

Genetic signature of natural selection in first Americans

Carlos Eduardo G. Amorim^{a,1}, Kelly Nunes^{b,1}, Diogo Meyer^b, David Comas^c, Maria Cátira Bortolini^d, Francisco Mauro Salzano^{d,2}, and Tábita Hünemeier^{b,2}

^aDepartment of Biological Sciences, Columbia University, New York, NY 10027; ^bDepartamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, 03178-200 São Paulo, Brazil; ^cInstitut de Biologia Evolutiva, Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, 08003 Barcelona, Spain; and ^dDepartamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, 91501-970 Porto Alegre, Brazil

Contributed by Francisco Mauro Salzano, January 2, 2017 (sent for review July 5, 2016; reviewed by Theodore G. Schurr and Anne C. Stone)

When humans moved from Asia toward the Americas over 18,000 y ago and eventually peopled the New World they encountered a new environment with extreme climate conditions and distinct dietary resources. These environmental and dietary pressures may have led to instances of genetic adaptation with the potential to influence the phenotypic variation in extant Native American populations. An example of such an event is the evolution of the fatty acid desaturases (*FADS*) genes, which have been claimed to harbor signals of positive selection in Inuit populations due to adaptation to the cold Greenland Arctic climate and to a protein-rich diet. Because there was evidence of intercontinental variation in this genetic region, with indications of positive selection for its variants, we decided to compare the Inuit findings with other Native American data. Here, we use several lines of evidence to show that the signal of *FADS*-positive selection is not restricted to the Arctic but instead is broadly observed throughout the Americas. The shared signature of selection among populations living in such a diverse range of environments is likely due to a single and strong instance of local adaptation that took place in the common ancestral population before their entrance into the New World. These first Americans peopled the whole continent and spread this adaptive variant across a diverse set of environments.

peopling of America | natural selection | genetics | first Americans

Human history in America begins about 18,000–15,000 y before the present (yBP) (1, 2), when the first Americans entered the New World and migrated southward, peopling all latitudes of the continent. However, the history of this human group begins well before this date, before the initial colonization. Around 25,000 yBP, the Last Glacial Maximum (LGM) drastically reduced sea levels. In the region where we currently find the Bering Strait, this drop in sea level exposed an unglaciated land extension known as the Bering land bridge, which joined Northeast Asia and North America, forming Beringia (3–5). Most of the studies on the peopling of the Americas agree that the genetic differentiation of the current Native American populations probably occurred in this area during a standstill in Beringia (6–9), where migrants from different regions in Asia had arrived about 23,000 yBP, inhabiting it for a few thousand years (5,000–8,000 y) (1).

Greenlandic Inuit are one of the descendant populations of the first Americans. A recent study analyzing 196,725 SNPs in 191 individuals from this population proposed that they have experienced genetic and physiological adaptations to a cold climate, obtained through their living for a long time in the extreme conditions of the Arctic and subsisting on a diet rich in protein and fatty acids (10). Using an approach based on a measure of genetic differentiation (F_{ST}), these authors provided evidence for positive selection at several sites in the *FADS* genes on chromosome 11 (*FADS1*, *FADS2*, and *FADS3*), which are involved in the metabolism of omega-3 polyunsaturated fatty acids (PUFAs).

A previous investigation (11) also studied *FADS1* and *FADS2*, with different SNPs from those of ref. 10, and found an

association between one haplotype (D) with PUFAs metabolism. This haplotype presented considerable intercontinental variation in its frequency, possibly indicating some level of local adaptation. Moreover, the observation that the climate and diet resources in Beringia during the LGM were probably similar to those in Greenland (4) suggested the possibility that this adaptive haplotype originated in a population ancestral to both Inuit and Native Americans. For this reason, we analyzed sequence variation in *FADS* genes in Native American populations from Central and South America and searched for sites in *FADS* genes implicated in cold adaptation.

