

# Reconciling migration models to the Americas with the variation of North American native mitogenomes

Alessandro Achilli<sup>a,1</sup>, Ugo A. Perego<sup>b,c</sup>, Hovirag Lancioni<sup>a</sup>, Anna Olivieri<sup>b</sup>, Francesca Gandini<sup>b</sup>, Baharak Hooshier Kashani<sup>b</sup>, Vincenza Battaglia<sup>b</sup>, Viola Grugni<sup>b</sup>, Norman Angerhofer<sup>c</sup>, Mary P. Rogers<sup>d</sup>, Rene J. Herrera<sup>e</sup>, Scott R. Woodward<sup>c,f</sup>, Damian Labuda<sup>g</sup>, David Glenn Smith<sup>h</sup>, Jerome S. Cybulski<sup>i</sup>, Ornella Semino<sup>b</sup>, Ripan S. Malhi<sup>d,j</sup>, and Antonio Torroni<sup>b,1</sup>

<sup>a</sup>Dipartimento di Chimica, Biologia e Biotecnologie, Università di Perugia, 06123 Perugia, Italy; <sup>b</sup>Dipartimento di Biologia e Biotecnologie “Lazzaro Spallanzani”, Università di Pavia, 27100 Pavia, Italy; <sup>c</sup>Sorenson Molecular Genealogy Foundation, Salt Lake City, UT 84115; <sup>d</sup>Department of Anthropology and <sup>e</sup>Institute for Genomic Biology, University of Illinois, Champaign, IL 61801; <sup>f</sup>Department of Biological Sciences, Florida International University, Miami, FL 33199; <sup>g</sup>AncestryDNA, Provo, UT 84604; <sup>h</sup>Département de Pédiatrie, Centre de Recherche du Centre Hospitalier Universitaire Sainte-Justine, Université de Montréal, Montréal, QC, Canada H3T 1C5; <sup>i</sup>Department of Anthropology, University of California, Davis, CA 95616; and <sup>j</sup>Canadian Museum of Civilization, Gatineau, QC, Canada K1A 0M8

Edited by Francisco Mauro Salzano, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil, and approved July 16, 2013 (received for review April 8, 2013)

In this study we evaluated migration models to the Americas by using the information contained in native mitochondrial genomes (mitogenomes) from North America. Molecular and phylogeographic analyses of B2a mitogenomes, which are absent in Eskimo–Aleut and northern Na-Dene speakers, revealed that this haplogroup arose in North America ~11–13 ka from one of the founder Paleo-Indian B2 mitogenomes. In contrast, haplogroup A2a, which is typical of Eskimo–Aleuts and Na-Dene, but also present in the easternmost Siberian groups, originated only 4–7 ka in Alaska, led to the first Paleo-Eskimo settlement of northern Canada and Greenland, and contributed to the formation of the Na-Dene gene pool. However, mitogenomes also show that Amerindians from northern North America, without any distinction between Na-Dene and non-Na-Dene, were heavily affected by an additional and distinctive Beringian genetic input. In conclusion, most mtDNA variation (along the double-continent) stems from the first wave from Beringia, which followed the Pacific coastal route. This was accompanied or followed by a second inland migratory event, marked by haplogroups X2a and C4c, which affected all Amerindian groups of Northern North America. Much later, the ancestral A2a carriers spread from Alaska, undertaking both a westward migration to Asia and an eastward expansion into the circumpolar regions of Canada. Thus, the first American founders left the greatest genetic mark but the original maternal makeup of North American Natives was subsequently reshaped by additional streams of gene flow and local population dynamics, making a three-wave view too simplistic.

Native Americans | human mtDNA

**P**leistocene glaciations—particularly the Last Glacial Maximum (LGM)—played an important role in shaping current patterns of animal and plant genetic diversity in many regions of the world. As for humans, the lowered sea level uncovered ~20% more terrain, which in some places acted as natural land-bridges that facilitated expansions to new and unexplored regions, including the American double-continent (1, 2). The general consensus is that modern Native Americans trace their ancestry to a limited number of original founders whose gene pool ultimately derived from Asian groups that peopled northeast Siberia, including parts of Beringia, before the LGM (3–7). The ancestral Beringian populations probably retreated into refugia during the Ice Age, where their genetic variation was reshaped not only because of drift, but also because of admixture with population groups newly arrived from regions located west of Beringia. Thus, pre-LGM haplotypes of Asian ancestry were differently preserved, modified, and lost in Beringian enclaves (8–13).

One very contentious issue is whether the settlement occurred by means of a single (14–16) or multiple (17–21) streams of migration. In developing hypotheses to address these questions, most

analyses of Native American genetic diversity have examined single loci, particularly the maternally inherited mitochondrial DNA (mtDNA), and some interpretations of modern and ancient data point to a major migratory wave from Beringia dated between 15 and 18 ka (14, 22–25). However, a recent genome-wide scan (364,470 SNPs) of Native American and Siberian groups (26) identified three different source populations for Native Americans that revived a more complex and long-debated scenario, also known as the tripartite migration model, which was originally proposed by combining anthropometric, genetic, and linguistic data (19, 27–29). This scenario postulates that the Americas were settled through three separate population movements whose identity was expressed in linguistic terms as Amerinds, Na-Dene, and Eskimo–Aleut speakers (*SI Text*).

Although recent (26) and some very early nuclear (28) and mtDNA (18, 30, 31) data would favor a minimum of three distinct streams of gene flow from Beringian/Siberian sources, neither this model, nor the alternative scenarios have been fully evaluated by using the information contained in the entire mitochondrial genome (mitogenome). To evaluate this issue, we focused on mitogenomes belonging to two haplogroups, known as A2a and B2a, which are characterized by peculiar geographic distributions. Among the numerous subclades radiating from the root of the pan-American haplogroup A2 (32–34), A2a (together with A2b) mtDNAs have been mostly identified only in Siberia, in Alaska, and the surrounding regions, and in Natives from the American Southwest (21, 31, 35–41). B2a radiates from the root of B2, another common pan-American haplogroup. However, this particular B2 branch has also been observed exclusively in North America, just to the south and more widespread than A2a (42). In this study we report 41 additional mitogenomes belonging to A2a and B2a as well as detailed phylogeographic analyses of the two haplogroups in the general context of North American mtDNA variation.

