How Genes Have Illuminated the History of Early Americans and Latino Americans

Andrés Ruiz-Linares

Department of Genetics, Evolution and Environment, University College London, London WC1E 6BT, United Kingdom

Correspondence: a.ruizlin@ucl.ac.uk

The American continent currently accounts for ~15% of the world population. Although first settled thousands of years ago and fitting its label as “the New World,” the European colonial expansion initiated in the late 15th century resulted in people from virtually every corner of the globe subsequently settling in the Americas. The arrival of large numbers of immigrants led to a dramatic decline of the Native American population and extensive population mixing. A salient feature of the current human population of the Americas is, thus, its great diversity. The genetic variation of the Native peoples that recent immigrants encountered had been shaped by demographic events acting since the initial peopling of the continent. Similarly, but on a compressed timescale, the colonial history of the Americas has had a major impact on the genetic makeup of the current population of the continent. A range of genetic analyses has been used to study both the ancient settlement of the continent and more recent history of population mixing. Here, I show how these two strands of research overlap and make use of results from other scientific disciplines to produce a fuller picture of the settlement of the continent at different time periods. The biological diversity of the Americas also provides prominent examples of the complex interaction between biological and social factors in constructing human identities and of the difficulties in defining human populations.

A multiplicity of research approaches have been used to explore the original settlement of the American continent, often focusing on three prominent questions: (1) the route of entry of the initial settlers, (2) their time of arrival, and (3) the pattern of subsequent migration. These questions have been approached with variable degrees of success using various types of genetic markers examined in “Native” populations, defined on anthropological grounds (particularly language). Early studies used information from blood groups and proteins

(Cavalli-Sforza et al. 1994) and were followed by DNA analyses mainly of mitochondrial DNA (mtDNA) (Forster et al. 1996; Tamm et al. 2007; Fagundes et al. 2008; Kitchen et al. 2008) and the Y chromosome (Lell et al. 1997; Bianchi et al. 1998; Karafet et al. 1999; Bortolini et al. 2003). The more recent studies have examined the human genome at increasing levels of resolution, from analyses with restricted sets of markers (Wang et al. 2007; Ray et al. 2010) to ongoing studies based on full genome sequences. Although a range of scenarios for the initial

A. Ruiz-Linares

peopling of the Americas have been envisaged, genetic evidence points to the continent being settled by people migrating into the northwestern tip of the continent from Asia. This migration would have been facilitated by the existence, at that time, of a land bridge connecting Siberia to Alaska, which later was submerged beneath the Bering Strait by the rising sea level at the end of the last glaciation, around 15,000 years ago (Fiedel 2000). Genetic support for an American settlement from Eastern Siberia includes the finding that Native Americans are genetically most similar to North Asians (Cavalli-Sforza et al. 1994; Wang et al. 2007) and the existence of a gradient of declining genetic diversity from northwest North America southward (Wang et al. 2007; Reich et al. 2012). This gradient extends beyond that seen in the “Old World” for populations at increasing distance from Africa, possibly resulting from a sequence of population contractions that occurred as small groups of humans moved from settled areas into uninhabited territories (Ramachandran et al. 2005; Handley et al. 2007; Wang et al. 2007). The American continent, being the last major landmass to have been settled by humans, shows a low genetic diversity as compared with all other continents (Wang et al. 2007).

Estimating the date of the initial settlement of the Americas has proven a difficult and contentious issue. Geological information provides a key reference point in that because of extensive ice sheets covering North America at the peak of the last glaciation (around 20,000 years ago), the continent would have been impenetrable then (Fig. 1). Therefore, this leaves two broad opportunities for settlement: before or after this last glacial maximum (LGM). Calculating the time of initial settlement of the continent from genetic information requires a number of assumptions of which the exact validity is difficult to assess, including variation in factors such as population demography, mutation rates, and the influence of selection. Perhaps, not surprisingly, the range of genetic estimates for the time of human settlement of America is quite wide, extending to both sides of the glacial maximum. It is, however, encouraging that most of the recent estimates, based on increasingly larger

amounts of data and more sophisticated statistical methods, point to a settlement not long after the LGM. These estimates show greater consistency with the archaeological evidence, which, although itself not devoid of controversy, points to a human presence in the Americas by ~14,000 years ago.

