Mitochondrial Population Genomics Supports a Single Pre-Clovis Origin with a Coastal Route for the Peopling of the Americas


It is well accepted that the Americas were the last continents reached by modern humans, most likely through Beringia. However, the precise time and mode of the colonization of the New World remain hotly disputed issues. Native American populations exhibit almost exclusively five mitochondrial DNA (mtDNA) haplogroups (A–D and X). Haplogroups A–D are also frequent in Asia, suggesting a northeastern Asian origin of these lineages. However, the differential pattern of distribution and frequency of haplogroup X led some to suggest that it may represent an independent migration to the Americas. Here we show, by using 86 complete mitochondrial genomes, that all Native American haplogroups, including haplogroup X, were part of a single founding population, thereby refuting multiple-migration models. A detailed demographic history of the mtDNA sequences estimated with a Bayesian coalescent method indicates a complex model for the peopling of the Americas, in which the initial differentiation from Asian populations ended with a moderate bottleneck in Beringia during the last glacial maximum (LGM), around ~23,000 to ~19,000 years ago. Toward the end of the LGM, a strong population expansion started ~18,000 and finished ~15,000 years ago. These results support a pre-Clovis occupation of the New World, suggesting a rapid settlement of the continent along a Pacific coastal route.

Introduction

In the complex history of human migrations, it is widely accepted that the New World continents were the ones colonized most recently by *Homo sapiens*, most likely from Asia through Beringia. A popular model for the peopling of the Americas suggests that the archaeological remains known as the Clovis complex (thought to be the oldest unequivocal evidence of humans in the Americas) represent the people that first colonized the continent after a late-glacial migration through the ice-free corridor that separated the Laurentide and Cordilleran ice sheets. However, the recently re-evaluated age of the Clovis sites to only between about 12.7 and 13.2 thousand years ago (kya) and the confirmed human presence at the Monte Verde site located in southern South America around 14.5 kya challenge this Clovis-first model and call for alternative hypotheses. Because the earlier date for Monte Verde implies that peopling of the Americas south of Beringia occurred before the ice-free corridor was formed, a first migration along the Pacific coast may have been a viable route. Unfortunately, archaeological verification of this scenario is very difficult because most of the late Pleistocene coast is currently underwater; the sea level has risen more than 120 m since the end of the last glacial maximum (LGM).

The maternally inherited mitochondrial DNA (mtDNA) has been widely used to understand the peopling of the Americas. Since the first studies, it has been found that extant Native American populations exhibit almost exclusively five mtDNA haplogroups (A–D and X) classified in the autochthonous haplogroups A2, B2, C1, D1, and X2a. Haplogroups A–D are found all over the New World and are frequent in Asia, supporting a northeastern Asian origin of these lineages. This distribution, together with the similar coalescence time for these haplogroups, was used to suggest a single-migration model. However, a different pattern of diversification and distribution of haplogroup B found in some studies led some authors to hypothesize that it could represent a later and separate migration from the joint arrival of haplogroups A, C, and D. The history of haplogroup X is more elusive; it is presently found in the New World at a relatively low frequency and only in North America, it is rare in West Eurasians, and it is almost absent in Siberia. In addition, some have claimed that Native American haplogroup X is less diverse and has a younger coalescence time than haplogroups A–D. These differential features have been cited to argue that haplogroup X represents an independent migration to the Americas from Asia or even Europe. More specifically, it has been used to support a putative connection between the European Soltreaan and the American Clovis lithic technologies. This so called “Soltread hypothesis” proposed the colonization of North America by Europeans through the North Atlantic, even...
though this interpretation is heavily debated (e.g., 19). All the five founding haplogroups have been shown to be present in Native Americans in pre-Columbian times. 12,20

In general, the studies on mtDNA control-region variation have been taken to support a pre-Clovis migration, between ~20 and 30 kya, before the LGM, for the single (or the most ancient) migration. 6,21 However, the uncertainties about and range around these dates are very large. One cause for this variation is the limited information content of the mtDNA control region, which is also too divergent to allow reliable substitution-rate estimation by comparison with the chimpanzee. 22 Alternatively, the complete coding region of the mtDNA is being increasingly used to circumvent these limitations in studies of human migrations (e.g., 22,23) but has not been used so far for studying the origin of Native Americans.