Results and Discussion

Genomic Screening for Mutations with Large Continental Frequency Differential. We first compared Native Americans (12) to other continental populations (Africans, Europeans, and East Asians; Table S1) to verify the alleles that had high frequencies in indigenous American populations and low frequencies elsewhere. We found three significant SNPs (Fisher's exact test with Bonferroni correction), two of which lay within *FADS2* (rs174570, $P < 0.0001$ and rs174556, $P < 0.001$) and another located on chromosome 22 (rs13054099, $P < 0.0000001$) in *SLC25A17* (Fig. 1). The latter gene encodes a peroxisomal membrane protein that belongs to the family of mitochondrial solute carriers. It is expressed in liver and is likely involved in membrane transport (13). Thus far, there is no a priori evidence that it plays a role in adaptation to diet or temperature.

Significance

There is much interest in understanding the role of natural selection in shaping physiological adaptations to climate, diet, and diseases in humans. We investigated this issue by analyzing genomic data from Native American populations inhabiting different ecological regions and ancient Native Americans. We found signals of natural selection at the fatty acid desaturases (*FADS*) genes not only in an Arctic population, as was previously found, but throughout the Americas, suggesting a single and strong adaptive event that occurred in Beringia, before the range expansion of the first Americans within the American continent and Greenland.

Author contributions: C.E.G.A., K.N., and T.H. designed research; C.E.G.A., K.N., and T.H. performed research; C.E.G.A., K.N., and T.H. analyzed data; and C.E.G.A., K.N., D.M., D.C., M.C.B., F.M.S., and T.H. wrote the paper.

Reviewers: T.G.S., University of Pennsylvania; and A.C.S., Arizona State University.

The authors declare no conflict of interest.

¹C.E.G.A. and K.N. contributed equally to this work.

²To whom correspondence may be addressed. Email: francisco.salzano@ufrgs.br or hunemeier@usp.br.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1620541114/-DCSupplemental.

