

# The Paleo-Indian Entry into South America According to Mitogenomes

Stefania Brandini,<sup>1</sup> Paola Bergamaschi,<sup>1,2</sup> Marco Fernando Cerna,<sup>3</sup> Francesca Gandini,<sup>1,4</sup> Francesca Bastaroli,<sup>1</sup> Emilie Bertolini,<sup>1</sup> Cristina Cereda,<sup>5</sup> Luca Ferretti,<sup>1</sup> Alberto Gómez-Carballa,<sup>6,7,8</sup> Vincenza Battaglia,<sup>1</sup> Antonio Salas,<sup>6,7</sup> Ornella Semino,<sup>1</sup> Alessandro Achilli,<sup>1</sup> Anna Olivieri,<sup>1</sup> and Antonio Torroni\*,<sup>1</sup>

<sup>1</sup>Dipartimento di Biologia e Biotechnologie, Università di Pavia, Pavia, Italy

<sup>2</sup>Servizio di Immunoematologia e Medicina Trasfusionale, Fondazione IRCCS Policlinico San Matteo, Pavia, Italy

<sup>3</sup>Biotechnology Laboratory, Salesian Polytechnic University of Ecuador, Quito, Ecuador

<sup>4</sup>Department of Biological Sciences, School of Applied Sciences, University of Huddersfield, Huddersfield, United Kingdom

<sup>5</sup>Genomic and Post-Genomic Center, National Neurological Institute C. Mondino, Pavia, Italy

<sup>6</sup>Departamento de Anatomía Patológica e Ciencias Forenses, Instituto de Ciencias Forenses, Facultade de Medicina, Universidade de Santiago de Compostela, Unidade de Xenética, Galicia, Spain

<sup>7</sup>GenPoB Research Group, Instituto de Investigaciones Sanitarias (IDIS), Hospital Clínico Universitario de Santiago (SERGAS), Unidade de Xenética, Galicia, Spain

<sup>8</sup>Grupo de Investigación en Genética, Vacunas, Infecciones y Pediatría (GENVIP), Hospital Clínico Universitario and Universidade de Santiago de Compostela, Galicia, Spain

\*Corresponding author: E-mail: antonio.torroni@unipv.it.

Associate editor: Anne Stone

## Abstract

Recent and compelling archaeological evidence attests to human presence ~14.5 ka at multiple sites in South America and a very early exploitation of extreme high-altitude Andean environments. Considering that, according to genetic evidence, human entry into North America from Beringia most likely occurred ~16 ka, these archeological findings would imply an extremely rapid spread along the double continent. To shed light on this issue from a genetic perspective, we first completely sequenced 217 novel modern mitogenomes of Native American ancestry from the northwestern area of South America (Ecuador and Peru); we then evaluated them phylogenetically together with other available mitogenomes (430 samples, both modern and ancient) from the same geographic area and, finally, with all closely related mitogenomes from the entire double continent. We detected a large number ( $N = 48$ ) of novel subhaplogroups, often branching into further subclades, belonging to two classes: those that arose in South America early after its peopling and those that instead originated in North or Central America and reached South America with the first settlers. Coalescence age estimates for these subhaplogroups provide time boundaries indicating that early Paleo-Indians probably moved from North America to the area corresponding to modern Ecuador and Peru over the short time frame of ~1.5 ka comprised between 16.0 and 14.6 ka.

**Key words:** Native Americans, mitochondrial DNA, mitochondrial genomes, haplogroups, first peopling of South America, Ecuador, Peru.

## Introduction

The initial peopling of the Americas is a long-standing topic of debate, which has been fueled over the years by findings at numerous archaeological sites all over the double continent. The sites of Monte Verde in southern Chile and Pedra Furada in northeastern Brazil have played a major role in this debate. Since their preliminary excavations, they raised the possibility of human presence in South America, as far south as the Southern Cone, by 14–15 ka or even earlier, during the final Pleistocene (Guidon and Delibrias 1986; Dillehay 1989; Parenti et al. 1990, 1996; Dillehay and Collins 1988). In the scenario that the first Southern Americans were an offshoot of the initial settlers who moved from Beringia first to North

America and then to Central America, a view supported by genetic evidence (Greenberg et al. 1986; Schroeder et al. 2009; Reich et al. 2012), the postulated dates of these sites implied that (1) the Clovis people of North America (~13 ka) were not the first Americans and (2) the spread of Paleo-Indians throughout the Americas could have occurred very rapidly.

In recent years, compelling archeological evidence has indeed shown that people occupied North America prior to Clovis (Gilbert et al. 2008a; Waters et al. 2011) and findings at multiple sites in South America have confirmed that Paleo-Indians not only had already spread through the southern subcontinent (Fraser 2014; Boëda et al. 2014; Aimola et al. 2014; Dillehay et al. 2017), but had also colonized extreme high-altitude Andean environments (Rademaker et al. 2014)

© The Author 2017. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

Open Access

at the time in which the Clovis technological complex was developing in North America.

In the genetic arena, studies of mitochondrial DNA (mtDNA) have extensively contributed to the current view that people entered North America ~16 ka (Schurr and Sherry 2004; Fagundes et al. 2008; O'Rourke and Raff 2010; Achilli et al. 2013; Llamas et al. 2016), after a period of isolation in Beringia (Tamm et al. 2007; Raghavan et al. 2015; Tackney et al. 2015; Hoffecker et al. 2016) that had a major role in the shaping of the first settlers' genetic diversity (Perego et al. 2009). As a single maternally-inherited locus, mtDNA often does not reflect the whole complexity of past demographic processes, but allows an extremely detailed reconstruction of the nesting relationships within a phylogeny, a feature that can be extremely informative for dating migration and population separation events, especially when the sequence variation of large data sets of entire mitogenomes is considered (Richards et al. 2016).

In the early 1990s, studies based on RFLPs and mtDNA control-region variation suggested that Native Americans exhibit a low variability when compared with other continental contexts, with only four haplogroups, initially named A, B, C, and D (Schurr et al. 1990; Torroni et al. 1992, 1993) later relabeled as A2, B2, C1, and D1 (Forster et al. 1996), encompassing the vast majority of mtDNAs in the entire double continent. Subsequent studies, mostly carried out at the level of entire mitogenomes, allowed the phylogenetic dissection of the four major haplogroups and the identification of additional rare haplogroups bringing the overall number of maternal founding lineages of Asian/Beringian origin to 16 (Tamm et al. 2007; Perego et al. 2010). Among these, eight (A2, B2, C1b, C1c, C1d, C1d1, D1, and D4h3a) are often defined as "pan-American," as they are found across the double continent. The others are either extremely rare (X2g and D4e1; Perego et al. 2009; Kumar et al. 2011) or generally restricted to the populations of the arctic and subarctic regions of North America. Among the latter, some haplogroups such as C4c, interestingly detected also in the Ijka-speakers from Colombia (Tamm et al. 2007), and X2a might have arrived from Beringia with Paleo-Indian groups that followed alternative migration routes (Perego et al. 2009; O'Rourke and Raff 2010; Hooshiar Kashani et al. 2012) whereas others (A2a, A2b, D2a, and D3) are the result of much later arrivals (Gilbert et al. 2008b; Achilli et al. 2013; Raghavan et al. 2014).

In addition to the founding haplogroups of Beringian/Asian origin, mitogenome analyses have identified a few subhaplogroups whose geographical distributions and estimated ages indicate an *in situ* origin in the Americas shortly after or within a few millennia from the initial peopling. Currently, they include: (1) B2b, which is shared between North and South America and has been preliminarily dated to ~21 ka on the basis of 14 complete mitogenomes (Taboada-Echalar et al. 2013); (2) B2a, dated to ~11–13 ka and whose distribution is restricted to the US and Mexico, with traces in Canada (Achilli et al. 2013); and (3) four subhaplogroups (D1g, D1j, C1b13, and B2i2; 11–16 ka) that have been identified in the Southern Cone of South America (Bodner et al. 2012; de Saint Pierre et al. 2012). The geographical distribution of B2b is best

explained with an origin in North American Paleo-Indians (Taboada-Echalar et al. 2013) while they were expanding southward. In this scenario B2b was carried first to Meso-America and then to South America together with A2, B2, C1b, C1c, C1d, C1d1, D1, and D4h3a of Beringian origin. The other five subhaplogroups instead most likely arose sometime later, B2a in North America and the others in the Southern Cone, after that the front of the expansion wave had already passed through, thus remaining mostly confined to the geographic area where each arose, especially if the processes of adaptation to the different environments, as suggested by archeological evidence (Rademaker et al. 2014), and tribalization (Torroni et al. 1993) of Paleo-Indian settlers began early.

A recent study of mitogenomes in the Mediterranean basin has shown that, if the identification and dating of autochthonous subhaplogroups that arose *in situ* shortly after the first peopling event is accompanied also by the identification and dating of a close upstream node in the phylogeny which instead arose somewhere else (outside the area of interest) prior to the colonization event, minimum and maximum times for the presence of autochthonous subhaplogroups in the area can be estimated (Olivieri et al. 2017). In other words, this approach can provide rather narrow time boundaries for the peopling event.

To shed light on the entry time of Paleo-Indians into South America from a genetic perspective, we applied the same rationale to the present study. Our objective was to identify and accurately date as many subhaplogroups as possible of the two classes mentioned above: those that arose in South America early after the peopling event, and those—as B2b—that instead arose in North or Central America after the human entry from Beringia, but prior to the peopling of South America. For this reason, we first completely sequenced 217 novel modern mitogenomes of Native American ancestry from the northwestern area of South America (Ecuador and Peru), we then evaluated them phylogenetically together with all available mitogenomes (both modern and ancient) from the same geographic area and, finally, with all closely related mitogenomes from the entire double continent. These analyses allowed the detection of numerous novel subhaplogroups belonging to two classes. Coalescence age estimates for these subhaplogroups are in agreement with archeological evidence attesting to human presence in South America ~14.5 ka and provide time boundaries indicating that early Paleo-Indians probably moved from North to South America over the short time frame of ~1.5 ka.