Another frequent controversy is about the size of the founding population during the peopling of the Americas. The initial results showing the existence of few founder haplogroups for the mtDNA and Y chromosome suggested a strong population bottleneck, 6 although this interpretation was not supported by further mtDNA studies. 21 However, a recent analysis of several genomic loci, including mtDNA, suggested that the Americas could have been founded by as few as 80 effective individuals, and even the largest values in the credible interval only comprise a few hundred effective individuals. 24 On the other hand, the study of other single genetic systems does not seem to support much loss of genetic diversity during the initial settlement of the continent; 25–28 instead, it concludes that a moderate-intensity bottleneck is the best scenario. Another recent genomic study using exclusively autosomal intergenic markers also suggested moderate values, with the Native American founding population consisting of around 500 effective individuals (95% confidence interval 74–1332). 29

Material and Methods

Subjects
DNA samples were obtained from 58 individuals from South and North American native populations and most have been collected directly by some of the authors (F.M.S., S.E.B.S., M.A.Z., or D.G.S.). Table 1 provides further details on the individuals studied. All ethical guidelines were followed, as stipulated by the institutions involved in the study.

PCR, Sequencing, and Contig Assembling
Given the low quantity of some of our DNA samples, we performed a genomic preamplification protocol by using the GenomiPhi kit (GE Healthcare) on these. The PCR amplifications, using primers covering the entire mitochondrial genome, were performed as described elsewhere. 30 Sequencing reactions covering the entire mitochondrial genome for both strands were read in a MegaBACE 1000 (GE Healthcare) with the ET Terminators cycle sequencing kit. Chromatograms were assembled in individual genomes with the Phred-Phrap-Consed package. 31,32 After an initial visual inspection for low-quality regions in the assembly, we aligned the contigs generated for every individual to each other and to the corrected Cambridge reference sequence (rCRS) 33,34 and checked all variable positions in the original chromatograms. Possible phantom mutations were again verified in the chromatograms and, whenever needed, resequenced from a new PCR product. 35 Although some mtDNAs have a partial sequence already published, 36 the whole genomes were mostly resequenced to ensure maximum quality.

Additional Data
To the 58 genomes obtained here, we added 28 complete mtDNA genomes published throughout the literature (see Table 1). This makes a dataset of 86 complete mtDNA genomes characterized from mainly Native American individuals. We have deliberately restricted our analysis to the populations known as “Amerinds,” leaving aside people from Eskimo-Aleuts and Na-Dené linguistic groups. We 31 and others (reviewed in 6) have already demonstrated that the latter two were part of the single founding population that gave origin to all Native Americans. However, there is also evidence 9 that the Eskimo-Aleuts and Na-Dené diverged from Amerinds > 10 kya and underwent independent population contractions and re-expansions around the circumarctic region. Methods such as Bayesian skyline plot, neutrality tests, etc., are only applicable to a group of populations that share the same demographic history. We have therefore not incorporated the haplogroup D2 in our study, because it is only found in Aleutians and in a few other Beringian populations in low frequency. 37 Two large-scale databases (ref. 38 and GenBank accession numbers DQ282387–DQ282487) encompassing Native American mtDNAs have not been used in the main analyses because they consist of data from nonnative individuals. However, to check the robustness of our conclusions, we also performed most of our analyses by adding these nonnative individuals to generate a dataset of 244 mtDNA genomes comprising all available sequences from the five Native American haplogroups. The differences between the results of the native (86 sequences) and those of the 244 sequence datasets are very small in all analyses. Therefore, our results with the 86 dataset are robust and authentically represent present-day mtDNA diversity in “Amerinds.”