Author contributions: A.A., U.A.P., R.S.M., and A.T. designed research; A.A., U.A.P., H.L., A.O., F.G., B.H.K., V.B., V.G., and M.P.R. performed research; A.A., R.J.H., S.R.W., D.L., D.G.S., J.S.C., O.S., R.S.M., and A.T. contributed new reagents/analytic tools; A.A., U.A.P., H.L., A.O., N.A., and A.T. analyzed data; and A.A., U.A.P., D.L., O.S., R.S.M., and A.T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. [KC710999–KC711039](https://doi.org/10.1093/ncbi/kct711039)).

<sup>1</sup>To whom correspondence may be addressed. E-mail: [alessandro.achilli@unipg.it](mailto:alessandro.achilli@unipg.it) or [antonio.torroni@unipv.it](mailto:antonio.torroni@unipv.it).

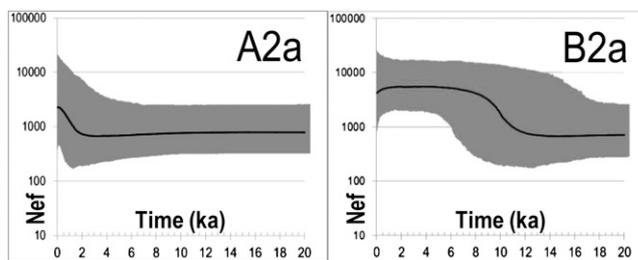
This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1306290110/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1306290110/-DCSupplemental).

## Results

**Phylogenies, Age Estimates, and Population Expansions.** The phylogenetic relationships of a total of 46 A2a and 38 B2a mitogenomes from this and previous studies (Table S1) are shown in Fig. 1 and Fig. S1 (see *SI Results* for further details). Age estimates for A2a based on the entire mitogenome suggest that it arose well after the LGM, possibly between ~4 and ~7 ka (maximum likelihood: 4.03 ka;  $\rho$ : 7.36 ka) (Table S2). The circumpolar subbranches A2a2 and A2a3 radiated almost concomitantly ~2.5 ka, whereas the two southern North American subclusters arose about 1 ka (A2a5) or even later (A2a4). Instead B2a appears to have arisen toward the end of Pleistocene between ~12 and ~13 ka (maximum likelihood: 12.07 ka;  $\rho$ : 13.00 ka) with three older branches (B2a1, B2a2, and B2a4) radiating in the early Holocene (~10 ka) and two younger ramifications (B2a3 and the “Southwestern” B2a5) dated less than ~5 ka.

To assess population expansions that might have involved the two haplogroups, comparable Bayesian skyline plots (BSPs) for both A2a and B2a were performed. The BSP for B2a points to a main episode of population growth dated at ~10 ka, whereas the steepest A2a expansion took place less than ~2 ka (Fig. 2). The age estimate obtained for the A2a overall tree was also used to run analyses for two different clusters of samples that might be geographically classified as Siberia/Alaska/Greenland and the remainder (mostly from the American Southwest), thus roughly separating circumpolar people (mainly Eskimos) from the others (Fig. S2). This process allowed us to assess whether the two subsets were characterized by similar or different expansion times. The emerging picture suggests that the circumpolar expansion took place much earlier, between 4 and 3 ka, than the diffusion into the Southwest (< 1–2 ka).

**Geographic Distributions of A2a and B2a mtDNAs.** To further investigate the geographical distributions of haplogroups A2a and B2a, we compared the control-region variation of the sequences in Fig. 1 with that of published data (mainly from hypervariable segment I; HVS-I) from both Native American groups and general-mixed national populations living in Siberia and North/



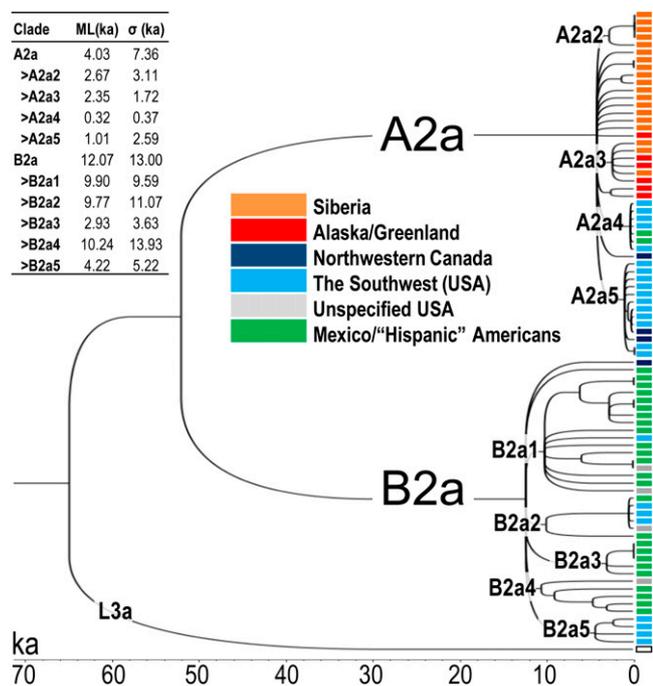
**Fig. 2.** BSP showing population size trends based on A2a and B2a mtDNAs. The thick solid line is the median estimate and the shading shows the 95% highest posterior density limits. The time axis is limited to 20 ka; beyond that time the curves remain linear.

Central America. In agreement with the geographic and the ethnic distribution of the A2a and B2a mtDNAs mainly identified in the Sorenson Molecular Genealogy Foundation Mitochondrial database (Fig. 1 and Table S1), it is apparent that the geographical distributions of A2a and B2a share some common features, as well as several distinctive peculiarities, which are shown in Fig. 3 (see Table S3 for a detailed sample list). Both A2a and B2a harbor a frequency peak in the American Southwest, but whereas A2a mtDNAs are mostly restricted to the typical Athapaskan territories, B2a mtDNAs are widely spread all over the Southwest both in mixed populations (present study) and indigenous populations, including Apache and Navajo (42). Traces of B2a are also found northward in one Turtle Mountain Chippewa (43) and in one Tsimshian from Canada (Fig. 1 and Table S1). There is some evidence that it might also be present in the Bella Coola tribe (5, 44) and among the Ojibwa (5, 45), based solely on the 16111 HVS-I marker. Looking southward, a significant incidence of B2a found in Mexico is confirmed both in tribal groups (42) and general populations (present study), but not in Central America (33, 42). Among the two analyzable subclades, defined by a control-region marker (Fig. S3), the B2a4 is widespread in North-Central Mexico and the American Southwest, whereas the B2a5 is mainly restricted only to the Yuman (~5%) and the Uto-Aztecan Pima and Papago populations from Arizona (~7%).