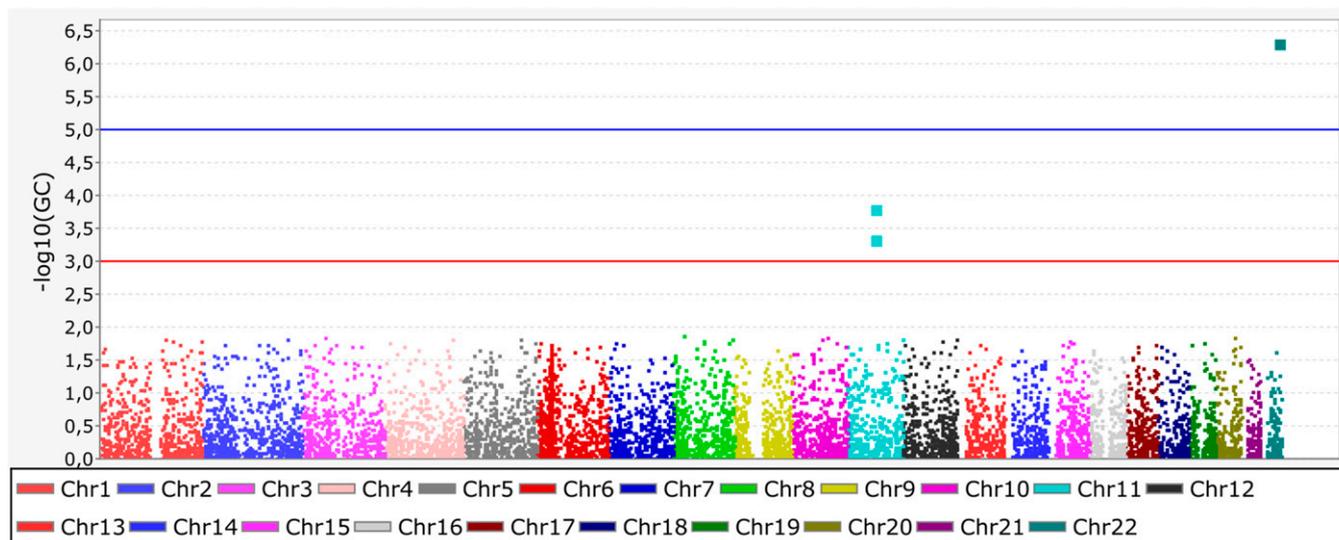


Fig. 1. Significant SNPs found in Native American populations compared with other continental populations (Africans, Asians, and Europeans). The comparison was tested using Fisher's exact test with Bonferroni correction. We found three significant SNPs, two within *FADS2* (rs174570, $P < 0.0001$ and rs174556, $P < 0.001$) and another located on chromosome 22 (rs13054099, $P < 0.0000001$) in *SLC25A17*.

Genetic Signatures of Natural Selection. Genes showing population-specific changes in allele frequency, such as *FADS2* and *SLC25A17* in Native Americans, are potential candidates for revealing adaptive evolution. We thus focused our subsequent analyses on chromosomes 11 and 22. To evaluate whether we could identify natural selection as the cause for these extreme allele frequency changes, we applied three cross-population test statistics to two different datasets with 374,470 (dataset 1, ref. 12) and 593,142 (dataset 2, ref. 14) SNPs in each. Two tests are F_{ST} -based, involving the comparison of differentiation between Native Americans, Europeans, and East Asians [the population branch statistics (PBS) (15) and a Bayesian method implemented with Bayescan (16)]. To formally test whether natural selection underlies the cases of extreme differentiation, PBS values were compared against those obtained with neutral coalescent simulations generated according to a plausible demographic scenario for the peopling of the New World (17). Bayescan corrects for demographic biases by differentiating the proportion of the F_{ST} variation that is locus-specific (due to natural selection) from that which is population-specific (due to demography) (16). In addition to these tests, we also performed the cross-population extended haplotype homozygosity test (XP-EHH) (18) between pairs of populations.

Fig. S1 *A* and *B* show the distribution of PBS statistics for the Native American branch across the extension of chromosome 11 in each dataset. The most extreme PBS value in dataset 1 corresponds to rs174570 (Fig. S1*A*), the same SNP that was highly significant in our preliminary analysis of extreme allele frequency differences across continents (Fig. 1). This result demonstrated that the SNP detection was consistent across typing platforms and different methods. In addition, this SNP was one of the sites found under positive selection in the Greenlandic Inuit (10).

For dataset 2, we found 15 SNPs with highly significant PBS values (Fig. S1*B*), three of them being located on *FADS* genes (rs74771917, rs7115739, and rs174570). Neutral coalescent simulations indicated that these deviations were statistically significant ($P < 0.0001$; Fig. S2), consistent with the action of positive selection as opposed to genetic drift in increasing the frequency of the derived allele in Native American populations. The other SNPs showing signals of natural selection in Native Americans were located at, or near, genes related with metabolism (Table S2). The Bayescan analysis corroborated all of the aforementioned

findings, especially regarding selection on *FADS* genes (Fig. S1 *C* and *D*). However, for the rs13054099 SNP on chromosome 22 we found weak or lack of evidence for selection with all methods.

Putatively Selected Haplotype in Extant and Ancient Americans. To investigate whether this signal of natural selection could be the result of adaptation to the conditions encountered by the ancestors of first Americans in Beringia, we compared the genotypes of living Native Americans and Inuit to those of four ancient humans. These included Saqqaq, a Paleo-Eskimo from Greenland who lived ~4,000 yBP (19); Anzick-1, an individual belonging to the classical North American Clovis culture who lived ~12,500 yBP (20); the Mal'ta boy who lived in Siberia ~24,000 yBP (21); and the Ust'-Ishim man, who lived in Siberia ~45,000 yBP (22) (Table 1). The haplotype found in the homozygous state in the Anzick-1 individual—which is a representative of first Americans, or close to them—is the same as the one present at high frequencies in extant Native Americans from diverse locations of the continent. The Paleo-Eskimo had a haplotype different from that present in Native Americans or the Inuit. The high frequency of the putatively selected *FADS* haplotype and its shared distribution in living and ancient Native Americans (Table S3) is therefore consistent with a scenario of intense selection during the Beringian standstill.