Data Analysis
All statistical analyses were done with the slowly evolving mtDNA coding region (positions 577–16022) only. Control-region sequence was used to confirm haplogroup assignment. To check for mutations separating Native American and Old World haplogroups, we compared our sequences with sequences belonging to Asian (haplogroups A–D) and European (haplogroup X) individuals available in the literature. 16,22,39–47 Basic diversity statistics, neutrality tests, and mismatch distributions were calculated with Arlequin 3.11. 48
Table 1. Individuals Used for the Analyses Whose mtDNAs Were Obtained in This Work or Gathered from Literature

<table>
<thead>
<tr>
<th>Hg</th>
<th>GenBank Accession</th>
<th>Tribe/Population</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>A2</td>
<td>AM17 DQ112832</td>
<td>Auca</td>
<td>47</td>
</tr>
<tr>
<td>A2</td>
<td>haplotype A AF382010</td>
<td>Canary</td>
<td>42</td>
</tr>
<tr>
<td>A2</td>
<td>AM17 DQ112832</td>
<td>Auca</td>
<td>47</td>
</tr>
<tr>
<td>A2</td>
<td>AM17 DQ112832</td>
<td>Auca</td>
<td>47</td>
</tr>
<tr>
<td>A2</td>
<td>AM17 DQ112832</td>
<td>Auca</td>
<td>47</td>
</tr>
<tr>
<td>A2</td>
<td>AM17 DQ112832</td>
<td>Auca</td>
<td>47</td>
</tr>
<tr>
<td>A2</td>
<td>AM17 DQ112832</td>
<td>Auca</td>
<td>47</td>
</tr>
<tr>
<td>A2</td>
<td>AM17 DQ112832</td>
<td>Auca</td>
<td>47</td>
</tr>
<tr>
<td>A2</td>
<td>AM17 DQ112832</td>
<td>Auca</td>
<td>47</td>
</tr>
<tr>
<td>A2</td>
<td>AM17 DQ112832</td>
<td>Auca</td>
<td>47</td>
</tr>
<tr>
<td>A2</td>
<td>AM17 DQ112832</td>
<td>Auca</td>
<td>47</td>
</tr>
<tr>
<td>A2</td>
<td>AM17 DQ112832</td>
<td>Auca</td>
<td>47</td>
</tr>
<tr>
<td>A2</td>
<td>AM17 DQ112832</td>
<td>Auca</td>
<td>47</td>
</tr>
<tr>
<td>A2</td>
<td>AM17 DQ112832</td>
<td>Auca</td>
<td>47</td>
</tr>
<tr>
<td>A2</td>
<td>AM17 DQ112832</td>
<td>Auca</td>
<td>47</td>
</tr>
<tr>
<td>A2</td>
<td>AM17 DQ112832</td>
<td>Auca</td>
<td>47</td>
</tr>
<tr>
<td>A2</td>
<td>AM17 DQ112832</td>
<td>Auca</td>
<td>47</td>
</tr>
<tr>
<td>A2</td>
<td>AM17 DQ112832</td>
<td>Auca</td>
<td>47</td>
</tr>
<tr>
<td>A2</td>
<td>AM17 DQ112832</td>
<td>Auca</td>
<td>47</td>
</tr>
<tr>
<td>A2</td>
<td>AM17 DQ112832</td>
<td>Auca</td>
<td>47</td>
</tr>
<tr>
<td>A2</td>
<td>AM17 DQ112832</td>
<td>Auca</td>
<td>47</td>
</tr>
<tr>
<td>A2</td>
<td>AM17 DQ112832</td>
<td>Auca</td>
<td>47</td>
</tr>
<tr>
<td>A2</td>
<td>AM17 DQ112832</td>
<td>Auca</td>
<td>47</td>
</tr>
<tr>
<td>A2</td>
<td>AM17 DQ112832</td>
<td>Auca</td>
<td>47</td>
</tr>
<tr>
<td>A2</td>
<td>AM17 DQ112832</td>
<td>Auca</td>
<td>47</td>
</tr>
<tr>
<td>A2</td>
<td>AM17 DQ112832</td>
<td>Auca</td>
<td>47</td>
</tr>
<tr>
<td>A2</td>
<td>AM17 DQ112832</td>
<td>Auca</td>
<td>47</td>
</tr>
<tr>
<td>A2</td>
<td>AM17 DQ112832</td>
<td>Auca</td>
<td>47</td>
</tr>
<tr>
<td>A2</td>
<td>AM17 DQ112832</td>
<td>Auca</td>
<td>47</td>
</tr>
<tr>
<td>A2</td>
<td>AM17 DQ112832</td>
<td>Auca</td>
<td>47</td>
</tr>
<tr>
<td>A2</td>
<td>AM17 DQ112832</td>
<td>Auca</td>
<td>47</td>
</tr>
<tr>
<td>A2</td>
<td>AM17 DQ112832</td>
<td>Auca</td>
<td>47</td>
</tr>
</tbody>
</table>
implemented in BEAST v1.4, which applies Markov Chain Monte Carlo integration for parameter estimation over the space of all equally likely trees. Population size dynamics through time (i.e., a Bayesian Skyline plot) were also estimated with this approach in BEAST. It is important to emphasize that in this method all genomes were analyzed simultaneously without the assumption of any phylogenetic structure a priori, such as the existence of haplogroups or the number of founding haplotypes. Estimations were carried out assuming HKY + G model with the same rate used for $\rho$ time estimations but with log-normal relaxation allowed. The analysis was run for 60 million iterations, with the first 10% discarded as burn-in. Genealogies and model parameters were sampled every 1,000 iterations thereafter.

