Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Genetics & Development

# A genomic view of the peopling of the Americas

Pontus Skoglund[1,2,3] and David Reich[1,2,4]

Whole-genome studies have documented that most Native American ancestry stems from a single population that diversified within the continent more than twelve thousand years ago. However, this shared ancestry hides a more complex history whereby at least four distinct streams of Eurasian migration have contributed to present-day and prehistoric Native American populations. Whole genome studies enhanced by technological breakthroughs in ancient DNA now provide evidence of a sequence of events involving initial migrations from a structured Northeast Asian source population with differential relatedness to present-day Australasian populations, followed by a divergence into northern and southern Native American lineages. During the Holocene, new migrations from Asia introduced the Saqqaq/Dorset Paleoeskimo population to the North American Arctic ~4500 years ago, ancestry that is potentially connected with ancestry found in Athabaskan-speakers today. This was then followed by a major new population turnover in the high Arctic involving Thule-related peoples who are the ancestors of present-day Inuit. We highlight several open questions that could be addressed through future genomic research.

**Addresses**
[1] Harvard Medical School, Boston, MA, USA
[2] Broad Institute of MIT and Harvard, Cambridge, MA, USA
[3] Department of Archaeology and Classical History, Stockholm, Sweden
[4] Howard Hughes Medical Institute, Boston, MA, USA

Corresponding author: Skoglund, Pontus (pontus.skoglund@gmail.com)

**Current Opinion in Genetics & Development** 2016, **41**:27–35

This review comes from a themed issue on **Genetics of human origin**

Edited by **Joshua M Akey** and **Anna Di Rienzo**

For a complete overview see the Issue and the Editorial

Available online 6th August 2016

http://dx.doi.org/10.1016/j.gde.2016.06.016



## Introduction

The peopling of the Americas represented the culmination of a Late Pleistocene expansion of anatomically modern humans out of Africa. Archaeological evidence indicates that groups subsisting on hunting lived in extreme northeast Siberia (71° N) by at least 28,000 years ago [1]. Human groups adapted to the mammoth steppe habitat were thus poised to enter Beringia — the landmass between Alaska and Eurasia that is now submerged — by this time [2,3]. The path from Beringia to the more temperate parts of the American continents, however, was blocked by the merged Laurentide and Cordilleran ice sheets that covered northern North America until the end of the Last Glacial Maximum. The ice retreated from parts of the Pacific coast ~16,000 years ago, raising the possibility of a coastal migration after this time, and within a few thousand years a habitable corridor through the center of the continent opened between the two ice sheets [4]. The first unambiguous evidence of modern humans in the Americas dates to between 14,000 and 15,000 years ago [5–8], and was likely the consequence of migration from Beringia.

Major debates about the peopling of the Americas have focused on the question of whether the first early human populations in the Americas are directly ancestral to present-day Native Americans, as well as on the number, mode, and timing of secondary migrations to the Americas. Advances in genomics have, within the last five years, enabled the collection of far more data from present-day Native Americans than were available previously, and have also made it possible for the first time to access DNA from ancient Native American remains. Analysis of these data has highlighted a minimum of four distinct streams of Asian ancestry, some of which were not clear from the archaeological evidence. We review the contributions of genomic data to understanding the prehistory of the Americas, and outline outstanding questions where it may be able to provide additional insight.

## The power of the whole genome

The first meaningful genetic insights about Native American population history came from mitochondrial DNA, a segment of about 16,500 base pairs (approximately 1/200,000th of the genome) that is passed exclusively along the maternal line. Mitochondrial DNA was one of the first parts of the genome to be heavily investigated to learn about human population history for several reasons. First, it is highly variable on a per-nucleotide level and thus sequencing only a short stretch can detect non-trivial amounts of human variation. Mitochondrial DNA's high variability and short length also meant that it was practical to sequence in large numbers of samples at a time when it was prohibitively expensive to generate genome scale data. For ancient DNA studies, mitochondrial DNA had the further advantage that it exists in about a thousand-fold higher copy number than any other single place in the genome. Since one of the main challenges of ancient DNA is obtaining sufficient amount of material from any position in the genome to be able to analyze, beginning with more material can be an advantage [9].