During the LGM, Beringia offered a feasible home for the Asian migrants, due to its ecological conditions (4) furnishing a relatively isolated place for their differentiation. We hypothesized that a putative selection event occurred after migrants from northeastern Siberia (23), as well as other Asian regions, reached Beringia. Presently, extant native Siberians present relatively lower frequencies (average 69%) of the putatively selected *FADS* haplotype (Fig. 2 and Table S3). Cardona et al. (24) surveyed the entirety of Siberian chromosome 11 and found no indication of selection at the *FADS* region. Therefore, the presence of the putatively selected *FADS* haplotype might have resulted from subsequent gene flow between Native Americans and Siberian populations occurring after the Beringian event. Note that the two ancient Siberians (Mal'ta and Ust'-Ishim) had haplotypes distinct from that found in the ancient Anzick-1 Native American individual but similar to the one found in the Saqqaq Paleo-Eskimo (Table 1).

ancestry with the first Americans, which reflects adaptive selection in that region of the genome.

Conclusions

Our results strongly suggest that our hypothesis that the signal of positive selection is not restricted to the Arctic is correct. We found a signature of natural selection at *FADS* loci in all 53 Native American populations studied here, including a strong signal in Amazonian populations that are extremely different from the Inuit regarding culture, environment, and diet. In our opinion, this genetic profile suggests a single and very strong adaptive event that occurred in Beringia, before the range expansion of the ancestors of the first Americans within the American continent and Greenland. This event was likely related to the metabolic adaptations to diet and cold weather required to subsist during glacial and periglacial conditions that existed during the Beringian standstill (30).

Materials and Methods

Populations. We analyzed data from 349 individuals from 44 Native American populations previously published by Reich et al. (12), as well as information on 48 individuals from nine Native American populations published by Skoglund et al. (14). These datasets were called dataset 1 and dataset 2, respectively. Detailed population information can be found in Table S1. In addition, data from 225 Asian individuals and 328 Europeans from the Human Genome Diversity Project (www.hagsc.org/hgdp/) were evaluated. The data were curated in the same way reported by Laurie et al. (31). We computationally phased these data using the SHAPEIT software (32) with default parameters.

PBS Test. PBS analysis was performed as described in ref. 15. For each SNP, an F_{ST} value was estimated between pairs of populations [Native Americans (NAM), East Asians (EAS), and Utah residents of European ancestry (CEU)], using Reynolds et al.'s (33) estimator. Nonpolymorphic sites in at least two populations were excluded. PBS estimates were performed between NAM and EAS, using CEU as an outgroup. The analyses were done with individual SNPs and with 20 SNPs windows overlapping 5 SNPs, as implemented by Fumagalli et al. (10).

Neutral Coalescent Simulations. Simulations were performed using the demographic model for the peopling of the New World inferred by Gutenkunst et al. (17). In short, these authors modeled the Eurasian divergence from the African population, with subsequent divergence of the Asian population from the European, at ~26 kya. The Americas were peopled by a subset of the Asian population with 800 individuals that went through exponential growth after entering the New World ~22 kya. Gene flow from Europeans to both Asian and Native American populations was allowed to happen, because the genetic dataset used to infer this model comprised Mexican individuals that are known to have some European genetic contribution (17). This approach allowed us to correct for possible bias due to genetic admixture, and indeed the generation time and population size estimates are in agreement with the literature (1). More details about this scenario can be found in figure 3 and table 2 from ref. 33.