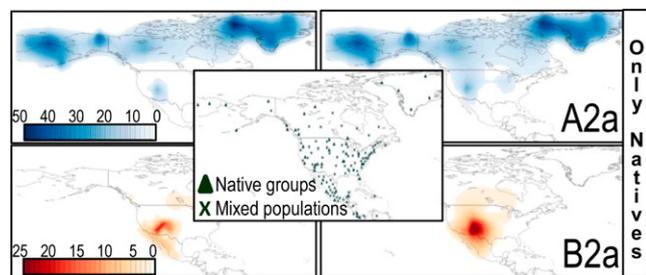
As expected, the frequency peaks of A2 (often higher than 40–50%) are located in the circumpolar regions from Siberia all of the way to Greenland, but the Southwest distribution is limited to the Athapaskan groups (Apache ~48% and Navajo ~13%) and the general population ( $\leq 7\%$ ) of New Mexico, Arizona, and California, in decreasing frequency order. The mtDNAs of these groups could be classified mostly as A2a5 (~95%), with the remainder falling into the A2a4 branch. As for the remaining A2a subclades, A2a2 has no control-region markers and A2a3 is essentially confined to northern North America (Fig. S3 and Table S3).

## Discussion

**Dating and Tracing the Origin and Spread of B2a.** A recent study, performed at the control-region level, proposed that B2a mtDNAs mark a population expansion within the American Southwest (42), and tested the hypothesis that this expansion was because of the spread of groups of Uto-Aztecan farmers who introduced maize agriculture into this area from Mesoamerica ~4 ka (46). Our analyses at the level of entire mitogenomes provide additional information that suggests a different scenario. The overall B2a coalescence time is at ~12–13 ka, and we obtained a similar estimate when combining the previously published B2a control-region data with the control-region sequences of our mitogenomes. The network in Fig. S4 shows that B2a mtDNAs diverge from the root (central node) by  $1.2 \pm 0.3$  control-region substitutions on average, which correspond to ~11 ka [95% confidence interval (CI) 5.5–16.2 ka] when using the molecular clock of Soares et al. (47). Therefore, the onset of B2a should be predated to 11–13 ka in the late



**Fig. 1.** Schematic phylogeny of complete mtDNA sequences belonging to haplogroups A2a and B2a. A maximum-likelihood (ML) time scale is shown. (Inset) A list of exact age values for each clade.



**Fig. 3.** Spatial frequency distributions (percent) of haplogroups A2a and B2a. The maps on the left include both Native American groups and general-mixed populations but only Native groups were considered on the right.

Pleistocene. This finding is further supported by our BSP analyses. Indeed, the main episode of population growth associated with B2a dates to 8–10 ka (Fig. 2); thus, it is likely too ancient to be associated with the farming/language dispersal hypothesis.

In search of alternative scenarios, one question immediately arises. Could B2a have been one of the many founder mtDNA haplogroups from Beringia? In the context of the Americas, this is a challenging issue, not only because the ancestral source populations in Siberia and Beringia no longer exists, but also because modern Asian and Siberian populations completely lack certain haplogroups. Therefore, the identification of founder mtDNA sequences from the analysis of modern individuals is based solely on the evaluation of two parameters: coalescence time and geographical distribution of the derived lineages from the postulated American founder.

Considering that B2 control-region haplotypes with the diagnostic B2a transition G16483A are not found either in present-day regions of former Beringia or further south than Central Mexico, and that our B2a mtDNAs coalesce at about 12–13 ka (more recently than the 15- to 18-ka time range reported for the 16 mitogenome founders identified to date), it is unlikely that B2a may have been a founder haplogroup from Beringia. Instead it is more likely that the diagnostic mutational motif of B2a evolved *in situ* (in North America) a few millennia after that B2 had already entered and spread along the double continent. If so, where did the first B2a mtDNA arise? One possibility is that it originated in the Pacific coastal regions from one of the local descendants of the B2 mitogenomes that had rapidly spread along the coast (48) and gave rise (at different stages of their spread) to other specific B2 subhaplogroups, such as B2i (49) and B2b (50) on their way from Beringia to South America. An alternative explanation is that this B2 mtDNA entered the American continent through the ice-free corridor with the same population groups that introduced the C4c and X2a mitogenomes (51, 52). Although at the moment there is no evidence that, in addition to the main Pacific coastal route, ancestral B2 mtDNAs also entered North America through the ice-free corridor, such a scenario cannot be completely ruled out. Further evidence for a B2a origin along the Pacific coast (possibly as far north as the Northwest Coast), sometime after that the Paleo-Indian migration wave had gone through, is provided by the Tsimshian (British Columbia) mitogenome that radiates directly from the B2a root (Fig. 1 and Table S1). Soon afterward, ~10 ka, the deepest and most represented sublineages of B2a—which we termed B2a1, B2a2, and B2a4—arose possibly in the (greater) American Southwest and Mexico. It is worth noting that the main expansion signal for B2a occurs at approximately 9 ka (Fig. 2), predating both the farming diffusion linked to maize domestication (53) and the Uto-Aztecan culture (54, 55) and, according to Brown (56), even the formation of the oldest Native American language families. In this view, the spread of agriculture into the Southwest is probably marked by some of the younger B2a branches, such as B2a5, whose age and geographic distribution are both consistent with such event (Fig. S3, and Tables S1 and S2).