The pattern of migration into the continent has also been the subject of considerable disagreement. An influential model put forward in the mid-1980s posited that the settlement of the continent occurred in three sequential migratory waves from Asia, corresponding to the three major linguistic stocks in which the linguist Joseph Greenberg classified Native American languages (Greenberg et al. 1986; Greenberg 1987; Ruhlen 1991). The first migration would have given rise to a very large Amerind linguistic family comprising populations living all over the continent, whereas two subsequent migrations, restricted to North America and the Arctic, would be associated with populations speaking languages of the Na-Dene and Eskimo-Aleut linguistic families, respectively. Although early blood group and protein data were interpreted in support of the Greenberg model (Cavalli-Sforza et al. 1994), subsequent mtDNA and Y-chromosome analyses have been mostly interpreted as indicative of a single migration wave into the continent (Bonatto and Salzano 1997; Tamm et al. 2007; Fagundes et al. 2008; Kitchen et al. 2008). The recent genome-wide surveys of diversity with increasing resolution have, however, provided a different view. These are inconsistent with the single migration model and are more in line with the occurrence of multiple migrations (Fig. 1). Particularly strong support for several ancient migrations comes from a study based on a large survey of populations and using data for hundreds of thousands of genetic markers. With this type of data, it is possible to estimate the ancestry of every segment of DNA along the genome and state whether such a segment is of African, European, or Native American origin. Analyses can then focus only on the Native American segments of the genome (Fig. 2). This means that Native American individuals, and populations, that previously had to be excluded

Figure 1. First peopling of the American continent. Settlement is thought to have occurred from Eastern Siberia through several waves of migration (arrows) across a land bridge connecting Siberia to Alaska, existing at the time. Crossing was impossible during the last glacial maximum (LGM) (~20,000 years ago) because of glaciers covering a large part of North America. Most genetic studies of contemporary Native Americans point to a settlement of the continents soon after the LGM, subsequent to the retreat of the ice sheets. Although classical studies associated initial settlement with the Clovis archaeological complex of North America (~13,000 years ago), older sites have been identified, including Monte Verde in South America (dated at ~15,000 years ago). The Native American populations placed on this map are those included in the phylogenetic tree shown in Figure 3. Analysis of genetic data from these populations is consistent with the important role of the coast during the initial settlement of the continent (Reich et al. 2012).

from study because of admixture with non-Natives can now be included, facilitating a more extensive population survey and reducing bias. These recent data have provided strong evidence that the Eskimo-Aleut, Na-Dene, and Amerind linguistic groups show evidence of differential gene flow from Asia, inconsistent with stemming from a discrete single colonization event with no subsequent migration (Fig. 1). Noticeably, although North American populations show evidence of multiple episodes of gene exchange with Asia, Native populations from

Mexico to the Southern tip of South America appear to stem from one colonization wave with no subsequent Asian gene flow. This observation agrees with the highly controversial proposal of grouping widely separated Native American languages into a single “Amerindian” linguistic family (Greenberg 1987; Ruhlen 1991). These data also confirm the correlation of population diversity with distance from the Bering Strait, in agreement with settlement in a north-to-south direction. Interestingly, this correlation increases when considering the

A. Ruiz-Linares

Figure 2. Inference of local ancestry along the two copies of chromosome 1 in an admixed Native American individual. The height of the thick line indicates local ancestry as the number of chromosome copies at that position that are estimated to be Native American (0, 1, or 2). Numbers on the x-axis refer to the position along chromosome 1 (in kilobases) of the genetic markers allowing inference of local ancestry. (Modified from data in Reich et al. 2012.)

coasts as facilitators of population movement, suggesting an important role of the coast during the initial population dispersals on the continent. A phylogenetic tree relating the Native American populations examined in that survey is also consistent with the north-to-south settlement of the continent, as it shows a sequence of major population splits separating groups of populations mostly along a north-to-south axis (Figs. 1 and 3). Consistent with some degree of parallel evolution for languages and genes (Cavalli-Sforza et al. 1994), resulting from population separation followed by relative isolation, the major clusters of populations in this genetic tree show a broad correspondence with the linguistic affiliation of the populations (Fig. 3).