**Results**

As expected, all mitochondrial genomes obtained here grouped in the five known haplogroups, as shown in the schematic trees in Figure 1. The diversity patterns within each Native American haplogroup, including haplogroup X, are remarkably alike. All haplogroups exhibit similar nucleotide diversity values, as well as a marked excess of low-frequency variants that is characteristic of a strong and recent population expansion as shown by significant negative values for Tajima’s D and Fu’s Fs statistics (Table 2), and single waves in the mismatch distribution graphics (Figure 2).

![Schematic Tree from the Five Native American mtDNA Haplogroups with Sequences Obtained Here and Indicating the Coding-Region Substitutions](image)

Letters following positions indicate transversions, and the others are transitions. Transition 3552 in the X2a haplogroup was absent in the individual Na3X from, so it is not considered to be a marker for this haplogroup.

**Table 2. Summary Statistics and Coalescence Times for the Five Native American Haplogroups**

<table>
<thead>
<tr>
<th>Haplogroup</th>
<th>n</th>
<th>$S$</th>
<th>$\pi$ (SD) %</th>
<th>Tajima’s D</th>
<th>Fu’s Fs</th>
<th>$\rho$ (95% CI) $^a$</th>
<th>Bayesian (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A2</td>
<td>16</td>
<td>58</td>
<td>0.0512 (0.0282)</td>
<td>−2.333**</td>
<td>−9.897**</td>
<td>20,552 (14,953–26,151)</td>
<td>21,290 (16,550–28,130)</td>
</tr>
<tr>
<td>B2</td>
<td>21</td>
<td>72</td>
<td>0.0504 (0.0273)</td>
<td>−2.468**</td>
<td>−15.997**</td>
<td>20,307 (15,246–25,369)</td>
<td>22,140 (17,570–28,730)</td>
</tr>
<tr>
<td>C</td>
<td>17</td>
<td>44</td>
<td>0.0417 (0.0233)</td>
<td>−2.097**</td>
<td>−7.200**</td>
<td>17,227 (11,461–22,994)</td>
<td>20,680 (16,830–26,260)</td>
</tr>
<tr>
<td>D1</td>
<td>20</td>
<td>44</td>
<td>0.0484 (0.0263)</td>
<td>−1.596**</td>
<td>−7.280**</td>
<td>21,580 (13,263–29,896)</td>
<td>21,430 (16,850–28,730)</td>
</tr>
<tr>
<td>X2a</td>
<td>12</td>
<td>20</td>
<td>0.0304 (0.0180)</td>
<td>−1.277*</td>
<td>−2.410*</td>
<td>19,783 (6,056–29,910)</td>
<td>20,730 (16,100–29,000)</td>
</tr>
<tr>
<td>Average</td>
<td>19</td>
<td>48</td>
<td>0.0480 (0.0247)</td>
<td>−2.468**</td>
<td>−8.977**</td>
<td>20,552 (14,953–26,151)</td>
<td>21,250 (16,550–28,130)</td>
</tr>
</tbody>
</table>

Summary statistics and coalescence times were based on median-joining calculation ($\rho$) and on Bayesian estimation. $n$ indicates number of sequences, $S$ indicates number of segregating sites, $\pi$ indicates nucleotide diversity, and CI indicates confidence interval.