## Results

### Mitogenome Variation in Northwestern South America

To survey mitogenome variation in northwestern South America, DNA was obtained from 217 Ecuadorians (93 Native Americans and 124 Mestizos), representatives of all major regions of Ecuador and ten (all from Mestizos) Peruvian individuals (supplementary tables S1 and S2, Supplementary Material online). An initial preliminary survey of the mtDNA

control-region showed that 208 (96%) of the Ecuadorians harbor mitogenomes of Native American ancestry. The founder pan-American haplogroup B2 is very common in northwestern South America (52.1%), but also all others (A2, C1b, C1c, C1d, D1, and D4h3a) were detected (supplementary table S3, Supplementary Material online). Only nine mitogenomes, all from Mestizos, were members of Old World haplogroups: L2a1 ( $N = 1$ ), L3e2b ( $N = 4$ ), R0a ( $N = 2$ ), U2d3 ( $N = 1$ ), and U5b3f ( $N = 1$ ). These findings are consistent with the multiple ancestral sources (Native Americans, Europeans, and Africans) that have contributed to the formation of the modern Ecuadorian population (González-Andrade et al. 2007; Santangelo et al. 2017). As for Peru, one of the ten mtDNA control regions was classified into an Old World haplogroup, the European U5a1a1.

The 217 mitogenomes of Native American ancestry mentioned above (208 from Ecuador and nine from Peru) underwent complete sequencing and were employed, together with 430 previously published Ecuadorian and Peruvian mitogenomes from both modern and ancient samples (supplementary table S4, Supplementary Material online), to reconstruct the phylogenies of the pan-American haplogroups A2, B2, C1b, C1c, C1d, D1, and D4h3a in northwestern South America.

### The Phylogenies of the Pan-American Haplogroups in Ecuador and Peru

The phylogenetic relationships of the 647 mitogenomes are illustrated in supplementary figures S1–S5, Supplementary Material online. Each figure includes one or more of the founding pan-American haplogroups, with the exception of supplementary figure S3, Supplementary Material online, which encompasses the mitogenomes belonging to B2b ( $N = 137$ ), a subbranch of B2 that is extremely common in Ecuador and Peru (21.2%). Within each of the pan-American haplogroups, we detected extensive differentiation into derived branches (supplementary table S5, Supplementary Material online), including a large number ( $N = 48$ ) of novel subhaplogroups (in blue in supplementary figures S1–S5, Supplementary Material online), often branching into further subclades. Note that new subhaplogroups were defined and named only when they encompassed a minimum of three haplotypes sharing at least one stable mutation.

The frequencies of subhaplogroups in modern Ecuadorians and Peruvians were determined by excluding mitogenomes from nonrandom population surveys (supplementary table S6, Supplementary Material online). Despite some differences, in particular for D1 (6.3% in Ecuador vs. 14.9% in Peru), frequencies at the level of founding pan-American haplogroups are rather similar in the two countries. We also assessed some mitogenome diversity parameters in the two geographic areas (supplementary table S7, Supplementary Material online) employing the same data set used to assess haplogroup frequencies. Also in this case we did not detect major differences.

We also included 68 previously published ancient mitogenomes in our analyses, all from Peru (supplementary table S4, Supplementary Material online; Gómez-Carballa et al. 2015;

Fehren-Schmitz et al. 2015; Llamas et al. 2016). Similar to present-day mitogenomes, they encompassed all founding pan-American haplogroups, except for the relatively rare D4h3a. The phylogenetic relationships between ancient mitogenomes and those from modern Ecuadorians and Peruvians are shown in supplementary figures S1–S5, Supplementary Material online. Only one ancient Peruvian B2 mitogenome (#242), dated at  $1,639 \pm 275$  years ago, turned out to be completely identical to a modern (Peruvian) mitogenome (#241).

After having classified modern and ancient mitogenomes from Ecuador and Peru into numerous subhaplogroups, we searched for their diagnostic mutational motifs (supplementary table S5, Supplementary Material online) in our in-house database that encompasses  $>1,700$  published mitogenomes of Native American origin. We identified 85 mitogenomes belonging or phylogenetically closely related to the subhaplogroups found in Ecuador and Peru (supplementary table S8, Supplementary Material online). Their phylogenetic and geographical evaluation revealed two classes of subhaplogroups: those restricted to South America and those with representatives also in the northern part of the double continent (supplementary table S5, Supplementary Material online).

It should be underscored that the distinguishing mutational motifs of the subhaplogroups in both classes are restricted to the Americas. This implies that they arose somewhere in the Americas during the (long) time frame that ranges from the initial human entry into the continent to very recent times. To estimate the minimum ages of these subhaplogroups, we calculated coalescence times with both Maximum Likelihood (ML) and BEAST (Bayesian Evolutionary Analysis Sampling Trees) computations (supplementary table S5, Supplementary Material online). As expected subhaplogroup ages varied widely, with some close to the postulated entry time of Paleo-Indians into North America and others which are extremely young ( $<1$  ka). The age estimates obtained with ML and BEAST were overall very similar and overlapping when considering standard errors, although BEAST ages tended to be older than those obtained with ML. Therefore, in order to be as conservative as possible in our dating, we employed ML ages in our evaluations. The subhaplogroups with ML ages that are equal to or older than 14 ka are listed in table 1; they include most (9 out of 11) of those also found in the northern part of the double continent and many subhaplogroups detected only in South America (supplementary table S5, Supplementary Material online).

### South American-Specific Subhaplogroups

Most subbranches within the pan-American haplogroups A2, B2, C1b, C1c, C1d, D1, and D4h3a shown in supplementary figures S1–S5, Supplementary Material online, turned out to encompass only mitogenomes from South America. Some of these subhaplogroups are apparently restricted to Ecuador (A2ac2, A2av1a, A2aw, B2b5a, B2b5b1a, B2b6a1a, B2b7, B2b8a, B2l1a, B2z, C1b23, C1b28, C1b29, and C1d1f), some to Peru (A2z2, A2as, A2at, A2au, B2b9a, B2b9b, B2b10, B2b11,

**Table 1.** Distribution of Subhaplogroups with ML Age Estimates Equal or Older Than 14 ka.

Subhaplogroups <sup>a</sup>	N <sup>b</sup>	ML Age Estimates ± SE (ka) <sup>c</sup>	Geographical Distribution
A2k <sup>d</sup>	20	15.32 ± 1.97	USA, Mexico, Venezuela, Wayuu (Colombia/Venezuela), Ecuador, Peru
<u>A2y</u> <sup>d</sup>	15	14.38 ± 1.64	Ecuador, Peru
A2z <sup>d</sup>	13	14.43 ± 2.00	USA, Puerto Rico, Peru
A2ar	5	15.20 ± 1.37	Guatemala, Ecuador, Peru
<u>A2as</u> <sup>e</sup>	6 (2)	15.88 ± 1.56	Peru
<u>A2at</u> <sup>e</sup>	10 (2)	15.89 ± 1.41	Peru
B2b	158 (11)	15.99 ± 0.92	USA, Puerto Rico, Mexico, Colombia, Bolivia, Brazil, Venezuela, Ecuador, Peru
>B2b2	4	14.76 ± 1.19	USA, Bolivia
>B2b3	8	14.77 ± 1.09	USA, Puerto Rico, Brazil, Venezuela
>B2b6 <sup>e</sup>	31	14.56 ± 1.11	Ecuador, Peru
>B2b11 <sup>e</sup>	15 (4)	13.99 ± 1.17	Peru
B2l <sup>e</sup>	10	15.64 ± 1.31	Mexico, Ecuador
>B2l1	9	14.23 ± 1.38	Mexico, Ecuador
<u>B2o1</u> <sup>d</sup>	5	14.19 ± 1.96	Bolivia, Ecuador
B2q	22 (1)	14.52 ± 1.86	USA, Mexico, Ecuador, Peru
B2aa <sup>e</sup>	15 (5)	14.60 ± 1.63	Mexico, Ecuador, Peru
<u>B2ab</u> <sup>e</sup>	15 (1)	14.28 ± 1.13	Bolivia, Peru
<u>B2ac</u> <sup>e</sup>	5	16.07 ± 1.27	Peru
<u>C1b21</u>	5 (2)	14.53 ± 1.71	Peru
<u>C1b24</u> <sup>e</sup>	4	16.27 ± 1.57	Peru
<u>C1b26</u> <sup>e</sup>	5	16.66 ± 1.21	Ecuador, Peru
>C1b26a <sup>e</sup>	4	15.55 ± 1.23	Ecuador, Peru
>>C1b26a1 <sup>e</sup>	3	14.47 ± 1.26	Peru
<u>C1b29</u> <sup>e</sup>	5	14.64 ± 2.38	Ecuador
<u>D4h3a11</u>	3	15.67 ± 1.81	Peru
D1f	27	18.05 ± 1.38	USA, Mexico, Brazil, Colombia, Venezuela, Ecuador, Peru
D1k	9	20.57 ± 2.76	USA, Mexico, Peru
>D1k1 <sup>e</sup>	5	15.81 ± 2.40	USA, Peru
<u>D1o</u> <sup>e</sup>	5 (1)	17.92 ± 1.63	Peru
<u>D1q</u> <sup>e</sup>	4	17.15 ± 1.52	Peru
<u>D1r</u>	6	17.86 ± 1.93	Peru
<u>D1t</u> <sup>e</sup>	3	18.81 ± 1.43	Peru

<sup>a</sup>The underlined subhaplogroups are restricted to South America.

<sup>b</sup>Number of all mitogenomes included in the analysis (both modern and ancient). The number of ancient mitogenomes is in brackets.

<sup>c</sup>ML age estimates are from the data set including both modern and ancient mitogenomes (supplementary table S5, Supplementary Material online).

<sup>d</sup>The sub-haplogroup nomenclature differs from that reported in PhyloTree (<http://www.phylotree.org>, last accessed July 10 2017).

<sup>e</sup>Sub-haplogroups defined for the first time in this study.