We implemented our simulations according to this scenario in ms (34) with the following command line:

```
ms Ntot 10000 -s 1 -l 4 0 NEuro NAsia NAm -n 1 1.682020 -n 2 2.424020
-n 3 4.185850 -n 4 7.942130 -eg 0 2 67.978337 -eg 0 3 109.406463 -eg 0 4
147.474095 -ema 0 4 x 0 0 0 0 x 3.960400 0 0 3.960400 x 0 0 0 0 x -ej
0.029475 4 3 -ema 0.029475 4 x 0.881098 0.561966 x 0.881098 x 3.960400 x
0.561966 3.960400 x x -ej 0.036114 3 2 -en 0.036114 2 0.287184 -ema 0.036114
4 x 7.293140 x x 7.293140 x x -ej 0.197963 2 1 -en 0.303500 1 1.
```

The total sample number (Ntot) and ethnic group samples number (NEuro, NAsia, and NAm) varied between datasets, being 904, 328, 358, and 218 for dataset 1 and 365, 129, 188, and 48 for dataset 2, respectively. We used 10,000 simulations according to this scenario to generate a null distribution of the PBS values to which empirical PBS values were compared. This procedure allowed us to evaluate the significance of the differences and get the outlier PBS values.

Detecting SNPs Under Positive Selection with a Bayesian Approach. Two versions of BayeScan (16, 35) were used to identify candidate targets for positive selection. Both versions assume an island model in which the subpopulations' allele frequencies are correlated through a common migrant gene pool from which they may differ in varying degrees. Based on this outcome, F_{ST} is calculated and decomposed into a population-specific component (β), shared by all loci, and a locus-specific component (α), shared by all populations (see refs. 15 and 35 and the corresponding manuals for details). Significant, positive values of α indicate an overly large level of differentiation of a given SNP, given the genomic average specific for each population (the β part of the F_{ST} mentioned above), indicating positive selection. With the older version of BayeScan we were able to identify outlier loci, and with the new version we identified those that were outliers in the Native American population, assigning posterior probabilities for each marker as being under positive selection. These posterior probabilities were then transformed into q-values to control for the false discovery rate.

A sliding-window approach was followed to generate Manhattan plots with the distribution of the q-values along chromosome 11. This approach was set with windows of size 500 kb, with a shifting increment of 25 kb at each step. The q-value associated to each window was assumed as the 95% quantile of the q-values calculated for all SNPs in the window.

Note that this approach slightly differs from the previous one (PBS). The different implementations repeated for each dataset were intended to eliminate biases related to a specific method, its implementation, and also any bias related to the genetic data.

Variation in Ancient Humans. We checked the *FADS* genotypes SNPs in the ancient individuals accessing the information from VCF or BAM files; for this purpose we used the VariantAnnotation (<https://bioconductor.org/packages/release/bioc/html/VariantAnnotation.html>) and Samtools (www.htslib.org/) packages.

Geographic Analysis. Geographic maps displaying spatial variation in populations were obtained with inverse distance weighted using the ArcGIS 9.3 software (www.esri.com/software/arcgis/).

ACKNOWLEDGMENTS. We thank Marcelo Zagonel de Oliveira for technical assistance in the geographic analysis. This work was supported by a Science Without Borders fellowship from Conselho Nacional de Desenvolvimento Científico e Tecnológico Grant PDE 201145/2015-4 (to C.E.G.A.) and Fundação de Amparo à Pesquisa do Estado de São Paulo, Brazil Grant 2012/09950-9 (to K.N.).

