA*, **285**, 2486–2497.
42. See, R.H., Caday-Malcolm, R.A., Singaraja, R.R., Zhou, S., Silverston, A., Huber, M.T., Moran, J., James, E.R., Janoo, R., Savill, J.M. *et al.* (2002) Protein kinase A site-specific phosphorylation regulates ATP-binding cassette A1 (ABCA1)-mediated phospholipid efflux. *J. Biol. Chem.*, **277**, 41835–41842.
43. Singaraja, R.R., Van Eck, M., Bissada, N., Zimetti, F., Collins, H.L., Brunham, L.R., Kang, M.H., Zannis, V.I., Chimini, G. and Hayden, M.R. (2006) Specific mutations in ABCA1 have discrete effects on ABCA1 function and lipid phenotypes both *in vivo* and *in vitro*. *Circ. Res.*, **99**, 389–397.
44. Wellington, C.L. (2002) ABCA1 mRNA and protein distribution patterns predict multiple different roles and levels of regulation. *Lab. Invest.*, **82**, 273–283.
45. Excoffier, L., Laval, G. and Schneider, S. (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinform. Online*, **1**, 47–50.
46. Campbell, L. (1997) *American Indian languages: the historical linguistics of Native America*. Oxford University Press, New York.
47. Bandelt, H.J., Forster, P. and Rohl, A. (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.*, **16**, 37–48.
48. Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z., Richter, D.J., Schaffner, S.F., Gabriel, S.B., Platko, J.V., Patterson, N.J., McDonald, G.J. *et al.* (2002) Detecting recent positive selection in the human genome from haplotype structure. *Nature*, **419**, 832–837.
49. Moreno-Estrada, A., Tang, K., Sikora, M., Marquès-Bonet, T., Casals, F., Navarro, A., Calafell, F., Bertranpetit, J., Stoneking, M. and Bosch, E. (2009) Interrogating 11 fast-evolving genes for signatures of recent positive selection in worldwide human populations. *Mol. Biol. Evol.*, **26**, 2285–2297.
50. Voight, B.F., Kudaravalli, S., Wen, X. and Pritchard, J.K. (2006) A map of recent positive selection in the human genome. *PLoS Biol.*, **4**, e72.
51. Pickrell, J.K., Coop, G., Novembre, J., Kudaravalli, S., Li, J.Z., Absher, D., Srinivasan, B.S., Barsh, G.S., Myers, R.M., Feldman, M.W. *et al.* (2009) Signals of recent positive selection in a worldwide sample of human populations. *Genome Res.*, **19**, 826–837.
52. Li, J.Z., Absher, D.M., Tang, H., Southwick, A.M., Casto, A.M., Ramachandran, S., Cann, H.M., Barsh, G.S., Feldman, M., Cavalli-Sforza, L.L. *et al.* (2008) Worldwide human relationships inferred from genome-wide patterns of variation. *Science*, **22**, 1100–1104.
53. Nielsen, R., Williamson, S., Kim, Y., Hubisz, M.J., Clark, A.G. and Bustamante, C. (2005) Genomic scans for selective sweeps using SNP data. *Genome Res.*, **15**, 1566–1575.
54. Mantel, N. and Haenszel, W. (1959) Statistical aspects of the analysis of data from retrospective studies of disease. *J. Natl Cancer Inst.*, **22**, 719–748.