B2b12b, B2b13, B2aa1a, B2ab1a1, B2ac, B2ad, B2ae, B2y2, B2ag, B2ah, C1b16, C1b19, C1b21, C1b24, C1b25, C1b26a1, C1b27, C1d1e, D4h3a11, D1k1a, D1o, D1p, D1q, D1r, D1s, D1t, and D1u), others are detected in both countries (A2y, A2av1, B2b6, B2b8, B2b9, B2b12, B2q1, C1b26), and some harbor representatives also from elsewhere in South America (B2ab, B2b5, B2o1; supplementary table S8, Supplementary Material online). These geographical distributions indicate that their distinguishing mutational motifs most likely arose *in situ* (in South America) sometime during or after the process of human entry and spread into the southern subcontinent. It is interesting to note that, among the 647 mitogenomes from Ecuador and Peru, we did not detect any belonging to the subhaplogroups (B2i2, C1b13, D1g, and D1j) previously identified in the Southern Cone of South America (de Saint Pierre et al. 2012; Bodner et al. 2012). This observation not only confirms that B2i2, C1b13, D1g, and D1j indeed arose in the Southern Cone in the terminal phase of the migration process from North to South, but also reveals that there was very limited maternal lineage gene flow, if any, from South to the North along the Pacific in the following millennia.

Many of the subhaplogroups observed in Ecuador and Peru encompass only a rather small number of mitogenomes, thus their estimated ages should be considered with some caution. However, among the oldest (table 1) there are four, each represented by at least 15 mitogenomes, with virtually identical coalescence ages: A2y ( $14.4 \pm 1.6$  ka), B2b6 ( $14.6 \pm 1.1$  ka), B2b11 ( $14.0 \pm 1.2$  ka), and B2ab ( $14.3 \pm 1.1$  ka).

### Subhaplogroups Found Also in North and Central America

This second class is made up by eleven subhaplogroups (A2k, A2z, A2ac, A2ar, A2av, B2b, B2l, B2q, B2aa, D1f, and D1k) that are detected also in North or Central America (supplementary tables S5 and S8, Supplementary Material online). Three explanations can be envisioned for their geographical distributions. The first is that they arose in North or Central America, prior to the arrival of humans in South America, and reached the southern continent with the first colonizers of the subcontinent. The second is that they arose in North or Central America but their presence in South America is accounted for by later migratory events. Alternatively, they

might have arisen in South America after human arrival in the subcontinent and have later spread towards the North.

Most of these subhaplogroups encompass only a limited number of mitogenomes, but they are characterized by some informative phylogenetic features that help to discriminate between the three possibilities listed above. For instance, B2q, B2aa, B2l, and D1k show subbranches (or haplotypes) that depart directly from the subhaplogroup root and are detected either only in Mexico (and the US) or only in Ecuador and/or Peru (fig. 1), a feature in agreement with the scenario that they arose in North America. Moreover, their coalescence times are all  $>14$  ka (table 1), even though the age estimate for subhaplogroup D1k ( $20.6 \pm 2.8$  ka) is probably affected by the excess of reversions in the published D1k1 mitogenomes (fig. 1, panel D). This indicates that they most likely reached South America with the first entry in the subcontinent. A similar branching pattern is also observed for subhaplogroup A2ar ( $15.2 \pm 1.4$  ka), with the difference here that the northern mitogenome departing from the root is not from Mexico or the US but from Guatemala (fig. 2, panel A).

Not all members of this class of subhaplogroups harbor only a few representatives, among these there is also B2b that, as mentioned above, is extremely frequent both in Ecuador and Peru. A phylogeny of B2b has been previously proposed (Taboada-Echalar et al. 2013). With the addition to the phylogeny of 134 novel B2b modern mitogenomes, we were able to identify nine new internal subclades (B2b5–B2b13), some already mentioned above, and obtained a more accurate estimate of its age. Among the 147 B2b mitogenomes included in figure 3, nine are from North and Central America: one from a Pomo of North California (EU095208), one from Mexico (HQ012137), three are from Hispanic subjects living in the US (KM102108, KM102111, KM102138) and four from Puerto Rico (HG00640, HG01079, HG01191, HG01198; supplementary table S8, Supplementary Material online). The remaining B2b mitogenomes ( $N = 138$ ) are from South America, but unlike the subhaplogroups mentioned above, they were found not only in the countries along the Pacific coast, but also in the Atlantic regions of the subcontinent (fig. 3).

Many of the mitogenomes from the US belonging to B2b and other subhaplogroups (supplementary table S8, Supplementary Material online) are from individuals classified as “Hispanics,” a rather generic term that refers to individuals of Cuban, Mexican, Puerto Rican, South or Central American origin. Thus, their ancestral ethnogeographic source remains essentially undefined unless genealogical information is available or an accurate evaluation of the phylogenetic relationships between members of the subhaplogroup is performed. Figure 3 shows that two of the mitogenomes from the US, including the one from the Pomo (North California), and the one from Mexico depart directly from the root of B2b, in agreement with the scenario that B2b originated in North America as well. Moreover, its age estimate ( $16.0 \pm 0.9$  ka) indicates an early occurrence after the human entry from Beringia. Another one of the US mitogenome clusters within the subbranch B2b2 together with three mitogenomes from Bolivia. It should be noted that B2b2 (fig. 3) is one of the

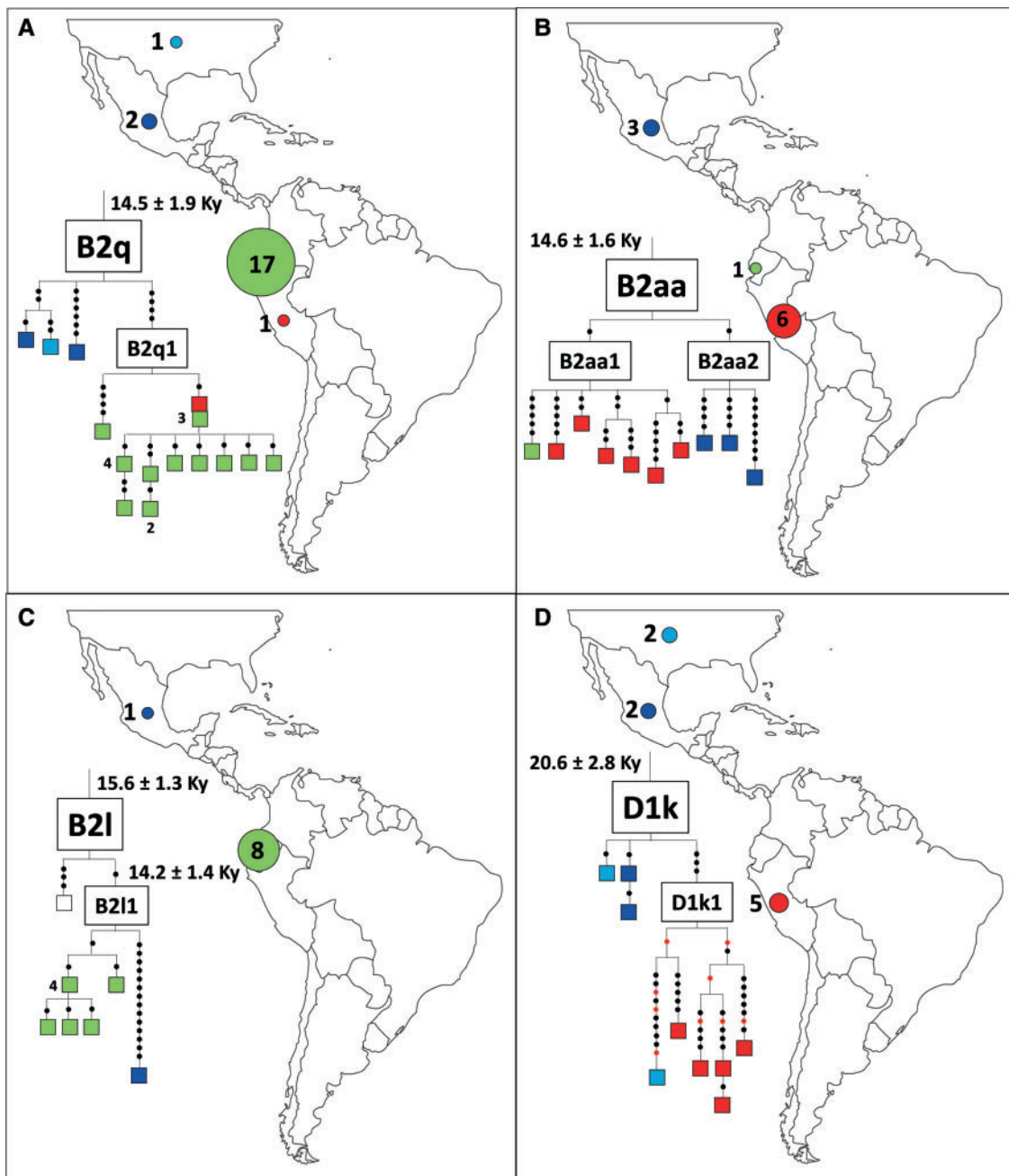
oldest subbranches ( $14.8 \pm 1.2$  ka) of B2b and that also in this case the mitogenome from the US diverges directly from the ancestral root. This suggests that not only B2b but also B2b2 might have arisen in North or Central America. The phylogenetic connections of the remaining B2b mitogenome from the US (KM102108), a member of B2b3 (fig. 3, supplementary table S8, Supplementary Material online), are instead clearly indicative of Puerto Rico as an ancestral source. Indeed it is a direct derivative of a haplotype found in two subjects from Puerto Rico, in turn deriving from another one found in both Puerto Rico and Venezuela that is distantly related to mitogenomes from Brazil (Kayapo and Yanomama). These phylogenetic links become clearer when considering the strong Taino mtDNA component in modern Puerto Ricans and that the Tainos were the final outcome of migratory events from the Orinoco river basin, which eventually reached the Greater Antilles  $\sim 5$  ka (Martínez-Cruzado et al. 2001). Thus, B2b3 most likely represents an additional ancient branch ( $14.8 \pm 1.1$  ka) that arose in South America at a very early stage of human spread into the southern subcontinent, a branch now found in the Greater Antilles because of ancient migrations, and in the US because of recent gene flow from Puerto Rico. It should also be mentioned that Martínez-Cruzado (2010) has raised the possibility, based also on the apparent lack (at least so far) of B2 mitogenomes in ancient DNA studies in the Caribbean, that most B2 mitogenomes in Puerto Ricans might have arrived during the postcontact era, a scenario that could also apply to B2b3. While, we cannot rule out this scenario at the moment, we can say that even if B2b3 arrived in Puerto Rico after the Taino period, its ancestral source remains unchanged (the Orinoco river basin).