$^a$ Estimated as $\rho \pm 2 \times$ standard deviation (SD); $^p < 0.10$; $^{**}p < 0.05$. 

---

Please cite this article in press as: Fagundes et al., Mitochondrial Population Genomics Supports a Single Pre-Clovis Origin with a Coastal Route for the Peopling of the Americas, The American Journal of Human Genetics (2008), doi:10.1016/j.ajhg.2007.11.013
Using the standard substitution rate of $1.26 \times 10^{-8}$ per site per year for the mtDNA coding region, all haplogroups show coalescence times around 20 kya, with both a strict ($\rho$) and a relaxed molecular clock method (Bayesian, Table 2), together with a close correspondence in the distribution of coalescence times for each haplogroup (Figure 3). Additionally, similar values are found when we compute a maximum likelihood tree with an external calibration point (i.e., without assuming a predetermined substitution rate, Figure 4A). For each haplogroup, all Native American sequences trace back to a single founder haplotype that can be distinguished from Old World haplogroups by the presence of exclusive mutations (Figure 1) or, in the case of haplogroup C, by specific control region sequence motifs, corroborating the results from Bandelt et al.\textsuperscript{7} We observed 2, 5, 0, 1, and 3 coding-region mutations which are markers for the five Native American haplogroups A–D, and X, respectively. The accumulation of these Native American specific mutations possibly reflect the duration of the transition period after the ancestral population divergence from Asians but before the within-haplogroups diversification.\textsuperscript{7,11,21,54} Because the standard rate for these mutations is equivalent to one substitution per 5138 years,\textsuperscript{39} it suggests that this transition period took several thousand years, likely $>\ 5$ kya. All the above results strongly suggest a scenario in which all five haplogroups were part of a single founding population that ultimately led to the peopling of the whole American continent.

To get a more realistic picture of the complex demographic history associated with the colonization of the New World we applied the Bayesian skyline plot approach\textsuperscript{53} to the whole Native American mtDNA genome sample, noting that this estimate is unaffected by the different proposals for the actual number of founder lineages among present day Native American mtDNAs.\textsuperscript{6} The skyline plot (Figure 4B) identifies a moderate effective population size reduction between ~23–19 kya reaching a minimum of ~1000 women followed by a strong (~100-fold) and rapid size expansion beginning ~19–18 kya and ending ~16–15 kya. It is noteworthy that the time of the population reduction correlates very well with the LGM (23–18 kya) while the expansion dates are in excellent agreement with the end of the LGM, dated around 19–17 kya.\textsuperscript{55,56} The Bayesian skyline plot of each haplogroup separately showed a similar pattern of population expansion (data not shown).

**Discussion**

Overall, the Native American mtDNA genomic data suggests the following scenario for the peopling of the Americas. The transition period between the separation of the Native American mtDNA haplogroups from their Asian ancestors and the start of their diversification and expansion into the Americas was estimated at $>\ 5000$ years. Adding this time to our estimates for the expansion indicates that the beginning of the divergence of the Native American founder population from its Asian ancestral population probably predates the LGM. There is evidence of human settlements in the Arctic around 30 kya.\textsuperscript{57} Therefore, it is possible that the precursors of Native American populations represent a human group that pioneered colonization of northeast Asia before the LGM, during a period of amelioration in the climatic conditions.\textsuperscript{58} Although it was not possible to determine where in northeast Asia this population stayed during this long period of isolation, Beringia represents the best candidate for that location, at least for the moderate bottleneck period (~20 kya) before the expansion. Toward the end of the Pleistocene, Beringia was mostly exposed, and even though
There is also strong evidence that the Beringian environment could sustain human populations during at least part of the LGM, being considered both climatically and ecologically a glacial refugium. mtDNA results further suggest that during the LGM the Native American founding population experienced a reduction in size (with a minimum of around 1000 women) that lasted for 3000–4000 yr. The y axis is the effective number of females. The thick solid line is the median estimate and the thin lines (blue) show the 95% highest posterior density limits estimated with 60 million chains. The time axis is limited to 30 kya.