Finally, the high frequency of B2a in both Southern Athapaskan groups (Apache: 14.8%; Navajo: 15.1%) and its absence in all other Na-Dene groups from North America remain to be explained. The fact that none of the detected B2a sublineages are Southern Athapaskan-specific confirms that their B2a mtDNAs are most likely the result of female gene flow from surrounding non-Na-Dene populations over the last 500–1,000 y that elapsed since the first Athapaskan groups entered into the Southwest moving from their northern ancestral homeland (5, 40). When the Europeans arrived in the area, trade between the long-established Pueblo people and the Southern Athapaskans was well established. Spanish records speak of the Athapaskans traveling to the Pueblos, or living in the vicinity of them, to exchange bison meat, hides, and stone goods for maize and woven cotton (57, 58). The Navajo willingness to adopt the Pueblo lifestyle, as well as the Apache raiding and capturing of neighboring women, probably brought the B2a haplogroup, as well as others, into the Southern Athapaskan mtDNA gene pool, a scenario envisioned also by very early studies with classic genetic markers (28, 59) and mtDNA restriction fragment length polymorphism haplotypes (18).

**Dating and Tracing the Origin and Spread of A2a.** Using a dataset combining both general-mixed and Native American populations, we have shown that the common ancestor of contemporary A2a mitogenomes originated much later than B2a, some 4–7 ka, in agreement with previous estimates (21). Such a temporal dissimilarity is also reflected in terms of their spatial distributions within the North American continent. Among the 130 populations considered in our control-region survey (Table S3), only seven showed both A2a and B2a mtDNAs. These populations include the two southern Athapaskan groups (Apache and Navajo), four general-mixed population samples from the American Southwest (Arizona and California) and northern Mexico (Chihuahua), and (perhaps) the Oneota specimens buried at the Norris Farms cemetery in Illinois. Most of the other A2a mitogenomes are found in the arctic regions of North America and Northeast Asia, with the three highest frequency peaks observed in northeastern Siberia, Alaska, and Greenland (Fig. 2). Most of these mtDNAs belong to the A2a3 (Fig. S3) and A2a2 subbranches (Fig. S1), even if for the latter clade, which is defined only by a coding-region marker, this is based only on a rather limited number of available mitogenomes (Fig. 1). A secondary peak was observed in the American Southwest, where most A2a mtDNAs belong to A2a4 and A2a5. However, A2a5 mtDNAs have also been observed in Alaska, the Pacific Northwest, and northwestern Canada, with a frequency peak in the Athapaskan Dogrib (26.7%) (Table S3). Even if a Siberian origin for A2a cannot be completely ruled out, these phylogeographic profiles support the scenario that one or more enclaves in Alaska (21) (or in the westernmost part of the Northwest Territories of Canada) might have functioned as incubators and ancestral homelands initially for A2a as a whole and later for its major sublineages. This finding is in agreement with the oldest archaeological evidence of the Arctic Small Tool tradition in the Americas, which is dated back 5.5 ky in the Kuzitrin Lake of Alaska (60). In the rather short period of 1–2 ky, the A2a2 and A2a3 variants, typical of contemporary circumpolar populations, evolved from the bearers of haplogroup A2a root, thus probably matching the beginnings of Paleo-Eskimo cultures. Our analyses reveal that the female effective size of these ancestral Paleo-Eskimo populations underwent a detectable bidirectional expansion temporally starting from approximately 4 ka (Fig. S2), covering all of the arctic regions both in northern Canada and Greenland (on the American side) and in northeastern Siberia (on the Asian side), as a result of the back-migration process proposed by Tamm et al. (8). This expansion event is the same that led to the spread of: (i) mtDNA haplogroup D2a1, found in the northernmost Eskimos (and the single individual available from the extinct

Paleo-Eskimo Saqqaq culture, dated to 3.4–4.5 ka; but hopefully other Paleo-Eskimo remains will be found and analyzed soon), Chukchi, Aleuts, Athapaskans, and possibly the Tlingit (21, 61); and (ii) Y-chromosome haplogroup Q1a\*-MEH2, found in the Koryaks and Chukchi of Northeast Siberia, as well as in the ancient Paleo-Eskimo already mentioned above (12, 62, 63). Such circumpolar population dynamics neither involved the pan-American lineages that were spread all over the Americas by the Paleo-Indians, nor the subsequent Neo-Eskimo (Thule) groups that probably brought the D3 (and perhaps A2b1) mtDNAs from Alaska to South Greenland only one millennium ago (21, 61, 64).

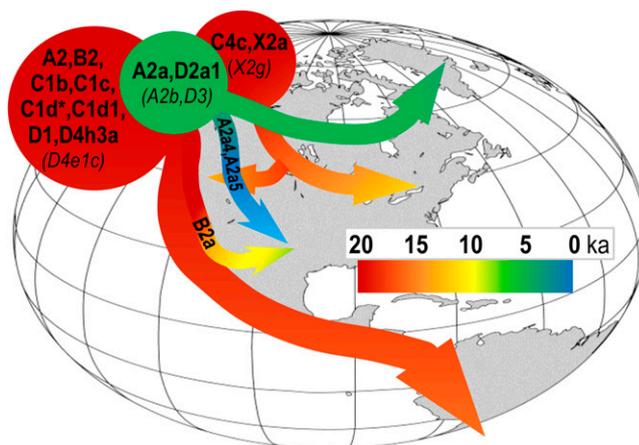
The expansion of the native populations of Alaska and north-west Canada is also confirmed by the current geographic distributions of subhaplogroups A2a4 and, especially, A2a5, which in turn reflect the southward migration of the Athapaskans, dated less than 1 ka, and confirmed also by close Y-chromosome relationships between Athapaskans from the Subarctic and the Southwest (65). Subhaplogroup A2a5 is common in the Athapaskans (9.5%) and Tlingit (15.6%) from Alaska, but harbors a frequency peak among the Dogrib of Northwest Canada (26.7%). Among the Southern Athapaskans, A2a5's average frequency is even higher, with a peak of 47.3% in the Apache (Fig. S3 and Table S3). Based on the available mtDNA data, the southward expansion of the ancestors of the Southern Athapaskans most likely occurred along the eastern side of the Rocky Mountains, but a parallel Pacific coastal route is also a possibility. In both cases, it is evident that the Paleo-Eskimos, who moved eastward into the American Arctic (or returned to Northeast Siberia), the Tlingit, and the Athapaskans share an ancestral A2a mitogenome that, at different times and locations, differentially evolved within distinct North American population groups.