Phylogeographic data suggest a similar Puerto Rican origin also for the US mitogenomes of subhaplogroup A2z (fig. 2, panel C), which is split into A2z1 and A2z2, with the former branch already identified in Puerto Rico and Cuba by assessing mtDNA control-region variation (Vilar et al. 2014). The phylogeny shows that all six US mitogenomes are members of A2z1, a branch that they share with three Puerto Rican mitogenomes and one previously published from Peru. Taking into account that three of the US mitogenomes are identical to two from Puerto Rico, it is likely that the six US members of A2z1 are indeed all from Puerto Rico, and that A2z ( $14.4 \pm 2.0$  ka), similarly to B2b3, also arose in South America.

Finally, there are two additional subhaplogroups, D1f and A2k, with phylogeographic features that are suggestive of a North American origin. Despite their much lower frequency, they both show geographic distributions very similar to that of B2b, with mitogenomes not only from North America and northwestern South America (Ecuador and Peru), but also from South American countries of the Atlantic area (fig. 2).

## Discussion

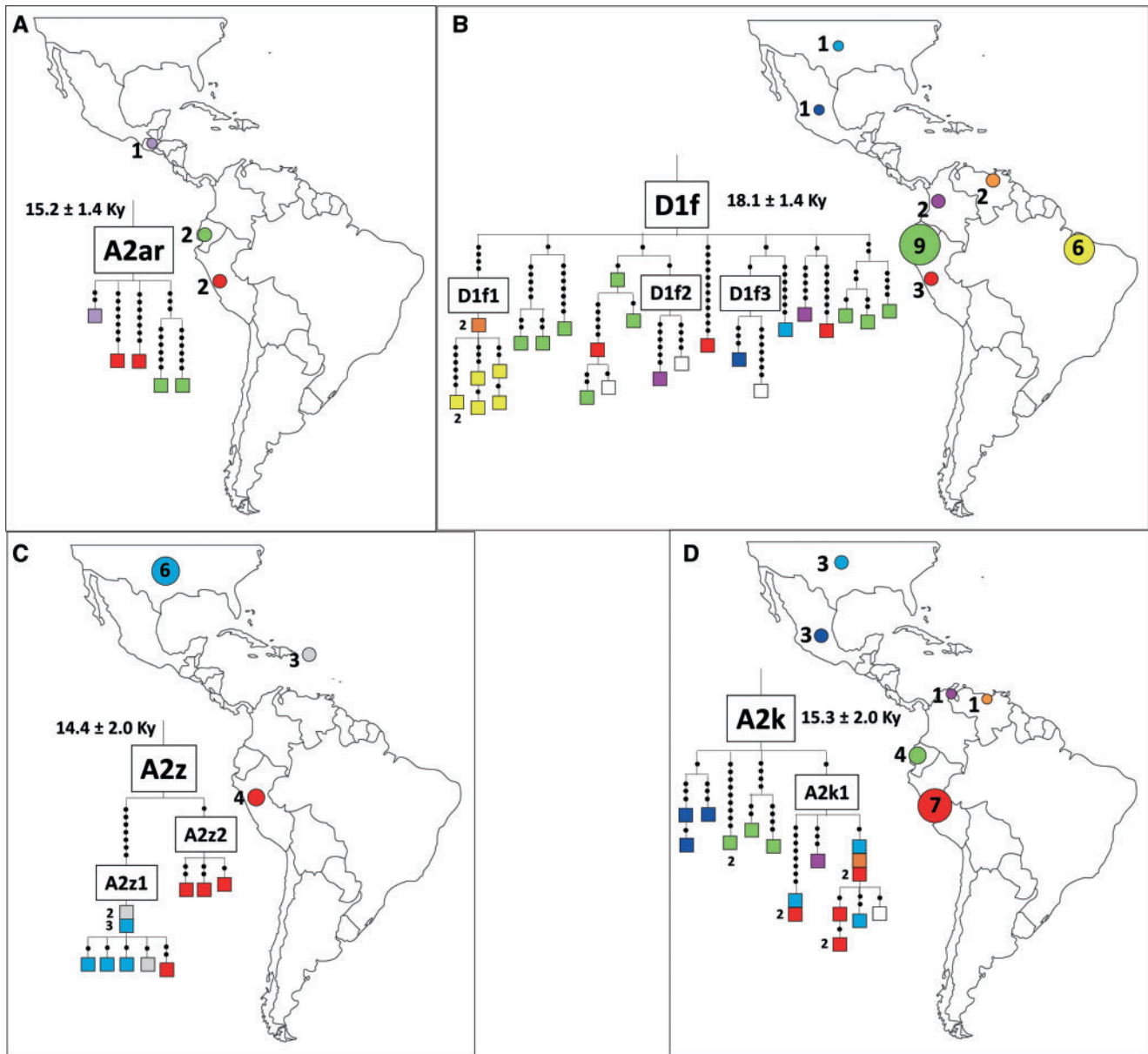
To date and in contrast with the Southern Cone (Bodner et al. 2012; de Saint Pierre et al. 2012), the nature and extent of mtDNA variation in northwestern South America have only been marginally evaluated at the level of entire mitogenomes.



**Fig. 1.** Phylogenetic relationships and geographical distributions of B2q (A), B2aa (B), B2l (C), and D1k (D) modern mitogenomes. Each color corresponds to a country, thus colors of mitogenomes (squares) in the schematic trees correspond to their (maternal) geographical origin, except for white that indicates an unknown source. In the tree, each square corresponds to one mitogenome unless otherwise indicated by the number close or below the square. Each black dot on a branch indicates a mutation. The red dots in the subbranch D1k1 (panel D) are all reversions. Their reliability is dubious and they might be the cause of the high age estimate of D1k. In the map (drawn by hand), the number of mitogenomes for each haplogroup is shown per country and the sizes of circles are proportional to the numbers of mitogenomes.

In this study, we analyzed 647 mitogenomes from Ecuador and Peru, detecting all pan-American haplogroups (A2, B2, C1b, C1c, C1d, D1, and D4h3a) that, according to previous studies (Tamm et al. 2007; Perego et al. 2009; Achilli et al. 2013), entered into North America from Beringia along the Pacific coast. This confirms that the founding haplotypes of these haplogroups were also involved in the human entry into South America. As expected, we did not detect representatives of the other founding Native American haplogroups for which either a different entry route (ice free

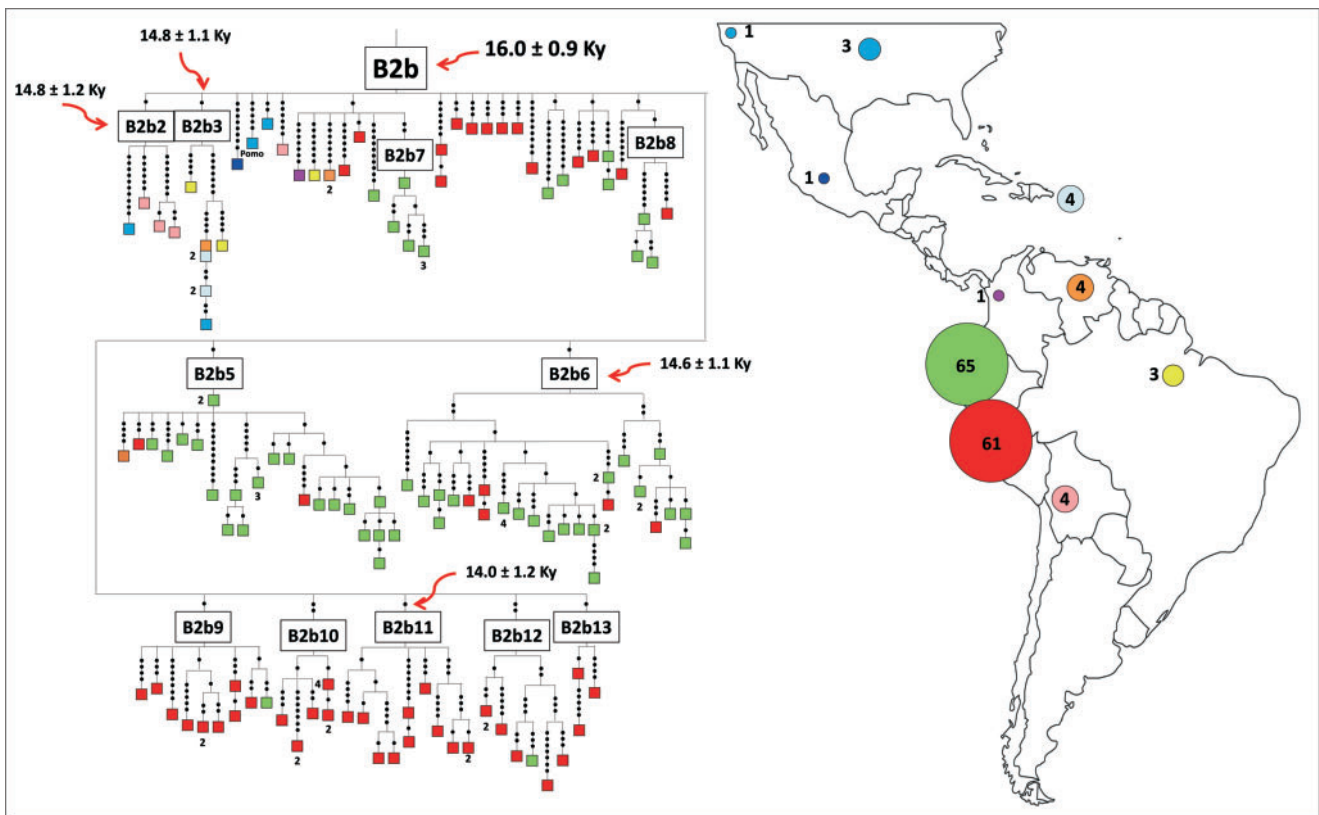
corridor) (X2a, X2g, and C4c) or a later arrival from Beringia/Alaska (A2a, A2b, D2a, and D3) have been proposed (Perego et al. 2009; Hooshiar Kashani et al. 2012; Achilli et al. 2013). Note that we also did not detect any mitogenomes of either Oceanian or South East Asian origin, a scenario that has been recently re-opened by two genome-wide studies that apparently identified signals of shared ancestry with Australo-Melanesians in some Amazonian populations (Surui, Karitiana, Xavante) of Brazil (Raghavan et al. 2015; Skoglund et al. 2015).