APÊNDICE II

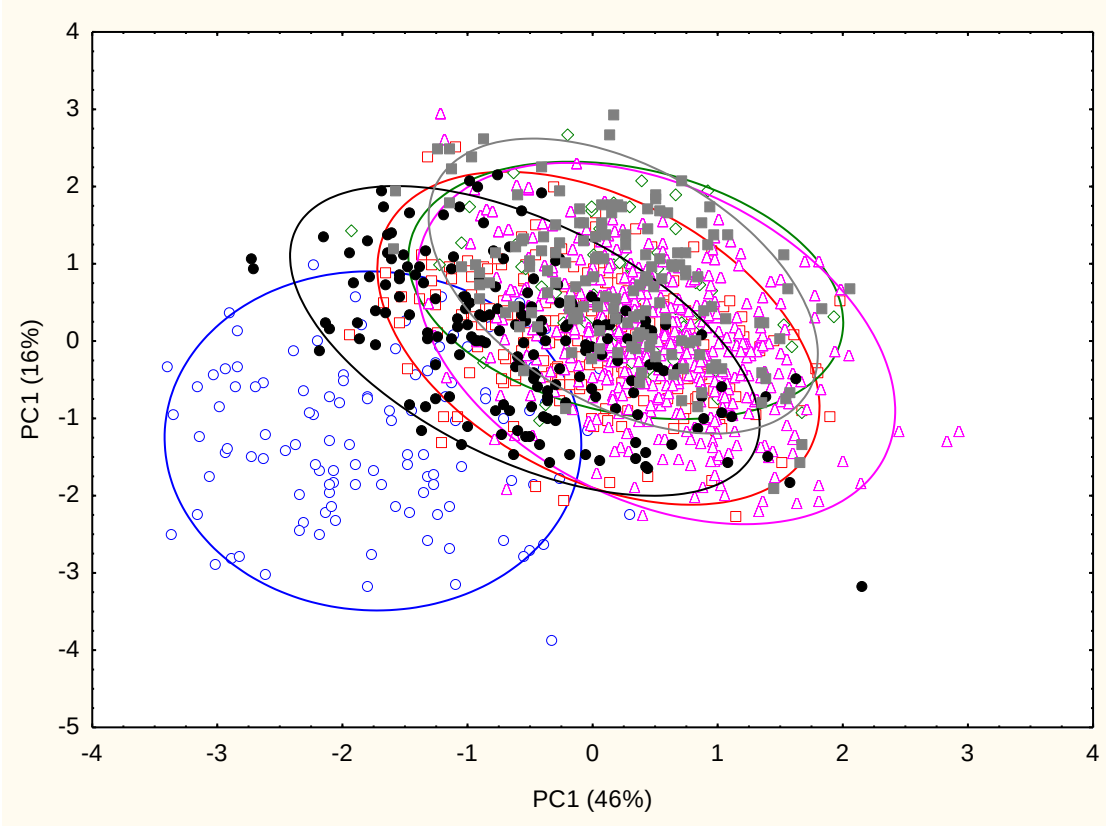


Figure 1. Análise de Componentes Principais das variáveis cefálicas considerando 6 tribos indígenas sul-americanas: Xavante (azul), Kaingang (preto), Kayapo (vermelho), Yanomama (cinza), Baniwa (verde) e Ticuna (rosa). O gráfico mostra grande separação dos Xavantes em relação às demais populações, incluindo outras populações Gê.

Tabela 1. Modelo Relethford-Blangero para as 6 populações exibidas no gráfico acima. Os Xavantes apresentam maior variação fenotípica interna; considerando todas as populações estudadas apresenta uma variação maior que o esperado para o modelo. *Mean Within-Group Phenotypic Variance* = 0,748.

Within-group Phenotypic Variance

Population	*r(ii)*	*Observed*	*Expected*	*Residual*
XAV	0,617994	0,888	0,350	0,538
KAY	0,032154	0,711	0,886	-0,176
BAN	0,176957	0,658	0,754	-0,096
TIC	0,119905	0,644	0,806	-0,162
KAI	0,033031	0,877	0,885	-0,009
YAN	0,117953	0,712	0,808	-0,095

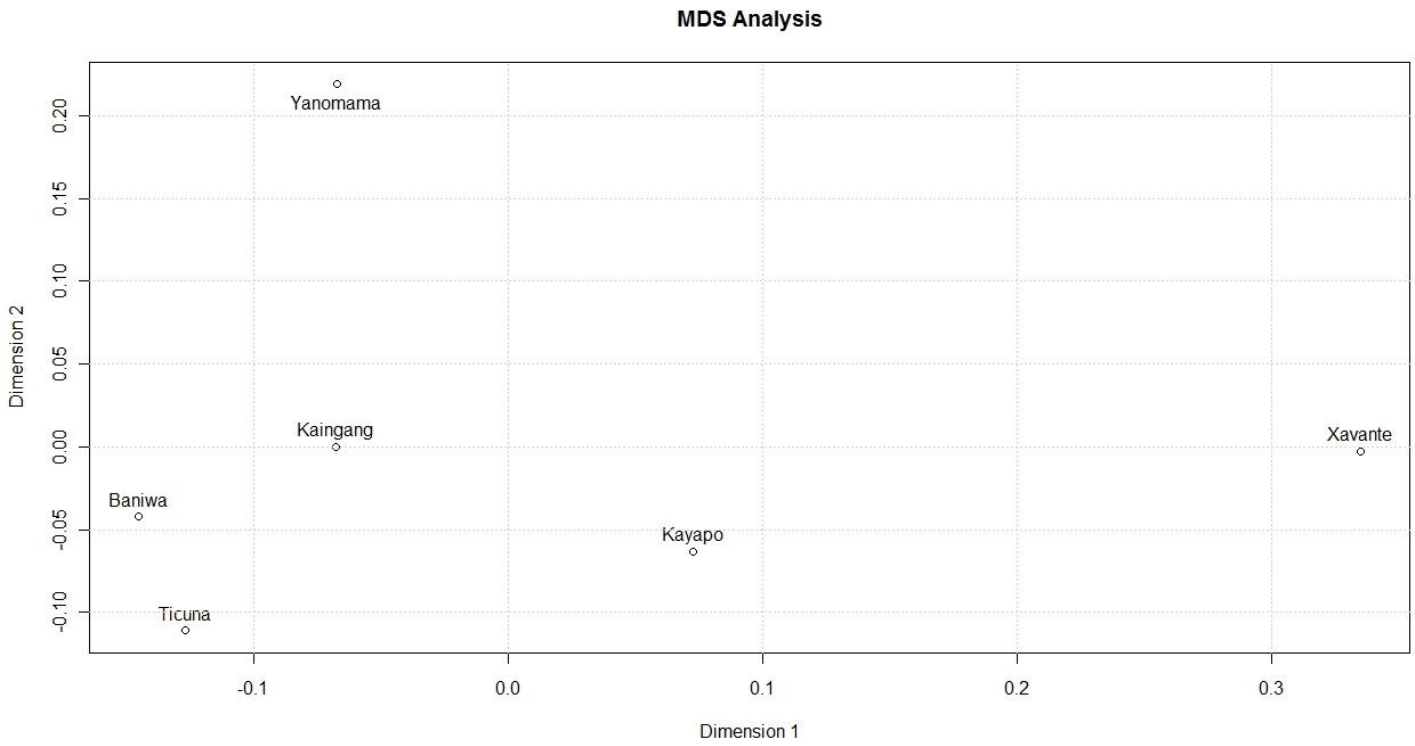


Figura 2. *Multidimensional Scaling Analysis* com dados de seqüência da região HVSI do DNA mitocondrial considerando as 6 populações fenotipificadas. Os Xavantes apresentam grande diferenciação em relação às outras populações em nível molecular.

UFRGS

PPGBM

Programa de Pós-Graduação
m Genética e Biologia Molecular
da UFRGS