**Correlating the Overall Variation of MtDNA Lineages in North America with Migration Models.** More than one migration wave was originally hypothesized to explain the introduction of distinct linguistic families, such as Eskimo–Aleut and Na-Dene, into the Americas. Recent Y-chromosome data supported the scenario that the Eskimoan- and Athapaskan-speaking groups were the result of two population expansions that occurred after the initial wave of American settlement (20). The old tripartite linguistic subdivision into Eskimo–Aleuts, Na-Dene, and Amerinds of Greenberg et al. (19) has been somehow recently revived to explain the pattern of nuclear diversity (over 364,000 SNPs) identified in 52 Native populations: 3 Eskimo–Aleutians, 1 Na-Dene (the North Athapaskan-speaking Chipewyan of the Northwest Territories of Canada), and 48 other Native American groups from North, Central, and South America (26). The whole-genome profiles were indeed interpreted as indicating that Native Americans descend from at least three streams of Asian/Beringian gene flow. However, we believe that such a stringent parallelism with the linguistic tripartite hypothesis should be regarded as challenging for at least two reasons. First, the reported data concern only a single Athapaskan population. Second, in the neighbor-joining tree [based on *Fst* distances and reported in figure 1C of Reich et al. (26)] the Chipewyan form a cluster with three non-Na-Dene groups (Cree, Ojibwa, and Algonquin) of the northern part of North America, the only other populations from that area included in the survey.

Taking into account these observations, could we reconcile in a different way the three-wave migration model suggested by nuclear data with the overall picture emerging from this and other recent studies which have assessed mtDNA variation at the level of entire mitogenome? As shown in Fig. 4, most of the contemporary Native American mtDNA variation (along the entire double-continent) stems from the same, presumably first wave of American settlement that was dated to 15–18 ka by applying the mitogenome molecular clock to different mtDNA founder lineages including the pan-American haplogroups A2 and B2 (52, 66, 67). In the present study we have investigated two of their

subbranches (A2a and B2a) that exhibit different distributions in North America, the only geographic area in which all three language clusters proposed by Greenberg (29) are present. The estimated age of B2a and its localized geographic distribution indicate that the mutational motif of B2a most likely evolved in one of the (probably numerous) derived population groups that arose locally in North America along the trail of the coastal migration route. The early B2a carriers might have initially diffused along the northwestern Pacific coastal region, eventually resulting in a major in situ expansion in the (greater) American Southwest and northern Mexico (Fig. 4). The mutational motifs of the B2a subhaplogroups were instead fully completed more recently. Therefore, these subclusters, which arose at different times and locations, might provide new clues on the process of local linguistic and ethnic differentiation. For example, the B2a5 mtDNAs were probably carried in the Southwest ~4–5 ka by the Uto-Aztecan farmers that introduced and diffused the maize agriculture. Such an early process of regional genetic differentiation was also evidenced by the pioneering analyses of classical nuclear markers (68–71).

On the other hand, A2a mtDNAs were certainly involved in the northernmost population movements that affected the Paleo-Eskimos and took place in the Arctic regions of North America (and Northeast Asia). Our phylogeographic data indicate that the present-day A2a mitogenomes coalesce to ~4–7 ka and the variants detected among the Eskimo–Aleut speakers of both Northeast Asia and northern North America experienced the steepest population expansion ~4 ka, thus confirming that this lineage is a good candidate marker for a separate and more recent migratory wave associated with the first Paleo-Eskimo settlement of the northernmost part of Canada and Greenland (Fig. 4), as also supported by the Y-chromosome prevalence (and diversity) of the Q1a6-NWT01 sublineage in the Western Canadian Inuit (Inuvialuit) (20). The ancestral A2a Paleo-Eskimo carriers undertook both a back-migration to Asia and an eastward path along the circumpolar region of Canada and Greenland. About 3,000 y later, from approximately the same northern regions, the Athapaskan populations moved to the Southwest carrying different A2a variants (A2a4 and A2a5). This information provides a scenario of how different languages might be associated with distinctive gene pools,



**Fig. 4.** In this schematic overview, the 16 mtDNA founder lineages are associated with three major migratory events. Note that the location of the three spheres is approximate. In parentheses (and in italics) are indicated those founder lineages that are not yet sufficiently analyzed. Two additional events are indicated by stealth arrows. The first arrow corresponds to the recent southward spread of the Athapaskans (marked by A2a4 and A2a5). The second arrow marks the major in situ expansion of B2a.

not only at the level of the nuclear genome, but also at the level of mtDNA haplogroups or subhaplogroups.

Finally, there is no reason to assume that the additional (intermediate) migration wave, identified by Reich et al. (26) for the moment only in the Chipewyan, should be restricted to the Na-Dene speakers. In the American portion of their neighbor-joining tree, after the first split of the Eskimo–Aleut speakers and the far-northeastern Siberians, the Athapaskan-speaking Chipewyan and three Algonquian-speaking populations create a distinctive northern North American branch, which is well separated from the major branch encompassing all other more southern Native Americans. Thus, it is indeed conceivable that this intermediate branch indicates a migration wave that was restricted not linguistically, but rather geographically. This theory is also consistent with the observation that Chipewyans derive 90% of their nuclear genome from Paleo-Indian ancestors (26), which would imply the closest genetic relationship (and admixture) with surrounding non–Na-Dene native groups. Evidence on this issue has been recently provided by studies of mitogenomes from North America. At the moment we know that 7 of the 16 recognized mtDNA founding lineages are restricted to North America: A2b, D2a, and D3 are mostly limited to the northernmost regions (similarly to A2a), thus probably being involved only in the population dynamics of the arctic areas (21, 38, 61); X2g has been identified until now only in a single Ojibwa subject; and C4c and X2a show peculiar geographic and ethnic distributions that require special attention (51, 52, 72). Phylogeographic analyses of these two latter haplogroups have led to the conclusion that they might have been carried to North America by Beringian populations, which arrived through the ice-free corridor between the Laurentide and Cordilleran ice sheets, either concomitantly or some time after the southward spread of the Beringian groups who were instead following the Pacific coastal route. Either the location of the ice-free corridor on the east side of the Rocky Mountains along the Mackenzie river basin, or a delayed arrival compared with the groups that were moving along the Pacific, would have limited the penetration into North America of the population(s) carrying haplogroups X2a and C4c. Indeed, the spatial distributions of X2a and C4c are not only limited to North America, but with an interesting exception (see below), they are restricted to the northern part of North America, encompassing numerous populations ranging from the Nuu-Chah-Nulth of Vancouver Island and the Yakima of Washington State on the Pacific coast, to the Algonquian-speakers, including the Mi'kmaq of the Atlantic regions of Canada (43, 45, 51, 52, 73). Most likely, the expansion of C4c and X2a in America occurred in the Great Plains region and our skyplot analysis suggests that the effective number of X2/C4c ancestral females began to increase from ~8–10 ka (Fig. S2) right after the start of the stable, warm Holocene, which signed major postglacial expansions worldwide (74, 75).