Recent data have shown that this coastal route was largely ice free by ~19 kya and that the environment improved rapidly, being capable of supporting bears as of ~15 kya. Interestingly, the end of the intense expansion period coincides with the age of the southern South American Monte Verde site, ~14.5 kya. The strong and rapid population growth suggested by our data is consistent with a model in which humans have traveled the >13,000 km along the coast from Alaska to the southern tip of Chile in a few thousand years. All of the above age estimates were calculated with the standard mtDNA coding-region substitution rate that was presented by Mishmar et al. and was used in the great majority of the recent mtDNA studies. A new rate using exclusively synonymous substitutions was recently suggested. We have applied this approach to our dataset, and the only changes in the results are the time estimates, which were in general ~5 kya more recent than those presented here. Because there are some questions concerning this rate and its application (e.g.,), and because it has been used in just a few studies to date, for the sake of comparison we presented here only the age estimated with the rate of Mishmar et al.

This model could help explain why some of the earliest known sites are in coastal South America whereas more recent sites are more frequently situated inland. Associated with the end of the ice age, sea level rose rapidly between ~18 and ~10 kya, inundating most of North America’s Pacific coast, which was exposed during the earliest expansion southward. Some of the earliest sites might occur along the much larger South American western coastal plain because large portions of its prehistoric coastline are still exposed. The human dispersal from the coast into the interior of the continent, perhaps driven by growing population density, depletion of coastal resources, and rising sea levels, was probably delayed by the need to cross the mountain ranges and change living strategies and technologies from those associated with coastal adaptations. Interestingly, a similar model was proposed for the first colonization of Asia, ~65 kya.

The moderate population reduction found here was also supported by recent results from biparental loci and by some earlier results from both mtDNA and nuclear
In Siberia, it has been described in only a few populations, none of which currently inhabit eastern Siberia. It is likely that this haplogroup is absent in eastern Siberian populations because of drift effects, which impact rare variants more strongly. Thus, its probability of being lost through random effects would be high. In support for this hypothesis, we note that current Siberian and Native American sequences belonging to the haplogroup X are distantly related, suggesting that the intermediate lineages have been lost. Finally, it is noteworthy that haplogroup X is not the only one of the Native American haplogroups that is more frequent in the New World than in Siberia; haplogroups A and B also show this pattern.

In the Americas, a likely explanation for the observation that haplogroup X has a much more restricted distribution would be that if we assume it was relatively rare in the founding population, then it could have been lost by successive founder effects and genetic drift as the expansion wave moved southward. Actually, it was recently shown that the probability that an allele (e.g., a founding haplotype) survives and expands spatially and in frequency by “surfing” on the wave of a range expansion depends on its presence in the wave of expansion, which in turn depends largely on its proximity to the edge of the wave. Therefore, using this framework, one could conceive that haplogroup X may have “failed” to expand simply as a result of its location in the expansion wave and/or its low initial frequency. A similar explanation may be used to account for the existence of other similarly rare haplogroups in the Americas, such as the “cayapa” subhaplogroup D, as well as the distribution of some rare Y chromosome haplogroups, without the need to postulate independent colonization events. In addition, the existence of additional, rare founding haplotypes agrees well with the moderate bottleneck estimated here. Such strong and old demographic expansion inferred from our data might also indicate that this was the most important time frame in which major changes in haplogroup composition could occur. Interestingly, two studies with ancient DNA samples scattered over most of the Holocene suggested regional continuity in the frequency of mtDNA haplogroups, indicating that in these populations drift has not played a major role in more recent times.

The fact that the five most common Native American mtDNA haplogroups display similar diversity patterns strongly indicates that they have not been much affected by natural selection. Because human mtDNA does not recombine, directional selection upon a specific substitution would favor the haplotype in which this variant occurs, mimicking a demographic expansion. It is very unlikely that in all haplogroups specific variants that would be favored by natural selection with similar intensity would have occurred by chance and at a similar time. Therefore, our results strongly indicate that the diversity pattern in Native American mtDNA results from a demographic expansion in the founding population in which all founding haplotypes were present.