The recent study by Reich et al. (2012) illustrates the potential of high-density genotyping for extending studies focused on the original settlement of the Americas to Native individuals with evidence of admixture with recent immigrants. This admixture is extensive across the continent and involves not only Natives but also the general population, particularly in the countries of what is now referred to as Latin America. Historically, a major driver behind population mixing in this region was the fact

that immigrants from Spain and Portugal, particularly in the early phases of the colonial expansion, were mostly men (Boyd-Bowman 1973). It is well documented that many Conquistadors had children with Native women, the most famous example possibly being that of the Conquistador of Mexico, Hernán Cortez, and the Nahua woman known as “Malinche” (Fig. 4). This “sex-biased” pattern of mating between immigrant men and Native women had been alluded to by historians (Mörner 1967), but it was only with mtDNA and Y-chromosome studies that the full genetic impact of this feature became apparent. Because mtDNA and the Y chromosome are only transmitted by mothers and fathers, respectively, they allow the inference of the maternal (mtDNA) and paternal (Y-chromosome) ancestry of individuals. One of the first such studies was performed in Antioquia (Colombia), a population traditionally considered as mainly of Spanish descent. Consistent with this view, it was found that >90% of men in Antioquia had Y-chromosome lineages of European origin (Carvajal-Carmona et al. 2000). Surprisingly, when examining their mtDNA, a sharply different picture was observed. In 90% of individuals, maternal

Genes and the History of Americans

Figure 3. Phylogenetic tree relating representative African, European, Oceanian, East Asian, and Native American populations based on high-density genetic marker data. For this analysis, Native American data used was restricted to genome segments of confirmed Native ancestry (determined as in Fig. 2). The tree branches are color-coded to represent the linguistic affiliation of the populations, as shown in the inset. Numbers in parentheses refer to sample size in each population. The length of the branches on this tree is proportional to a measure of genetic differentiation (F_{ST}). (From Reich et al. 2012; reprinted, with permission, from the author.)

A. Ruiz-Linares

Figure 4. The Native American woman known as “Malinche” or Malintzin (her Nahuatl name) was the interpreter and mistress of the Spanish Conquistador, Hernán Cortés. In 1523, she gave him a son, Martín, who is one of the first recorded individuals of mixed Native–European ancestry born in the Americas. Such offspring between immigrant men and Native women were a common occurrence in early colonial Latin America. The drawing shown is from the late 16th century “Codex Tlaxcala” and represents a meeting between the Mexican ruler, Moctezuma, and Hernán Cortés, with Malintzin (on the right) translating.

ancestry was Native American (Carvajal-Carmona et al. 2000). Similar observations have now been made in many Latin American countries (Alves-Silva et al. 2000; Green et al. 2000; Carvalho-Silva et al. 2001; Marrero et al. 2007), although with a considerable variation in ancestry proportions between them (Fig. 5). These studies, in addition, show a higher African ancestry with mtDNA than the Y chromosome, indicating that, historically, admixture with Africans has also mostly involved African women.

The large variation in ancestry seen across Latin America relates to differences in pre-Columbian Native population density and the pattern of recent immigration into specific regions of the continent. For instance, most studies performed so far have mainly focused on areas with little documented African immigration and consistently show a relatively low African

genetic ancestry. In these population samples, the variation in individual European and Native American ancestry is very large, to the extent that it overlaps with that seen in Native population samples (Fig. 6). The variation in individual ancestry seen in these samples thus effaces their designation as “Native” or “non-Native.” This observation punctuates the interest of incorporating admixed Latin American populations, traditionally considered non-Native, into studies on the initial settlement of the continent. Similar to what has been performed in the recent survey by Reich et al. (2012), the inference of ancestry of each genome segment in Latin Americans could be used to focus solely on Native American segments of the genome. This is an avenue of research that is just beginning to be explored and shows great potential for the future. It promises to be of particular importance for the analysis of regions where anthropologically recognizable Native populations and individuals are virtually nonexistent, as they have been absorbed into the current mixed population. This is the case for the many areas that were relatively sparsely populated in pre-Columbian times and, subsequently, received a large flow of immigrants, such as from the Caribbean and many parts of North and South America. Consequently, estimation of individual ancestry along the genome will facilitate denser demographic history analyses across the Americas, as well as a reexamination of the original settlement of the continent based on a more comprehensive population sampling.