At the moment there are no data concerning the distribution of X2 and C4c, as well as other haplogroups, in most of the northern Na-Dene populations, but the exception mentioned above concerns the southern Athapaskans, in which X2a has been indeed found (41, 43, 52). Overall, these observations suggest that haplogroups X2a and C4c might be present also in the northern Athapaskan groups to whom Apache and Navajo are closely related. If so, this finding would imply that X2a and C4c mtDNAs are shared between Na-Dene and non–Na-Dene of the northernmost regions of North

America. Thus, the “intermediate” migration highlighted by nuclear data in the Chipewyan by Reich et al. (26) would be part of a larger-scale migratory event corresponding to the spread of the haplogroups X2a and C4c into North America, an event that did not affect only the ancestor of modern Na-Dene. An analogous differentiated male ancestry for northern Amerindians relative to more southern Native American groups is also supported by the diversity of Y-chromosome lineages (76–78). This is also the highest likelihood scenario obtained by applying a copying model simulation on 2,540 linked SNPs across 32 autosomal regions (79, 80), by analyzing the diversity of X-chromosome *dys44* haplotypes (10) and even by some early classic genetic data on albumin showing that the Naskapi (AL\*Naskapi) variant was shared and limited to Athapaskan and Algonquian-speaking populations (59).

As shown (and dated) in the present study, the distinct, possibly intermediate, genetic input marked on the matrilineal side by mtDNA haplogroups X2a and C4c and the subsequent population expansion extensively predated the spread of A2a in the circumpolar regions of North America. Moreover, the observation that A2a, among all Amerindian populations of northern North America, is very common and almost exclusively present in the Na-Dene, with limited occurrences in other local Amerindian groups, suggests a distinctive, at least dual, genetic contribution to the formation of the Na-Dene: (i) an “older” Beringian genetic component arrived through the interior corridor and shared only with surrounding non–Na-Dene populations of northern North America, and (ii) a more recent genetic input of Alaskan origin maternally marked by mtDNA haplogroup A2a.

As a final remark, it is evident that the arrival of the first American founders, when the territory was empty, left the greatest genetic mark (81), but the original maternal and paternal makeups of North American Natives were subsequently reshaped by additional streams of gene flow and local population dynamics, making even a three-wave view by far too simplistic.

## Materials and Methods

Detailed materials and methods are provided in *SI Materials and Methods*.

**Analysis of Entire Mitochondrial Genomes.** A total of 41 candidate mtDNAs (22 A2a and 19 B2a) (Table S1) were completely sequenced after being selected based on the presence of the A2a/B2a control-region motifs relative either to the revised Cambridge Reference Sequence (82) or to the Revised Sapiens Reference Sequence (83). Complete sequencing and phylogeny construction was achieved as previously described (52).

**Coalescence and Expansion Times.** Coalescence times were estimated by both maximum likelihood (PAML 4.5) and the  $\rho$ -statistics (performed by considering all substitutions or only synonymous mutations). Mutational distances were converted into years using corrected molecular clock proposed by (47). Population expansions were assessed through BSPs (84) from BEAST 1.7.4 (85). An L3a mitogenome was used as an outgroup (86) by assuming an L3 age of 65 ka (95% CI: 60–70 ka) as a consistent internal calibration point (87). Finally, frequency maps were built with Surfer 9 (Golden Software).

**ACKNOWLEDGMENTS.** This research was supported in part by the Sorenson Molecular Genealogy Foundation (U.A.P. and S.R.W.); National Science Foundation Grant BCS-0745459 (to R.S.M.); and the Italian Ministry of Education, University and Research: Progetti Futuro in Ricerca 2008 (RBF08U07M) and 2012 (RBF08U07M) (to A.A. and A.O.) and Progetti Ricerca Interesse Nazionale 2009 and 2012 (to A.A., A.T., and O.S.).

- Curry A (2012) Ancient migration: Coming to America. *Nature* 485(7396):30–32.
- Goebel T, Waters MR, O'Rourke DH (2008) The late Pleistocene dispersal of modern humans in the Americas. *Science* 319(5869):1497–1502.
- Mazières S (2011) Towards a reconciling model about the initial peopling of America. *C R Biol* 334(7):497–504.
- 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.
- Torroni A, et al. (1993) Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am J Hum Genet* 53(3):563–590.
- O'Rourke DH, Raff JA (2010) The human genetic history of the Americas: The final frontier. *Curr Biol* 20(4):R202–R207.
- Dulik MC, et al. (2012) Mitochondrial DNA and Y chromosome variation provides evidence for a recent common ancestry between Native Americans and Indigenous Altaians. *Am J Hum Genet* 90(2):229–246.
- Tamm E, et al. (2007) Beringian standstill and spread of Native American founders. *PLoS ONE* 2(9):e829.
- Ray N, et al. (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(2):337–345.