- Fagundes NJ, 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* 82(3):583–592.
- Poznik GD, et al.; 1000 Genomes Project Consortium (2016) Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. *Nat Genet* 48(6):593–599.
- Hoffecker JF, Elias SA, O'Rourke DH (2014) Anthropology. Out of Beringia? *Science* 343(6174):979–980.
- Meiri M, et al. (2013) Faunal record identifies Bering isthmus conditions as constraint to end-Pleistocene migration to the New World. *Proc Biol Sci* 281(1776):20132167.
- Hopkins DM (1967) *The Bering Land Bridge* (Stanford Univ Press, Stanford, CA).
- Szathmari EJE (1993) Genetics of aboriginal North Americans. *Evol Anthropol* 1: 202–220.
- Bonato SL, Salzano FM (1997) A single and early migration for the peopling of the Americas supported by mitochondrial DNA sequence data. *Proc Natl Acad Sci USA* 94(5):1866–1871.
- Tamm E, et al. (2007) Beringian standstill and spread of Native American founders. *PLoS One* 2(9):e829.
- Kitchen A, Miyamoto MM, Mulligan CJ (2008) A three-stage colonization model for the peopling of the Americas. *PLoS One* 3(2):e1596.
- Fumagalli M, et al. (2015) Greenlandic Inuit show genetic signatures of diet and climate adaptation. *Science* 349(6254):1343–1347.
- Ameur A, et al. (2012) Genetic adaptation of fatty-acid metabolism: A human-specific haplotype increasing the biosynthesis of long-chain omega-3 and omega-6 fatty acids. *Am J Hum Genet* 90(5):809–820.
- Reich D, et al. (2012) Reconstructing Native American population history. *Nature* 488(7411):370–374.
- Visser WF, van Roermund CW, Waterham HR, Wanders RJ (2002) Identification of human PMP34 as a peroxisomal ATP transporter. *Biochem Biophys Res Commun* 299(3):494–497.
- Skoglund P, et al. (2015) Genetic evidence for two founding populations of the Americas. *Nature* 525(7567):104–108.
- Yi X, et al. (2010) Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* 329(5987):75–78.
- Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics* 180(2):977–993.

17. Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet* 5(10):e1000695.
18. Szpiech ZA, Hernandez RD (2014) selscan: An efficient multithreaded program to perform EHH-based scans for positive selection. *Mol Biol Evol* 31(10):2824–2827.
19. Rasmussen M, et al. (2010) Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* 463(7282):757–762.
20. Rasmussen M, et al. (2014) The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* 506(7487):225–229.
21. Fu Q, et al. (2014) Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* 514(7523):445–449.
22. Raghavan M, et al. (2014) Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* 505(7481):87–91.
23. Dulik MC, et al. (2012) Mitochondrial DNA and Y chromosome variation provides evidence for a recent common ancestry between Native Americans and Indigenous Altaians. *Am J Hum Genet* 90(2):229–246.
24. Cardona A, et al. (2014) Genome-wide analysis of cold adaptation in indigenous Siberian populations. *PLoS One* 9(5):e98076.
25. Harrit RK, et al. (1998) Paleo-Eskimo beginnings in North America: A new discovery at Kuzitrin Lake, Alaska. *Etud Inuit* 22:61–81.
26. Gilbert MT, et al. (2008) Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. *Science* 320(5884):1787–1789.
27. Hünemeier T, et al. (2012) Evolutionary responses to a constructed niche: Ancient Mesoamericans as a model of gene-culture coevolution. *PLoS One* 7(6):e38862.
28. González-José R, Bortolini MC, Santos FR, Bonatto SL (2008) The peopling of America: Craniofacial shape variation on a continental scale and its interpretation from an interdisciplinary view. *Am J Phys Anthropol* 137(2):175–187.
29. Bortolini MC, González-José R, Bonatto SL, Santos FR (2014) Reconciling pre-Columbian settlement hypotheses requires integrative, multidisciplinary, and model-bound approaches. *Proc Natl Acad Sci USA* 111(2):E213–E214.
30. Hoffecker JF, Elias SA, O'Rourke DH, Scott GR, Bigelow NH (2016) Beringia and the global dispersal of modern humans. *Evol Anthropol* 25(2):64–78.
31. Laurie CC, et al.; GENEVA Investigators (2010) Quality control and quality assurance in genotypic data for genome-wide association studies. *Genet Epidemiol* 34(6):591–602.
32. Delaneau O, Zagury JF, Marchini J (2013) Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods* 10(1):5–6.
33. Reynolds J, Weir BS, Cockerham CC (1983) Estimation of the coancestry coefficient: Basis for a short-term genetic distance. *Genetics* 105(3):767–779.
34. Hudson RR (2002) Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* 18(2):337–338.
35. Foll M, Gaggiotti OE, Daub JT, Vatsiou A, Excoffier L (2014) Widespread signals of convergent adaptation to high altitude in Asia and America. *Am J Hum Genet* 95(4):394–407.