Other than being informative for addressing questions of population history, the study of Latin American populations promises to facilitate the genetic characterization of biological attributes differentiated among the populations that participated in admixture on the continent. For instance, a range of facial features differ between Native Americans and Europeans, and the genetic study of admixed Latin Americans promises to help in the identification of genes explaining variation in facial appearance. Such research is of interest for understanding disorders of craniofacial development and could also have forensic applications. Another example is type 2 diabetes (T2D), a dis-

Genes and the History of Americans

Figure 5. Proportion of Native American, European, and African ancestry in 13 Latin American populations estimated using mtDNA and Y-chromosome markers. Samples from urban centers in five countries were examined (Mexico—Mexico City; Guatemala—Oriente; Costa Rica—Central Valley of Costa Rica [CVCR]; Colombia—Peque, Medellín, and Cundinamarca; Chile—Paposo and Quetalmahue; Argentina—Salta, Tucuman, and Tacamarca; and Brazil—Rio Grande do Sul [RGS]). Ancestry proportion (fraction of the pie chart) is color-coded: African (green), European (blue), and Native American (red). (Data for 20 individuals per population are from Wang et al. 2007, Yang et al. 2010, and NN Yang et al. unpubl.)

ease that has a very high frequency in Native Americans and for which a higher risk is associated with increased Native American ancestry. This observation led to the proposal of the “thrifty genotype” hypothesis, which posits that the increased risk of T2D in Native Americans results from genetic adaption to a low-calorie/high-exercise way of life that became detrimental with the recent change to a high-calorie/

low-exercise lifestyle (Pollard 2008). The study of large, carefully characterized samples from Latin American populations offers a unique opportunity for conducting a detailed assessment of this hypothesis. The identification of genes explaining the variable frequency of diseases between populations (such as T2D) will be an important step forward in the development of novel, more effective (even individualized) dis-

Figure 6. Distribution of individual Native ancestry estimated in samples from Native Americans (shown in blue) and two Latin American urban centers (850 residents of Medellín, Colombia shown in red, and 220 residents of Mexico City, Mexico shown in green). For each of the three population samples, the x-axis indicates the proportion of Native ancestry (in 0.1 unit intervals) and the y-axis indicates the proportion of individuals with that ancestry estimate. The Native American group includes 225 individuals from North, Central, and South America. Ancestry proportions were estimated using autosomal genetic markers. (Data from Florez et al. 2009, Campbell et al. 2011, and Reich et al. 2012.)

A. Ruiz-Linares

Figure 7. Box plots displaying the relationship of individual genetic ancestry estimates to self-perceived ancestry in 7342 Latin Americans (from Mexico, Colombia, Peru, Chile, and Brazil). Self-perception was categorized into 20% bands for African, European, and Native American ancestry. There is a highly significant correlation between the genetic estimate and self-perception for each continental ancestry component. However, there is a trend at higher Native American and African ancestry for self-perception to exceed the genetic estimates. Correspondingly, at lower European ancestry, there is a trend for the genetic estimates to exceed self-perception. Further analyses show that pigmentation impacts on these differences. Individuals with lower skin pigmentation tend to overestimate their European ancestry, whereas individuals with higher pigmentation overestimate their Native American and African ancestries. Orange lines indicate the median and the blue boxes are delimited by the 25th and 75th percentiles. (From AR Ruiz-Linares et al., in press; with permission from the author.)

ease-management strategies that account for human population diversity.