10. Bourgeois S, et al. (2009) X-chromosome lineages and the settlement of the Americas. *Am J Phys Anthropol* 140(3):417–428.
11. Arnaiz-Villena A, et al. (2010) The origin of Amerindians and the peopling of the Americas according to HLA genes: Admixture with Asian and Pacific people. *Curr Genomics* 11(2):103–114.
12. Rasmussen M, et al. (2010) Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* 463(7282):757–762.
13. Wang S, Bedoya G, Labuda D, Ruiz-Linares A (2010) Brief communication: Patterns of linkage disequilibrium and haplotype diversity at Xq13 in six Native American populations. *Am J Phys Anthropol* 142(3):476–480.
14. Fagundes NJ, et al. (2008) Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. *Am J Hum Genet* 82(3):583–592.
15. Merriwether DA, Rothhammer F, Ferrell RE (1995) Distribution of the four founding lineage haplotypes in Native Americans suggests a single wave of migration for the New World. *Am J Phys Anthropol* 98(4):411–430.
16. Bonatto SL, Salzano FM (1997) A single and early migration for the peopling of the Americas supported by mitochondrial DNA sequence data. *Proc Natl Acad Sci USA* 94(5):1866–1871.
17. Hubbe M, Harvati K, Neves W (2011) Paleoamerican morphology in the context of European and East Asian late Pleistocene variation: Implications for human dispersion into the New World. *Am J Phys Anthropol* 144(3):442–453.
18. Torroni A, et al. (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.
19. Greenberg JH, et al. (1986) The settlement of the Americas: A comparison of the linguistic, dental, and genetic evidence. *Curr Anthropol* 27(5):477–497.
20. Dulik MC, et al.; Genographic Consortium (2012) Y-chromosome analysis reveals genetic divergence and new founding native lineages in Athapaskan- and Eskimoan-speaking populations. *Proc Natl Acad Sci USA* 109(22):8471–8476.
21. Volodko NV, et al. (2008) Mitochondrial genome diversity in arctic Siberians, with particular reference to the evolutionary history of Beringia and Pleistocene peopling of the Americas. *Am J Hum Genet* 82(5):1084–1100.
22. Raff JA, Bolnick DA, Tackney J, O'Rourke DH (2011) Ancient DNA perspectives on American colonization and population history. *Am J Phys Anthropol* 146(4):503–514.
23. Schroeder KB, et al. (2009) Haplotypic background of a private allele at high frequency in the Americas. *Mol Biol Evol* 26(5):995–1016.
24. Balter M (2011) Archaeology. Do island sites suggest a coastal route to the Americas? *Science* 331(6021):1122.
25. Balter M (2011) Human evolution. Tracing the paths of the first Americans. *Science* 333(6050):1692.
26. Reich D, et al. (2012) Reconstructing Native American population history. *Nature* 488(7411):370–374.
27. González-José R, Bortolini MC, Santos FR, Bonatto SL (2008) The peopling of America: Craniofacial shape variation on a continental scale and its interpretation from an interdisciplinary view. *Am J Phys Anthropol* 137(2):175–187.
28. Williams RC, et al. (1985) GM allotypes in Native Americans: Evidence for three distinct migrations across the Bering land bridge. *Am J Phys Anthropol* 66(1):1–19.
29. Greenberg JH (1987) *Language in the Americas* (Stanford Univ Press, Stanford, CA), p 438.
30. Wallace DC, Torroni A (1992) American Indian prehistory as written in the mitochondrial DNA: A review. *Hum Biol* 64(3):403–416.
31. Torroni A, et al. (1993) mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans. *Am J Hum Genet* 53(3):591–608.
32. Perego UA, et al. (2010) The initial peopling of the Americas: A growing number of founding mitochondrial genomes from Beringia. *Genome Res* 20(9):1174–1179.
33. Perego UA, et al. (2012) Decrypting the mitochondrial gene pool of modern Panamanians. *PLoS ONE* 7(6):e38337.
34. Salas A, et al. (2009) Mitochondrial echoes of first settlement and genetic continuity in El Salvador. *PLoS ONE* 4(9):e6882.
35. Starikovskaya YB, Sukernik RI, Schurr TG, Kogelink AM, Wallace DC (1998) mtDNA diversity in Chukchi and Siberian Eskimos: Implications for the genetic history of Ancient Beringia and the peopling of the New World. *Am J Hum Genet* 63(5):1473–1491.
36. Rubicz R, Schurr TG, Babb PL, Crawford MH (2003) Mitochondrial DNA variation and the origins of the Aleuts. *Hum Biol* 75(6):809–835.
37. Shields GF, et al. (1993) mtDNA sequences suggest a recent evolutionary divergence for Beringian and northern North American populations. *Am J Hum Genet* 53(3):549–562.
38. Helgason A, et al. (2006) mtDNA variation in Inuit populations of Greenland and Canada: Migration history and population structure. *Am J Phys Anthropol* 130(1):123–134.
39. Schurr TG, et al.; Genographic Consortium (2012) Clan, language, and migration history has shaped genetic diversity in Haida and Tlingit populations from Southeast Alaska. *Am J Phys Anthropol* 148(3):422–435.
40. Malhi RS, et al. (2003) Native American mtDNA prehistory in the American Southwest. *Am J Phys Anthropol* 120(2):108–124.
41. Budowle B, et al. (2002) HVI and HVII mitochondrial DNA data in Apaches and Navajos. *Int J Legal Med* 116(4):212–215.
42. Kemp BM, et al. (2010) Evaluating the Farming/Language Dispersal Hypothesis with genetic variation exhibited by populations in the Southwest and Mesoamerica. *Proc Natl Acad Sci USA* 107(15):6759–6764.
43. Malhi RS, Schultz BA, Smith DG (2001) Distribution of mitochondrial DNA lineages among Native American tribes of Northeastern North America. *Hum Biol* 73(1):17–55.
44. Malhi RS, et al. (2004) Patterns of mtDNA diversity in northwestern North America. *Hum Biol* 76(1):33–54.
45. Brown MD, et al. (1998) mtDNA haplogroup X: An ancient link between Europe/Western Asia and North America? *Am J Hum Genet* 63(6):1852–1861.