Our detailed demographical model for the earlier settlement of the Americas has implications for explaining some of the high level of disagreement that has been found among studies from different disciplines, especially in relation to an “exact” date for the peopling of the New World. Our results indicate that, strictly speaking, we will probably never be able to pinpoint a single and precise date for the entering of the Americas, because it occurred when Asia and America were not divided but were connected by the huge land mass of the subcontinent Beringia, and because it lasted several millennia, beginning with the isolation from the Asian ancestors and ending with the population size and range expansion into the continent. Under our model, three periods that may define a date for the peopling of the Americas can be delineated: (1) the colonization of Beringia (because about half of it was “America” at that time) by the founding population; (2) the movement out of Beringia—characterized by the fast colonization of the continental Pacific coastal plain—south of the ice sheets; and (3) the more recent and more extensive colonization of inland continental masses. Furthermore, the probability of coalescence of mtDNA lineages within a population and the chance of finding ancient archeological evidence go in opposite directions. Consider, for example, period 1 in Beringia: The many millennia of isolation followed by reduced population size accelerated the coalescence of mtDNA lineages into the haplogroups founding haplotypes (consequently determining their ages), but these conditions, aggravated by the fact that most of Beringia is now underwater, make it more difficult to find good archaeological
evidence of this period. On the contrary, period 3 should present a much higher probability of finding archaeological sites, but it was very difficult to distinguish from the previous period by the conventional methods of historical demography. Therefore, perhaps some of the heated debates about dissimilar “colonization dates” inferred by different disciplines may actually reflect the estimation of distinct periods as described above. More specifically, the generally more recent colonization dates estimated from extensive archaeological sites from inland North America (e.g.,²) may reflect the third, latest event, whereas the usually more ancient coalescence dates from mtDNA haplogroups⁶ actually reflect the earliest isolation of the founding population in Beringia. Finally, the intermediate genetic dates usually estimated from population expansion times as well as the fact the most ancient archeological sites have been found in the coast may reflect the early and fast settlement along the coast of the continent. Our results emphasize the necessity that the increasing amount of population genetic data should be analyzed with methods that provide more realistic pictures for similarly complex evolutionary histories.

While our manuscript was under review, another study on mtDNA genomic diversity of Native American haplogroups has been published by Tamm et al.⁷³ Their paper also suggests that the major Native American mtDNA haplogroups were part of a single demographic event, dating back to ~15 kya, after a period of isolation in Beringia, supporting our previous model.¹¹ Our study differs from the one of Tamm et al. in several issues (e.g., Tamm et al. do not date the age of haplogroup X expansion in the Americas, and they use a synonymous substitution rate⁴⁷ that furnishes slightly younger dates), but they provide similar conclusions about the colonization process.

Acknowledgments

Grant support was from the Brazilian Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Fundação de Amparo a Pesquisa do Rio Grande do Sul (FAPERGS) and by a CAPES scholarship (N.J.R.F.). We are also grateful to Instituto de Milênio e Programas de Apoio a Núcleos de Excelência for extra support (I.M.S.) and to the National Institutes of Health (D.G.S.). Research was developed with help from CENAPAD-SP supercomputer center. Thanks to Cladina R. Sartori, Ronaldo R. Ferreira, Luana Cardoso-Silva, Renata Schmitt, André Schnorr, Gabrielle D. Salton, Mariana Magalhães, and Marina O. Favarini for technical help, and to Kim Hill, A. Magdalena Hurtado, Ramiro Barrantes, and Luis Rodriguez-Delfín for sample donations, as well as to all individuals who, by contributing their own samples, made this study possible. We thank Claudio Bravi for help with checking mutations and Eduardo Eizirik and three anonymous reviewers for their suggestions.

Received: August 15, 2007
Revised: November 13, 2007
Accepted: November 29, 2007
Published online: February 28, 2008

Web Resources

Accession numbers and URLs for data presented herein are as follows:
BEAST, http://beast.bio.ed.ac.uk/

Accession Numbers

The 51 sequences reported in this paper have been deposited in GenBank with the accession numbers EU095194–EU095251.

References


8 The American Journal of Human Genetics 82, 1–10, March 2008


63. Kelly, R.L. (2003). Maybe we do know when people first came to North America; and what does it mean if we do? Quaternary Int. 109–110, 133–145.