**Fig. 2.** Phylogenetic relationships and geographical distributions of A2ar (A), D1f (B), A2z (C), and A2k (D) modern mitogenomes. Each color corresponds to a country, thus colors of mitogenomes (squares) in the schematic trees correspond to their (maternal) geographical origin, except for white that indicates an unknown source. In the tree, each square corresponds to one mitogenome unless otherwise indicated by the number close or below the square. Each black dot on a branch indicates a mutation. In the map, the number of mitogenomes for each haplogroup is shown per country and the sizes of circles are proportional to the numbers of mitogenomes.

Within each of the pan-American haplogroups, we observed extensive differentiation into derived branches, including 48 newly defined subhaplogroups (supplementary figs. S1–S5, Supplementary Material online). When we searched for the diagnostic mutational motifs of these derived branches in a database of  $\sim 1,700$  previously published mitogenomes of Native American origin, we identified 85 mitogenomes belonging or closely related to these subhaplogroups (supplementary table S8, Supplementary Material online). An overall phylogeographic assessment revealed two classes of subhaplogroups: those restricted to South America and those with representatives also in the northern part of the double continent (supplementary table S5, Supplementary Material online).

It should be mentioned here that the geographical distributions of subhaplogroups that we observed in this study might be in some cases inaccurate for at least two reasons. The first is the number of published mitogenomes ( $\sim 1,700$ ) available for comparisons that is not large enough to provide a good representation of the presence/absence of each subhaplogroup in the different regions of the double continent. The second is that many Native American populations experienced a dramatic size reduction at the time of European contact, with genetic drift playing a major role in shaping the remaining genetic variation. Thus geographical distributions of subhaplogroups prior to the arrival of Europeans might be very different from the current ones. We cannot exclude, for instance, that some of the subhaplogroups that we now



**Fig. 3.** Phylogenetic relationships and geographical distributions of B2b modern mitogenomes. Each color corresponds to a country, thus colors of mitogenomes (squares) in the schematic tree correspond to their (maternal) geographical origin. In the tree, each square corresponds to one mitogenome unless otherwise indicated by the number close or below the square. Each black dot on a branch indicates a mutation. In the map, the number of mitogenomes is shown per country and the sizes of circles are proportional (except for Ecuador and Peru) to the numbers of mitogenomes.

consider as restricted to South America might have indeed unsampled representatives in living individuals for North and/or Central America, or that they were present in North or Central America at the time of European arrival and then they went extinct.

Taking into account the limitations mentioned above, the simplest explanation, at least for the moment, for subhaplogroups with a geographical distribution restricted to South America is that they arose in South America. If so, the oldest of these subhaplogroups would have been the first to arise and their coalescence ages would provide a lower boundary for human presence in the southern subcontinent. Interestingly among these, the four most represented (A2y, B2b6, B2b11, B2ab; each  $\geq 15$  mitogenomes) harbor virtually identical ages (table 1), indicating that humans were already in South America by 14.0–14.6 ka. This time frame is further supported by the ages of A2z ( $14.4 \pm 2.0$  ka) and B2b3 ( $14.8 \pm 1.1$  ka), which most likely also arose in South America at a very early stage of human spread into the southern subcontinent and are now also found in the Greater Antilles because of ancient migrations and in the US because of recent gene flow from Puerto Rico.

The subhaplogroups found also in the northern part of the double continent are not as numerous as those restricted to South America, but almost all harbor old coalescence ages (figs. 1–3), as would be expected in the scenario that they

arose in North or Central America prior to the spread of Paleo-Indians into South America. These include B2q, B2aa, B2l, D1k, A2ar, D1f, and A2k (figs. 1 and 2) and B2b (fig. 3). These most likely arrived together with the founder haplotypes of haplogroups A2, B2, C1b, C1c, C1d, D1, and D4h3a, thus indicating that overall at least 15 different founding mtDNA haplotypes arrived to South America at the time of human entry in the subcontinent and could have exploited the selective advantage of being part of an expanding wave front (Moreau et al. 2011). Among these, B2b and its derivatives are extremely informative. They reveal that the early Paleo-Indian carriers of B2b probably moved from North America to the area corresponding to modern Ecuador and Peru over the short time frame of  $\sim 1.5$  ka comprised between  $16.0 \pm 0.9$  ka and  $14.6 \pm 1.1$  ka, corresponding to the ages of B2b and its oldest South American branch in the northwestern part of the subcontinent, respectively—a finding that fits with archaeological evidence attesting to human presence at the Monte Verde site in southern Chile at least 14.5 ka (Dillehay et al. 2015) and the conclusions of some earlier mitogenome studies (Bodner et al. 2012; de Saint Pierre et al. 2012).

An entry time into South America between these two time boundaries is also supported by the Bayesian Skyline Plot (BSP) analysis shown in figure 4, which includes not only the mitogenomes from Ecuador and Peru but also all modern



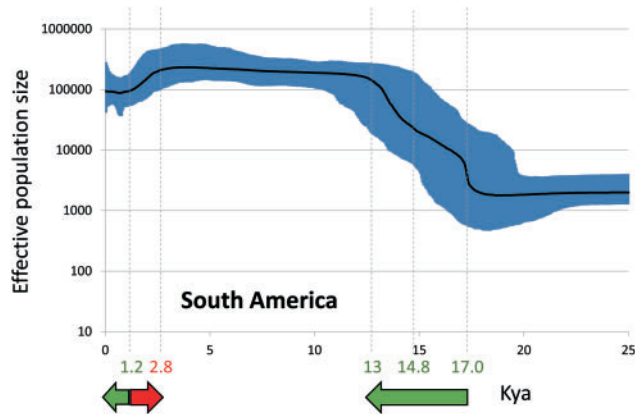
Native American mitogenomes from South America that are currently available in the literature—overall 1,053 subjects. The BSP shows a rather sharp increase in population size beginning  $\sim 17.0$  ka and a second demographic growth at  $\sim 14.8$ – $15$  ka, followed first by a very long period of overall demographic stability, from  $\sim 13$  to  $\sim 2.8$  ka, and then by a period of  $\sim 1,500$  years of demographic contraction. The starting growth probably reflects the initial entry and spread of the first Americans into the northern part of the double continent whereas the following one at  $\sim 14.8$ – $15$  ka is possibly associated with the later arrival and further spread into South America. We could not link the decline beginning  $\sim 2.8$  ka to any known climatic or cultural event, but a similar decline has been previously reported by employing mitogenomes from the Americas (O’Fallon and Fehren-Schmitz 2011) as well as data from South American archaeological sites (Goldberg et al. 2016).

Finally, some of the subhaplogroups that arose in North or Central America and later spread into South America also provide valuable information concerning the routes of diffusion of the first South Americans. In particular, the geographical distributions of the A2k, B2b and D1f mitogenomes (figs. 2 and 3) indicate that the first settler population(s) might have undergone an early split in the northern part of South America (Wang et al. 2007), followed by diffusion along both the Pacific and Atlantic coastal regions (fig. 5). It should be underlined that such a scenario is also compatible with an early finding that until now has not been fully evaluated. One of the previously published mitogenomes (GenBank FJ68754) belonging to D4h3a, a haplogroup so far considered a marker of the Paleo-Indian spread along the Pacific coast, has been found in the northeastern part of Brazil (Maranhão state; Perego et al. 2009). We now know that this mitogenome does not belong to any of the several D4h3a subbranches that characterize the western part of South America (Lindo et al. 2017); instead, it departs directly from the haplogroup root. Therefore a postpeopling event of gene flow from the western part of the subcontinent is a rather unlikely explanation for its presence in northeastern Brazil. In contrast, its detection there is fully compatible with the scenario that D4h3a is an additional haplogroup that, similar to A2k, B2b, and D1f, was present in both population subsets that moved along the Pacific and Atlantic coasts after the initial split in the northern part of South America.

## Materials and Methods

### Sample

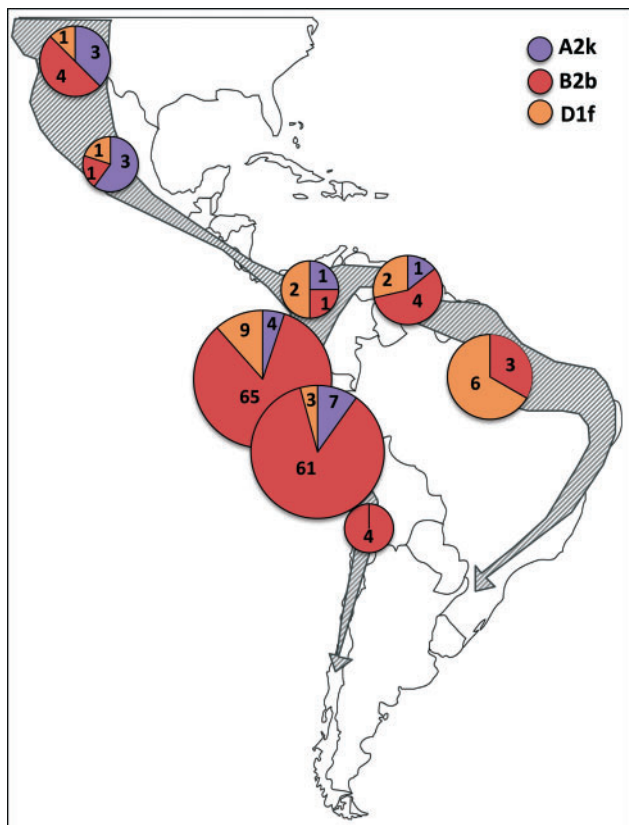
A total of 217 Ecuadorians (93 Native Americans and 124 Mestizos), representatives of all major regions of Ecuador (supplementary table S2, Supplementary Material online), and ten Peruvians (all Mestizos; supplementary table S3, Supplementary Material online) were enrolled in the study. Ethnicity and genealogical information were ascertained by direct interview. For all individuals an appropriate written informed consent was obtained, with protocols approved by the Ethic Committee for Clinical Experimentation of the University of Pavia, Board minutes of the April 11, 2013.