The overlap of individual genetic ancestry estimates, seen in Latin and Native American populations (Fig. 5), raises the question of the relationship of these estimates to the perception that individuals have of their own ancestry. A recent analysis of a large sample of individuals from five Latin American countries found a highly significant correlation between self-perception and genetically estimated ancestry (Fig. 7). However, this study also found evidence that self-perception is biased. A particularly clear bias involves pigmentation: individuals with greater pigmentation tend to overestimate their Native and African ancestry, whereas individuals with lighter pigmentation tend to overestimate their European ancestry (AR Ruiz-Linares, in press). Statistically significant differences were also observed between countries, pointing to the influence of social factors in self-perception. Consistent with this observation, social scientists have argued that, in Latin America, self-identification as Native or non-Native is often strongly influenced by social cues (Wade 2010).

The insights into the initial settlement of the continent provided by the genetic study of Native Americans illustrate the fact that a popula-

tion sampling that maximizes diversity based on anthropological grounds (such as language) can facilitate investigations concerned with the first settlement of the continent. However, the analysis of intercontinental admixture both in Native and non-Native Latin American populations show some of the complexities of defining human groups, with population labels suggesting a potentially misleading genetic singularity. The biological reality is that of a gradient in the genetic makeup of these populations (and individuals) involving various degrees of mixture between the initial settlers of the continent and more recent immigrants. The genetic diversity of Latin Americans is, thus, a prominent example of the fuzzy meaning of the labels used to refer to human populations. Although these labels can assist in study design and facilitate certain historical inferences, ethnicity, race, and other such terms are social constructs devoid of a clear-cut biological meaning.

REFERENCES

- Alves-Silva J, da Silva Santos M, Guimarães PE, Ferreira AC, Bandelt HJ, Pena SD, Prado VE. 2000. The ancestry of Brazilian mtDNA lineages. *Am J Hum Genet* **67**: 444–461.
- Bianchi NO, Catanesi CI, Bailliet G, Martinez-Marignac VL, Bravi CM, Vidal-Rioja LB, Herrera RJ, López-Camelo JS.