The greatest contribution of mitochondrial DNA studies to the understanding of Native American prehistory has been in the area of reconstruction of population size history. Mitochondrial DNA analyses were the first to document that the ancestry of most Native Americans derives from a population that experienced a profound founder event [10], with a relatively small number of individuals giving rise to a large number of descendants today. The evidence for this is that all Native American mitochondrial DNA lineages today descend from just five founding maternal lines [11–13] that each had a common ancestor around 18,000 to 15,000 years ago, implying a population size bottleneck around this time [14–18]. The evidence for a profound population bottleneck has since been confirmed and its intensity measured more accurately with genome scale data [19–22,23••], but it is important to note that there are still challenges with disentangling the number of founder individuals from the duration of the population size reduction using all the reported methods.

A second finding about Native American population history based on mitochondrial DNA data is that the founder event may have been preceeded by an extended period (many thousands of years) of little or no shared ancestry with non-Native American mitochondrial DNA lineages. This suggested to some researchers the hypothesis of a 'Beringian standstill', whereby the first founding population of the Americas was isolated from Eurasian populations before its radiation into a multitude of subpopulations in America [15].

At the same time, some observations from mitochondrial DNA studies of the Americas have deceptive. For example, the mitochondrial DNA subtype called D4h3a is today almost entirely restricted to Pacific coastal populations, both in North and South America. This observation led to the hypothesis that D4h3a was carried south of the ice sheets along a coastal route, in a migratory movement that was distinctive from the one that led to many other Native American populations [24]. However, ancient DNA studies have since found the same mitochondrial DNA type in a ~12,600 year old individual from present-day Montana, which based on its genome-wide data is unambiguously from the main ancestral lineage leading to most Native Americans [25••].

It is now clear that so many founder events and fluctuations in population size have occurred before, during, and after the peopling of the Americas that the evidence from one position in the genome — mitochondrial DNA, the Y chromosome, or any other location — is too subject to random changes in frequency (genetic drift) to provide a complete picture by itself. Only by taking the independent testimony of many locations in the genome simultaneously can we obtain a high-resolution picture of the deep past. The remainder of this article focuses on insights from whole genome studies of Native American population history. While these studies are still in their early days, they have already upended our understanding of key events. Application of ancient DNA technology promises further insights in years to come.

## Sources of Native American ancestry

Under the hypothesis that Native American ancestry stems from a single founder population that separated earlier from Eurasian populations, differences in allele frequencies between Native American groups should have developed independently from Eurasian allele frequencies. This simple null hypothesis makes it possible to explicitly test hypotheses about the number of American founder populations. Reich *et al.* [23••] applied this idea to the first comprehensive genome-wide data from Native American populations (52 populations, but none from the continental United States), and concluded that at least three ancestral populations were required to explain the similarities between Native Americans and East Asians. According to the initial study [23••], all Native American groups from Central and South America fit a model of a single founder population. An additional source of ancestry was necessary to account for genetic variation in Eskimo-Aleut speakers. In addition, analysis of the Athabaskan-speaking Chipewyan revealed that they could not solely have their ancestry from the same founding population as other Northern-, Meso- and South American populations.

## The main ancestral stream giving rise to Native American ancestry

One of the most important pieces of genetic evidence relevant to the peopling of the Americas was the sequencing of a genome from the remains of a child ('Anzick-1') buried with Clovis artifacts in western Montana and directly dated to 12,600 before present (BP) [25••]. This child was consistent with deriving all of his ancestry from the same founding population as Central and South Americans (Figure 1), contradicting the 'Solutrean hypothesis' of transatlantic migration from Upper Paleolithic Europe giving rise to the Clovis industry [26] which thus posits genetic discontinuity between the makers of the Clovis industry and present-day Native Americans [27]. The most surprising finding was that the Anzick individual is from a population more closely related to Central Americans and South Americans than to some northern North Americans (including all speakers of Algonquian languages studied to date), despite the apparent common ancestral origin of Native Americans across the continents. This suggests that the present-day structuring of the main ancestry in Native Americans into northern and southern lineages [23••] dates back to more than 12,600 years ago [25••]. This event divided the ancestry of present-day Native Americans into two main streams, one of which includes the ancestors of present-day Northern Native Americans