**Fig. 4.** Bayesian skyline plot (BSP) analysis of South American mitogenomes. This analysis included not only the mitogenomes from Ecuador and Peru, but also all modern Native American mitogenomes from South America currently available in the literature for a total of 1,053 subjects. Mitogenomes of Old World ancestry were excluded. The thick solid line is the median estimate and the blue shading shows the 95% highest posterior density limits. The red arrow indicates the population size decrease between 2.8 and 1.2 ka whereas the green one indicates the population size growth between  $\sim 17$  and  $\sim 13$  ka. The mitogenomes included in the BSP analysis are from the following sources: 129 from Colombia (Tamm et al. 2007; Hartmann et al. 2008, Direct Submission; Perego et al. 2010; Greenspan 2011, Direct Submission; Behar et al. 2012; Lippold et al. 2014; Rieux et al. 2014; Zheng et al. 2014, Direct Submission), 225 from Ecuador (this study; Tamm et al. 2007; Perego et al. 2010; Cardoso et al. 2012; Greenspan 2015, Direct Submission), 354 from Peru (this study; Zheng et al. 2014; Perego et al. 2009, 2010; Tito et al. 2012, Direct Submission; Fehren-Schmitz et al. 2015), 15 from Bolivia (Fagundes et al. 2008; Perego et al. 2009; Taboada-Echalar et al. 2013), 85 from Chile (Perego et al. 2009, 2010; Bodner et al. 2012; Behar et al. 2012; Perego et al. 2012; Greenspan 2012, Direct Submission; de Saint Pierre et al. 2012; Rieux et al. 2014), 51 from Argentina (Perego et al. 2010; Bodner et al. 2012; de Saint Pierre et al. 2012; Greenspan 2013, Direct Submission), 36 from Venezuela (this study; Ingman et al. 2000; Gómez-Carballa et al. 2012; Lee and Merriwether 2015), 146 from Brazil (Ingman et al. 2000; Fagundes et al. 2008; Perego et al. 2009, 2010; Bodner et al. 2012; Behar et al. 2012; Greenspan 2012, Direct Submission; Lippold et al. 2014; Zheng et al. 2014; Rieux et al. 2014), 5 from Paraguay (Fagundes et al. 2008; Perego et al. 2010; Rieux et al. 2014), and 7 from Uruguay (Perego et al. 2010; Sans et al. 2012, 2015).

Genomic DNA was extracted and purified from either buccal swabs (202 Ecuadorians), or mouthwash (one Peruvian and four Ecuadorians), or cord blood (nine Peruvians and 11 Ecuadorians; supplementary table S4, Supplementary Material online) following standard phenol/chloroform methods.

MtDNAs of Native American ancestry were identified and selected through a preliminary survey of the mtDNA control region from np 16024 to np 300 following a standard Sanger protocol (Karachanak et al. 2012). The identified mutational motifs, relative to the revised Cambridge Reference Sequence (rCRS; Andrews et al. 1999), allowed the classification of mtDNAs into Native American and Old World haplogroups (data not shown). The 217 subjects (208 from Ecuador and nine from Peru) harboring diagnostic mutational motifs of



**Fig. 5.** Diffusion routes of first South Americans as suggested by the geographical distributions of A2k, B2b, and D1f mitogenomes. The number of mitogenomes belonging to haplogroups A2k, B2b, and D1f is reported per country and the sizes of circles are proportional (except for B2b in Ecuador and Peru) to the numbers of mitogenomes.

Native American haplogroups underwent sequencing of the entire mitogenome.

### Sequencing of Entire Mitogenomes

The entire sequences of the 217 Native American mtDNAs were obtained by using either Next Generation Sequencing (NGS) with an Illumina MiSeq (a total of 150 mitogenomes: 141 from Ecuador and 9 from Peru) or Sanger sequencing (the remaining 67 mitogenomes from Ecuador).

For NGS, two overlapping long range PCR fragments covering the whole mtDNA sequence were first amplified with primer pairs 5871 for (5'-GCTTCACTCAGCCATTTTACCT-3') and 13829rev (5'-AGTCCTAGGAAAGTGACAGCGA-3') for the first fragment (7,959 bp), and 13477 for (5'-GCAGGAATACCTTTTCTCACAG-3') and 6151rev (5'-ACTAGTCAGTTGCCAAAGCCTC-3') for the second one (9,244 bp). The amplification was performed with 10–50 ng of template DNA in 50  $\mu$ l of reaction mix containing 1 $\times$  GoTaq LongPCR Master Mix (Promega) and 0.2  $\mu$ M of each primer, according to manufacturer's instructions. The PCR program included an initial denaturation step at 94  $^{\circ}$ C for 2 min, and 30 cycles with the following thermal profile: 94  $^{\circ}$ C for 30 s, 55  $^{\circ}$ C for 30 s, 65  $^{\circ}$ C for 9 min, with a final extension step at 72  $^{\circ}$ C for 10 min. The two PCR products were purified with Wizard SV Gel and PCR Clean-Up System (Promega)

according to manufacturer's protocol and quantified with a Quantus Fluorometer (Promega).

About 1.5 ng of PCR product (0.75 ng for each PCR) was used for the set-up of a sequencing library with the Nextera XT DNA sample preparation kit (Illumina) following the manufacturer's protocol. Sequencing reactions were carried out on a MiSeq System (Illumina) by using the MiSeq Reagent Nano Kit, v2 (300 cycles). On-board software created results in FASTQ format, which were analyzed with the Geneious software (version 8.1). This software was used to compare mitogenome sequences with both the Revised Sapiens Reference Sequence (RSRS; Behar et al. 2012) and the rCRS (Andrews et al. 1999) and to create a report of sequence variants (nucleotide substitutions and indels). The threshold used to detect heteroplasmies was 20% of mutated bases and the average depth of the obtained reads was  $\sim$ 4,000 $\times$ .

The 67 mitogenomes analyzed with the Sanger approach were completely sequenced following a well-established protocol (Torroni et al. 2001). We aligned, assembled and compared their sequences using Sequencher 5.0 (Gene Codes Corporation), also in this case, relative to both RSRS and rCRS.

### Phylogenetic Analyses and Haplogroup Age Estimates

In addition to the 217 novel Native American mitogenomes from Ecuador and Peru obtained in this study, 362 previously published modern mitogenomes (17 from Ecuador and 345 from Peru; Tamm et al. 2007; Perego et al. 2009, 2010; Tito et al. 2012, direct submission; Cardoso et al. 2012; Zheng et al. 2014; Greenspan 2015, direct submission) as well as 68 ancient ones from Peru dated to the Early/Middle Holocene (Fehren-Schmitz et al. 2015) and preColumbian times (Gómez-Carballa et al. 2015; Llamas et al. 2016), were included in the phylogenetic analyses (supplementary table S4, Supplementary Material online).

We built maximum-parsimony (MP) trees, one each for macrohaplogroups A2, B2 (without B2b), C1 (including C1b, C1c, and C1d), and D4 (including D4h3a and D1), and one for haplogroup B2b (supplementary figs. S1–S5, Supplementary Material online). These trees encompassed both the new and previously published Ecuadorian (225, all modern) and Peruvian (68 ancient and 354 modern) mitogenomes. The MP phylogenetic trees were obtained by using the mtPhyl software (<https://sites.google.com/site/mtphyl/home>) and hand-corrected with reference to PhyloTree (van Oven and Kayser 2009).

New haplogroups/subhaplogroups (in blue in supplementary figures S1–S5, Supplementary Material online and listed in supplementary table S5, Supplementary Material online) were defined when encompassing a minimum of three different haplotypes sharing at least one stable mutation (not recurrent in the tree), and were named following the nomenclature of the PhyloTree database build 17 (at <http://www.phylotree.org/>; van Oven and Kayser 2009). In some cases, the presence of new mitogenomes branching prior to a previously defined haplogroup node (i.e. A2k, A2y, A2z, and B2o1) forced us to redefine the nomenclature of the branches as

well as their diagnostic mutational motifs (supplementary table S5, Supplementary Material online).

Haplogroup coalescence ages were obtained using the 647 mitogenomes from Ecuador and Peru (supplementary table S4, Supplementary Material online) together with additional 85 listed in supplementary table S8, Supplementary Material online. Among these, 79 mitogenomes were previously published, whereas six mitogenomes from Venezuela (GenBank accession numbers: KY681011–KY681016) were extracted from cord blood and sequenced by NGS in this study. Also for these additional samples, written informed consent was obtained, and the protocols were approved by the Ethic Committee for Clinical Experimentation of the University of Pavia, Board minutes of the April 11, 2013. Coalescence times were estimated using two methods: ML and BEAST. We performed these calculations considering all substitutions except those at nps 16182, 16183, and 16519.

ML estimations were performed using the software PAMLX 1.3.1 (Yang 1997), assuming the HKY85 mutation model (two parameters in the model of DNA evolution) with gamma-distributed rates (approximated by a discrete distribution with 32 categories). They were performed on two data sets, one including only modern mitogenomes and the other with both modern and ancient mitogenomes. The estimated ages of macrohaplogroups M and N reported in Behar et al. (2012) were used as fixed priors for both data sets, whereas the dates of the ancient samples reported in supplementary table S1, Supplementary Material online of Llamas et al. (2016) were used as tip calibration points in the data set including ancient samples (supplementary tables S4 and S5, Supplementary Material online).