46. Bellwood PS, Renfrew C (2002) *Examining the Farming/Language Dispersal Hypothesis* (McDonald Institute for Archaeological Research, Cambridge, UK).
47. Soares P, et al. (2009) Correcting for purifying selection: An improved human mitochondrial molecular clock. *Am J Hum Genet* 84(6):740–759.
48. Raff J, Tackney J, O'Rourke DH (2010) South from Alaska: A pilot aDNA study of genetic history on the Alaska Peninsula and the eastern Aleutians. *Hum Biol* 82(5-6):677–693.
49. de Saint Pierre M, et al. (2012) Arrival of Paleo-Indians to the southern cone of South America: New clues from mitogenomes. *PLoS ONE* 7(12):e51311.
50. Taboada-Echalar P, et al. (2013) The genetic legacy of the pre-colonial period in contemporary Bolivians. *PLoS ONE* 8(3):e58980.
51. Hooshair Kashani B, et al. (2012) Mitochondrial haplogroup C4c: A rare lineage entering America through the ice-free corridor? *Am J Phys Anthropol* 147(1):35–39.
52. Perego UA, et al. (2009) Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. *Curr Biol* 19(1):1–8.
53. Merrill WL, et al. (2009) The diffusion of maize to the southwestern United States and its impact. *Proc Natl Acad Sci USA* 106(50):21019–21026.
54. Hill JH (2001) Proto-Uto-Aztecan: A community of cultivators in Central Mexico? *Am Anthropology* 103(4):913–934.
55. Gregory DA, Wilcox DR (2007) *Zuni Origins: Toward a New Synthesis of Southwestern Archaeology* (Univ of Arizona Press, Tucson, AZ).
56. Brown CH (2010) Lack of linguistic support for Proto-Uto-Aztecan at 8900 BP. *Proc Natl Acad Sci USA* 107(11):E34–author reply E35–E36.
57. Pritzker B (2000) *A Native American Encyclopedia: History, Culture, and Peoples* (Oxford Univ Press, New York).
58. Sturtevant WC (2004) *Handbook of North American Indians: Southeast* (Smithsonian Institution, Washington, D.C.).
59. Smith DG, et al. (2000) Implications of the distribution of Albumin Naskapi and Albumin Mexico for new world prehistory. *Am J Phys Anthropol* 111(4):557–572.
60. Harritt RK (1998) Paleo-Eskimo beginnings in North America: A new discovery at Kuzitrin Lake, Seward Peninsula, Alaska. *Etudes Inuit Stud* 22(1):61–81.
61. Gilbert MTP, et al. (2008) Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. *Science* 320(5884):1787–1789.
62. Regueiro M, Alvarez J, Rowold D, Herrera RJ (2013) On the origins, rapid expansion and genetic diversity of Native Americans from hunting-gatherers to agriculturalists. *Am J Phys Anthropol* 150(3):333–348.
63. Malychuk B, et al. (2011) Ancient links between Siberians and Native Americans revealed by subtyping the Y chromosome haplogroup Q1a. *J Hum Genet* 56(8):583–588.
64. Derenko M, et al. (2007) Phylogeographic analysis of mitochondrial DNA in northern Asian populations. *Am J Hum Genet* 81(5):1025–1041.
65. Malhi RS, et al. (2008) Distribution of Y chromosomes among native North Americans: a study of Athapaskan population history. *Am J Phys Anthropol* 137(4):412–424.
66. Achilli A, et al. (2008) The phylogeny of the four pan-American mtDNA haplogroups: Implications for evolutionary and disease studies. *PLoS ONE* 3(3):e1764.
67. Bodner M, 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.
68. Cavalli-Sforza LL, Menozzi P, Piazza A (1993) Demic expansions and human evolution. *Science* 259(5095):639–646.
69. Salzano FM, Callegari-Jacques SM (1988) *South American Indians: A Case Study in Evolution* (Oxford Univ Press, Cambridge, New York).
70. Crawford MH (1998) *The Origins of Native Americans: Evidence from Anthropological Genetics* (Cambridge Univ Press, Cambridge, New York).
71. Salzano FM, Bortolini MC (2002) *The Evolution and Genetics of Latin American Populations* (Cambridge Univ Press, Cambridge, New York).
72. Malhi RS, et al. (2010) Brief communication: Mitochondrial haplotype C4c confirmed as a founding genome in the Americas. *Am J Phys Anthropol* 141(3):494–497.
73. Scozzari R, et al. (1997) mtDNA and Y chromosome-specific polymorphisms in modern Ojibwa: Implications about the origin of their gene pool. *Am J Hum Genet* 60(1):241–244.
74. Soares P, et al. (2010) The archaeogenetics of Europe. *Curr Biol* 20(4):R174–R183.
75. Pala M, et al. (2012) Mitochondrial DNA signals of late glacial recolonization of Europe from near eastern refugia. *Am J Hum Genet* 90(5):915–924.
76. Bisso-Machado R, et al. (2011) Distribution of Y-chromosome Q lineages in Native Americans. *Am J Hum Biol* 23(4):563–566.
77. Bolnick DA, Bolnick DI, Smith DG (2006) Asymmetric male and female genetic histories among Native Americans from Eastern North America. *Mol Biol Evol* 23(11):2161–2174.
78. Bortolini MC, et al. (2003) Y-chromosome evidence for differing ancient demographic histories in the Americas. *Am J Hum Genet* 73(3):524–539.
79. Hellenthal G, Auton A, Falush D (2008) Inferring human colonization history using a copying model. *PLoS Genet* 4(5):e1000078.
80. Conrad DF, et al. (2006) A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. *Nat Genet* 38(11):1251–1260.
81. Moreau C, et al. (2011) Deep human genealogies reveal a selective advantage to be on an expanding wave front. *Science* 334(6059):1148–1150.
82. Andrews RM, et al. (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23(2):147.
83. Behar DM, et al. (2012) A “Copernican” reassessment of the human mitochondrial DNA tree from its root. *Am J Hum Genet* 90(4):675–684.
84. Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol* 22(5):1185–1192.
85. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7:214.
86. Torroni A, Achilli A, Macaulay V, Richards M, Bandelt H-J (2006) Harvesting the fruit of the human mtDNA tree. *Trends Genet* 22(6):339–345.
87. Soares P, et al. (2012) The Expansion of mtDNA Haplogroup L3 within and out of Africa. *Mol Biol Evol* 29(3):915–927.