1998. Characterization of ancestral and derived Y-chromosome haplotypes of New World native populations. *Am J Hum Genet* **63**: 1862–1871.
- 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* **94**: 1866–1871.
- Bortolini MC, Salzano FM, Thomas MG, Stuart S, Nasanen SP, Bau CH, Hutz MH, Layrisse Z, Petzl-Erler ML, Tsuneto LT, et al. 2003. Y-chromosome evidence for differing ancient demographic histories in the Americas. *Am J Hum Genet* **73**: 524–539.
- Boyd-Bowman P. 1973. Patterns of Spanish emigration to the New World (1493–1580). Council on International Studies, State University of New York at Buffalo, Buffalo, NY.
- Carvajal-Carmona LG, Soto ID, Pineda N, Ortiz-Barrientos D, Duque C, Ospina-Duque J, McCarthy M, Montoya P, Alvarez VM, Bedoya G, et al. 2000. Strong Amerind/white sex bias and a possible Sephardic contribution among the founders of a population in northwest Colombia. *Am J Hum Genet* **67**: 1287–1295.
- Carvalho-Silva DR, Santos FR, Rocha J, Pena SD. 2001. The phylogeography of Brazilian Y-chromosome lineages 1. *Am J Hum Genet* **68**: 281–286.
- Cavalli-Sforza LL, Menozzi P, Piazza A. 1994. *The history and geography of human genes*. Princeton University Press, Princeton, NJ.
- Fagundes NJ, Kanitz R, Eckert R, Valls AC, Bogo MR, Salzano FM, Smith DG, Silva WA Jr, Zago MA, Ribeiro-dos-Santos AK, 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**: 583–592.
- Fiedel SJ. 2000. The peopling of the New World: Present evidence, new theories, and future directions. *J Archaeol Res* **8**: 39–103.
- Forster P, Harding R, Torroni A, Bandelt HJ. 1996. Origin and evolution of Native American mtDNA variation: A reappraisal. *Am J Hum Genet* **59**: 935–945.
- Green LD, Derr JN, Knight A. 2000. mtDNA affinities of the peoples of North-Central Mexico. *Am J Hum Genet* **66**: 989–998.
- Greenberg JH. 1987. *Language in the Americas*. Stanford University Press, Stanford, CA.
- Greenberg JH, Turner CG II, Zegura SL. 1986. The settlement of the Americas: A comparison of the linguistic, dental, and genetic evidence. *Curr Anthropol* **27**: 477–497.
- Handley LJ, Manica A, Goudet J, Balloux F. 2007. Going the distance: Human population genetics in a clinal world. *Trends Genet* **23**: 432–439.
- Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, Goldman D, Klitz W, Harihara S, de Knijff P, et al. 1999. Ancestral Asian source(s) of new world Y-chromosome founder haplotypes. *Am J Hum Genet* **64**: 817–831.
- Kitchen A, Miyamoto MM, Mulligan CJ. 2008. A three-stage colonization model for the peopling of the Americas. *PLoS ONE* **3**: e1596.
- Lell JT, Brown MD, Schurr TG, Sukernik RI, Starikovskaya YB, Torroni A, Moore LG, Troup GM, Wallace DC. 1997. Y chromosome polymorphisms in Native American and Siberian populations: Identification of Native American Y chromosome haplotypes. *Hum Genet* **100**: 536–543.
- Marrero AR, Bravi C, Stuart S, Long JC, Pereira das Neves Leite F, Kommers T, Carvalho CM, Pena SD, Ruiz-Linares A, Salzano FM, et al. 2007. Pre- and post-Columbian gene and cultural continuity: The case of the Gaúcho from southern Brazil. *Hum Hered* **64**: 160–171.
- Morner M. 1967. *Race mixture in the history of Latin America*. Little Brown, New York.
- Pollard TM. 2008. *Western diseases: An evolutionary perspective*. Cambridge University Press, New York.
- 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* **102**: 15942–15947.
- Ray N, Wegmann D, Fagundes NJ, Wang S, Ruiz-Linares A, Excoffier L. 2010. A statistical evaluation of models for the initial settlement of the American continent emphasizes the importance of gene flow with Asia. *Mol Biol Evol* **27**: 337–345.
- Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, Ray N, Parra MV, Rojas W, Duque C, Mesa N, et al. 2012. Reconstructing Native American population history. *Nature* **488**: 370–374.
- Ruhlen M. 1991. *A guide to the world's languages*. Stanford University Press, Stanford, CA.
- Tamm E, Kivisild T, Reidla M, Metspalu M, Smith DG, Mulligan CJ, Bravi CM, Rickards O, Martinez-Labarga C, Khusnutdinova EK, et al. 2007. Beringian standstill and spread of Native American founders. *PLoS ONE* **2**: 1–6.
- Wade P. 2010. *Race and ethnicity in Latin America*. Pluto, London.
- Wang S, Lewis CM, Jakobsson M, Ramachandran S, Ray N, Bedoya G, Rojas W, Parra MV, Molina JA, Gallo C, et al. 2007. Genetic variation and population structure in native Americans. *PLoS Genet* **3**: e185.

Cold Spring Harbor Perspectives in Biology

How Genes Have Illuminated the History of Early Americans and Latino Americans

Andrés Ruiz-Linares

Cold Spring Harb Perspect Biol 2015; doi: 10.1101/cshperspect.a008557 originally published online September 25, 2014

Subject Collection [Human Variation](#)

Perspectives on Human Variation through the Lens of Diversity and Race

Aravinda Chakravarti

A Genomic View of the Peopling and Population Structure of India

Partha P. Majumder and Analabha Basu

Demographic Events and Evolutionary Forces Shaping European Genetic Diversity

Krishna R. Veeramah and John Novembre

Social Diversity in Humans: Implications and Hidden Consequences for Biological Research

Troy Duster

How Genes Have Illuminated the History of Early Americans and Latino Americans

Andrés Ruiz-Linares

Can Genetics Help Us Understand Indian Social History?

Romila Thapar

Genetic Variation and Adaptation in Africa: Implications for Human Evolution and Disease

Felicia Gomez, Jibril Hirbo and Sarah A. Tishkoff

What Type of Person Are You? Old-Fashioned Thinking Even in Modern Science

Kenneth M. Weiss and Brian W. Lambert

For additional articles in this collection, see <http://cshperspectives.cshlp.org/cgi/collection/>

**All Modifications and
Oligo Types Synthesized**
Long Oligos • Fluorescent • Chimeric • DNA • RNA • Antisense

Oligo Modifications?
Your wish is our command.