To calculate Bayesian age estimates, we employed BEAST 1.8.3 (Drummond and Rambaut 2007) on both data sets used for the ML analyses. Also in this case, the estimated ages of macrohaplogroups M and N were used as fixed priors for both data sets, whereas the dates of the ancient samples reported in supplementary table S1, Supplementary Material online of Llamas et al. (2016) were used as tip calibration points in the data set including ancient samples. The program was run under the HKY substitution model (gamma-distributed rates plus invariant sites) with a fixed molecular clock (Olivieri et al. 2017). Taking into account that the clock rate is linear in BEAST and that the timeframe for the appearance of the Native American subhaplogroups is <20 ka, the corrected molecular clock (Soares et al. 2009) was set at  $2.33 \pm 0.2 \times 10^{-8}$  base substitution per nucleotide per year over the entire mitogenome, which corresponds to ~2,650 years for a mutation to happen. The chain length was established at 50,000,000 iterations, with samples drawn every 10,000 Markov chain Monte Carlo (MCMC) steps, after a discarded burn-in of 5,000,000 steps, as in previous studies (Olivieri et al. 2017).

### Mitogenome Diversity

Haplotype (H) and nucleotide ( $\pi$ ) diversities and average number of nucleotide differences (M) in Ecuador and Peru were calculated using the software package DnaSP 5 (Librado and Rozas 2009). Only mitogenomes from random surveys of

modern populations were considered (supplementary table S4, Supplementary Material online). This led to the exclusion of 13 A2 (Tamm et al. 2007; Cardoso et al. 2012), one B2 (Tamm et al. 2007), one C1b (Greenspan 2015, direct submission), nine C1d (Perego et al. 2010), and 12 D4h3a (Perego et al. 2009) mitogenomes from the data set including all available Ecuadorian and Peruvian mitogenomes.

### BSP Analysis

BSP analysis was performed with the BEAST software using a strict molecular clock (lognormal distribution across branches and uncorrelated between them) and an HKY85-type model with  $\gamma$ -distributed rates. The ages of haplogroups N and M (Behar et al. 2012) were considered as consistent internal calibration points. The chain length was established at 50,000,000 iterations, with samples drawn every 10,000 MCMC steps, after a discarded burn-in of 5,000,000 steps. BSPs were visualized in a plot with Tracer v1.5 and then converted into an Excel graph by assuming a generation time of 25 years.

### Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

### Acknowledgments

We are grateful to all the volunteers who generously participated in this study and made this research possible. A special thanks to Ugo A. Perego and Ann Turner for providing us the Native American mitogenome database and to Ilaria Sbarsi, Anna Luisa Cremaschi, Marco Guarene, Rosalia Cacciatori, and Annamaria Pasi for their contributions to the study. This study received support from the University of Pavia strategic theme “Towards a governance model for international migration: An interdisciplinary and diachronic perspective” (MIGRAT-IN-G) (to A.O., A.A., O.S., and A.T.), and the Italian Ministry of Education, University and Research: Progetti Futuro in Ricerca 2012 (RBFR126B8I) (to A.O. and A.A.) and Progetti Ricerca Interesse Nazionale 2012 (to A.A., O.S., and A.T.). The sequence data for the mitogenomes are available in GenBank with the accession numbers KY680794–KY681016.

### References

- Achilli A, Perego UA, Lancioni H, Olivieri A, Gandini F, Hooshiar Kashani B, Battaglia V, Grugni V, Angerhofer N, Rogers MP. 2013. Reconciling migration models to the Americas with the variation of North American native mitogenomes. *Proc Natl Acad Sci U S A*. 110(35):14308–14313.
- Aimola G, Andrade C, Mota L, Parenti F. 2014. Final Pleistocene and Early Holocene at Sitio do Meio, Piauí-Brazil: stratigraphy and comparison with Pedra Furada. *J Lithic Stud*. 1(2):5–24.
- Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet*. 23(2):147.
- Behar DM, van Oven M, Rosset S, Metspalu M, Loogväli EL, Silva NM, Kivisild T, Torroni A, Villems R. 2012. “Copernican” reassessment of the human mitochondrial DNA tree from its root. *Am J Hum Genet*. 90(4):675–684.

- Boëda E, Clemente-Conte I, Fontugne M, Lahaye C, Pino M, Daltrini Felice G, Guidon N, Hoeltz S, Lourdeau A, Pagli M, et al. 2014. A new late Pleistocene archaeological sequence in South America: the Vale da Pedra Furada (Piauí, Brazil). *Antiquity* 88(341):927–955.
- Bodner M, Perego UA, Huber G, Fendt L, Röck AW, Zimmermann B, Olivieri A, Gómez-Carballa A, Lancioni H, Angerhofer N, et al. 2012. Rapid coastal spread of first Americans: novel insights from South America's Southern Cone mitochondrial genomes. *Genome Res.* 22(5):811–820.
- Cardoso S, Alfonso-Sánchez MA, Valverde L, Sánchez D, Zarrabeitia MT, Odriozola A, Martínez-Jarreta B, de Pancorbo MM. 2012. Genetic uniqueness of the Waorani tribe from the Ecuadorian Amazon. *Heredity* 108(6):609–615.
- de Saint Pierre M, Gandini F, Perego UA, Bodner M, Gómez-Carballa A, Corach D, Angerhofer N, Woodward SR, Semino O, Salas A, et al. 2012. Arrival of Paleo-Indians to the Southern Cone of South America: new clues from mitogenomes. *PLoS ONE* 7(12):e51311.
- Dillehay TD. 1989. Monte Verde: a late Pleistocene settlement in Chile. Volume I: The paleo-environment and site context. Washington (DC): Smithsonian Institution Press.
- Dillehay TD, Collins MB. 1988. Early cultural evidence from Monte Verde in Chile. *Nature* 332(6160):150–152.
- Dillehay TD, Ocampo C, Saavedra J, Sawakuchi AO, Vega RM, Pino M, Collins MB, Scott Cummings L, Arregui I, Villagran XS, et al. 2015. New archaeological evidence for an early human presence at Monte Verde, Chile. *PLoS One* 10(11):e0141923.
- Dillehay TD, Goodbred S, Pino M, Vásquez Sánchez VF, Tham TR, Adovasio J, Collins MB, Netherly PJ, Hastorf CA, Chiou KL, et al. 2017. Simple technologies and diverse food strategies of the Late Pleistocene and Early Holocene at Huaca Prieta, Coastal Peru. *Sci Adv.* 3(5):e1602778.
- Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol.* 7:214.
- Fagundes NJ, Kanitz R, Eckert R, Valls AC, Bogo MR, Salzano FM, Smith DG, Silva WA Jr, Zago MARibeiro-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(3):583–592.
- Fehren-Schmitz L, Llamas B, Lindauer S, Tomasto-Cagigao E, Kuzminsky S, Rohland N, Santos FR, Kaulicke P, Valverde G, Richards SM, et al. 2015. A re-appraisal of the early Andean human remains from Lauricocha in Peru. *PLoS ONE* 10(6):e0127141.
- Forster P, Harding R, Torroni A, Bandelt HJ. 1996. Origin and evolution of Native American mtDNA variation: a reappraisal. *Am J Hum Genet.* 59(4):935–945.
- Fraser B. 2014. The first South Americans: extreme living. *Nature* 514(7520):24–26.
- Gilbert MT, Jenkins DL, Götherstrom A, Naveran N, Sanchez JJ, Hofreiter M, Thomsen PF, Binladen J, Higham TF, Yohe RM 2nd, et al. 2008a. DNA from pre-Clovis human coprolites in Oregon, North America. *Science* 320:786–789.
- Gilbert MT, Kivisild T, Grønnow B, Andersen PK, Metspalu E, Reidla M, Tamm E, Axelsson E, Götherström A, Campos PF, et al. 2008b. Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. *Science* 320:1787–1789.
- Greenberg JH, Turner CG II, Zegura SL, Campbell L, Fox JA, Laughlin WS, Szathmari EJE, Weiss KM, Woolford E. 1986. The settlement of the Americas: a comparison of the linguistic, dental, and genetic evidence [and comments and reply]. *Curr Anthropol.* 27(5):477–497.
- Greenspan B. 2011. Direct submission. Family Tree DNA.
- Greenspan B. 2012. Direct submission. Family Tree DNA.
- Greenspan B. 2013. Direct submission. Family Tree DNA.
- Greenspan B. 2015. Direct submission. Family Tree DNA.
- Goldberg A, Mychajliw AM, Hadly EA. 2016. Post-invasion demography of prehistoric humans in South America. *Nature* 532(7598):232–235.
- Gómez-Carballa A, Ignacio-Veiga A, Álvarez-Iglesias V, Pastoriza-Mourelle A, Ruiz Y, Pineda L, Carracedo A, Salas A. 2012. A melting pot of multicontinental mtDNA lineages in admixed Venezuelans. *Am J Phys Anthropol.* 147(1):78–87. Direct submission.
- Gómez-Carballa A, Catelli L, Pardo-Seco J, Martín-Torres F, Roewer L, Vullo C, Salas A. 2015. The complete mitogenome of a 500-year-old Inca child mummy. *Sci Rep.* 5:16462.
- González-Andrade F, Sánchez D, González-Solórzano J, Gascón S, Martínez-Jarreta B. 2007. Sex-specific genetic admixture of Mestizos, Amerindian Kichwas, and Afro-Ecuadorans from Ecuador. *Hum Biol.* 79(1):51–77.
- Guidon N, Delibrias G. 1986. Carbon-14 dates point to man in the Americas 32,000 years ago. *Nature* 321(6072):769–771.
- Hartmann A, Thieme M, Nanduri LK, Stempf T, Moehle C, Kivisild T, Oefner OJ. 2008. Direct Submission.
- 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.
- Hooshiar Kashani B, Perego UA, Olivieri A, Angerhofer N, Gandini F, Carossa V, Lancioni H, Semino O, Woodward SR, Achilli A, et al. 2012. Mitochondrial haplogroup C4c: a rare lineage entering America through the ice-free corridor? *Am J Phys Anthropol.* 147(1):35–39.
- Ingman M, Kaessmann H, Pääbo S, Gyllensten U. 2000. Mitochondrial genome variation and the origin of modern humans. *Nature* 408(6813):708–713.
- Karachanak S, Carossa V, Nesheva D, Olivieri A, Pala M, Hooshiar Kashani B, Grugni V, Battaglia V, Achilli A, Yordanov Y, et al. 2012. Bulgarians vs the other European populations: a mitochondrial DNA perspective. *Int J Legal Med.* 126(4):497–503.
- Kumar S, Bellis C, Zlojutro M, Melton PE, Blangero J, Curran JE. 2011. Large scale mitochondrial sequencing in Mexican Americans suggests a reappraisal of Native American origins. *BMC Evol Biol.* 11:293.
- Lee EJ, Merriwether DA. 2015. Identification of whole mitochondrial genomes from Venezuela and implications on regional phylogenies in South America. *Hum Biol.* 87(1):29–38.
- Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25(11):1451–1452.
- Lindo J, Achilli A, Perego UA, Archer D, Valdiosera C, Petzelt B, Mitchell J, Worl R, Dixon EJ, Fifield TE, et al. 2017. Ancient individuals from the North American Northwest Coast reveal 10,000 years of regional genetic continuity. *Proc Natl Acad Sci U S A.* 114:4093–4098.
- Lippold S, Xu H, Ko A, Li M, Renaud G, Butthof A, Schröder R, Stoneking M. 2014. Human paternal and maternal demographic histories: insights from high-resolution Y chromosome and mtDNA sequences. *Investig Genet.* 5:13.
- Llamas B, Fehren-Schmitz L, Valverde G, Soubrier J, Mallick S, Rohland N, Nordenfelt S, Valdiosera C, Richards SM, Rohrlach A, et al. 2016. Ancient mitochondrial DNA provides high-resolution time scale of the peopling of the Americas. *Sci Adv.* 2(4):e1501385.
- Martínez-Cruzado JC. 2010. The history of Amerindian mitochondrial DNA lineages in Puerto Rico. In: Fitzpatrick SM, Ross AH, editors. Island shores, distant pasts: archaeological and biological approaches to the Pre-Columbian settlement of the Caribbean. Gainesville (FL): University Press of Florida. p. 54–80.
- Martínez-Cruzado JC, Toro-Labrador G, Ho-Fung V, Estévez-Montero MA, Lobaina-Manzanet A, Padovani-Claudio DA, Sánchez-Cruz H, Ortiz-Bermúdez P, Sánchez-Crespo A. 2001. Mitochondrial DNA analysis reveals substantial Native American ancestry in Puerto Rico. *Hum Biol.* 73(4):491–511.
- Moreau C, Bhéer C, Vézina H, Jomphe M, Labuda D, Excoffier L. 2011. Deep human genealogies reveal a selective advantage to be on an expanding wave front. *Science* 334(6059):1148–1150.
- O'Fallon BD, Fehren-Schmitz L. 2011. Native Americans experienced a strong population bottleneck coincident with European contact. *Proc Natl Acad Sci U S A.* 108:20444–20448.
- Olivieri A, Sidore C, Achilli A, Angius A, Posth C, Furtwängler A, Brandini S, Capodiferro MR, Gandini F, Zoledziewska M, et al. 2017. Mitogenome diversity in Sardinians: a genetic window onto an island's past. *Mol Biol Evol.* 34(5):1230–1239.
- O'Rourke DH, Raff JA. 2010. The human genetic history of the Americas: the final frontier. *Curr Biol.* 20:R202–R207.

- Parenti F, Mercier N, Valladas H. 1990. The oldest hearths of Pedra Furada, Brasil: thermoluminescence analysis of heated stones. *Curr Res Pleistocene*. 7:36–37.
- Parenti F, Fontugue M, Guérin C. 1996. Pedra Furada in Brazil and its “presumed” evidence: limitations and potential of the available data. *Antiquity* 70(268):416–421.
- Perego UA, Achilli A, Angerhofer N, Accetturo M, Pala M, Olivieri A, Hooshiar Kashani B, Ritchie KH, Scozzari R, Kong QP, et al. 2009. Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. *Curr Biol*. 19(1):1–8.
- Perego UA, Angerhofer N, Pala M, Olivieri A, Lancioni H, Hooshiar Kashani B, Carossa V, Ekins JE, Gómez-Carballa A, Huber G, et al. 2010. The initial peopling of the Americas: a growing number of founding mitochondrial genomes from Beringia. *Genome Res*. 20(9):1174–1179.
- Perego UA, Lancioni H, Tribaldos M, Angerhofer N, Ekins JE, Olivieri A, Woodward SR, Pascale JM, Cooke R, Motta J, et al. 2012. Decrypting the mitochondrial gene pool of modern Panamanians. *PLoS ONE*. 7(6):e38337.
- Rademaker K, Hodgins G, Moore K, Zarrillo S, Miller C, Bromley GRM, Leach P, Reid DA, Alvarez WY, Sandweiss DH. 2014. Paleoindian settlement of the high-altitude Peruvian Andes. *Science* 346(6208):466–469.
- Raghavan M, DeGiorgio M, Albrechtsen A, Moltke I, Skoglund P, Korneliusson TS, Grønnow B, Appelt M, Gulløv HC, Friesen TM, et al. 2014. The genetic prehistory of the New World Arctic. *Science* 345(6200):1255832.
- Raghavan M, Steinrücken M, Harris K, Schiffels S, Rasmussen S, DeGiorgio M, Albrechtsen A, Valdiosera C, Ávila-Arcos MC, Malaspina AS, et al. 2015. POPULATION GENETICS. Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* 349(6250):aab3884.
- 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(7411):370–374.
- Rieux A, Eriksson A, Li M, Sobkowiak B, Weinert LA, Warmuth V, Ruiz-Linares A, Manica A, Balloux F. 2014. Improved calibration of the human mitochondrial clock using ancient genomes. *Mol Biol Evol*. 31(10):2780–2792.
- Richards MB, Soares P, Torroni A. 2016. Palaeogenomics: mitogenomes and migrations in Europe’s past. *Curr Biol*. 26(6):R243–R246.
- Sans M, Figueiro G, Hidalgo PC. 2012. A new mitochondrial C1 lineage from the prehistory of Uruguay: population genocide, ethnocide, and continuity. *Hum Biol*. 84(3):287–305.
- Sans M, Figueiro G, Hughes CE, Lindo J, Hidalgo PC, Malhi RS. 2015. A South American prehistoric mitogenome: context, continuity, and the origin of haplogroup C1d. *PLoS ONE*. 10(10):e0141808.
- Santangelo R, González-Andrade F, Børsting C, Torroni A, Pereira V, Morling N. 2017. Analysis of Ancestry Informative Markers in three main ethnic groups from Ecuador supports a trihybrid origin of Ecuadorians. *Forensic Sci Int Genet*. 31:29–33.
- Schroeder KB, Jakobsson M, Crawford MH, Schurr TG, Boca SM, Conrad DF, Tito RY, Osipova LP, Tarskaia LA, Zhadanov SI, et al. 2009. Haplotypic background of a private allele at high frequency in the Americas. *Mol Biol Evol*. 26(5):995–1016.
- Schurr TG, Ballinger SW, Gan YY, Hodge JA, Merriwether DA, Lawrence DN, Knowler WC, Weiss KM, Wallace DC. 1990. Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. *Am J Hum Genet*. 46(3):613–623.
- Schurr TG, Sherry ST. 2004. Mitochondrial DNA and Y chromosome diversity and the peopling of the Americas: evolutionary and demographic evidence. *Am J Hum Biol*. 16(4):420–439.
- Skoglund P, Mallick S, Bortolini MC, Chennagiri N, Hünemeier T, Petzl-Erler ML, Salzano FM, Patterson N, Reich D. 2015. Genetic evidence for two founding populations of the Americas. *Nature* 525(7567):104–108.
- Soares P, Ermini L, Thomson N, Mormina M, Rito T, Röhl A, Salas A, Oppenheimer S, Macaulay V, Richards MB. 2009. Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am J Hum Genet*. 84(6):740–759.
- Taboada-Echalar P, Alvarez-Iglesias V, Heinz T, Vidal-Bralo L, Gómez-Carballa A, Catelli L, Pardo-Seco J, Pastoriza A, Carracedo A, Torres-Balanza A, et al. 2013. The genetic legacy of the pre-colonial period in contemporary Bolivians. *PLoS ONE*. 8:e58980.
- Tackney JC, Potter BA, Raff J, Powers M, Watkins WS, Warner D, Reuther JD, Irish JD, O’Rourke DH. 2015. Two contemporaneous mitogenomes from terminal Pleistocene burials in eastern Beringia. *Proc Natl Acad Sci U S A*. 112(45):13833–13838.
- Tamm E, Kivisild T, Reidla M, Metspalu M, Smith DG, Mulligan CJ, Bravi CM, Rickards O, Martínez-Labarga C, Khusnutdinova EK, et al. 2007. Beringian standstill and spread of Native American founders. *PLoS ONE*. 2(9):e829.
- Tito RY, Polo SI, Lewis CM. 2012. Direct Submission.
- Torroni A, Schurr TG, Yang CC, Szathmary EJ, Williams RC, Schanfield MS, Troup GA, Knowler WC, Lawrence DN, Weiss KM. 1992. Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. *Genetics* 130(1):153–162.
- Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, Vullo CM, Wallace DC. 1993. Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am J Hum Genet*. 53(3):563–590.
- Torroni A, Rengo C, Guida V, Cruciani F, Sellitto D, Coppa A, Calderon FL, Simionati B, Valle G, Richards M, et al. 2001. Do the four clades of the mtDNA haplogroup L2 evolve at different rates? *Am J Hum Genet*. 69(6):1348–1356.
- van Oven M, Kayser M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat*. 30(2):E386–E394.
- Vilar MG, Melendez C, Sanders AB, Walia A, Gieski JB, Owings AC, Schurr TG and Genographic Consortium. 2014. Genetic diversity in Puerto Rico and its implications for the peopling of the Island and the West Indies. *Am J Phys Anthropol*. 155(3):352–368.
- 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(11):e185.
- Waters MR, Forman SL, Jennings TA, Nordt LC, Driese SG, Feinberg JM, Keene JL, Halligan J, Lindquist A, Pierson J, et al. 2011. The Buttermilk Creek Complex and the origins of Clovis at the Debra L. Friedkin Site, Texas. *Science* 331(6024):1599–1603.
- Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci*. 13(5):555–556.
- Zheng H-X, Qin ZD, Jin L, Jin L. 2014. Direct Submission.