Mitochondrial haplogroup M discovered in prehistoric North Americans

Ripan S. Malhi a,*, Brian M. Kemp b, Jason A. Eshleman b, Jerome Cybulski c, David Glenn Smith b,d, Scott Cousins e, Harold Harry e

a Department of Anthropology, University of Illinois Urbana-Champaign, IL 61801, USA
b Department of Anthropology, UC Davis, CA 95616, USA
c Canadian Museum of Civilization, Gatineau, Quebec, Canada
d California National Primate Research Center, CA, USA
e Canoe Creek Band, BC, Canada

Received 21 April 2006; received in revised form 8 July 2006; accepted 11 July 2006

Abstract

We analyzed two mid-Holocene (~5000 years before present) individuals from North America that belong to mitochondrial DNA (mtDNA) haplogroup M, a common type found in East Asia, but one that has never before been reported in ancient or living indigenous populations in the Americas. This study provides evidence that the founding migrants of the Americas exhibited greater genetic diversity than previously recognized, prompting us to reconsider the widely accepted five-founder model that posits that the Americas were colonized by only five founding mtDNA lineages. Additional genetic studies of prehistoric remains in the Americas are likely to reveal important insights into the early population history of Native Americans. However, the usefulness of this information will be tempered by the ability of researchers to distinguish novel founding lineages from contamination and, as such, we recommend strategies to successfully accomplish this goal.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: Native American; DNA; Population; Molecular dating; Peopling of the Americas; Ancient DNA

1. Introduction

Extensive geographic surveys of contemporary Native American genetic diversity have been completed in the past decade (Bortolini et al., 2003; Malhi et al., 2002; Merriwether et al., 1995b; Tarazona-Santos et al., 2001; Zegura et al., 2004). These studies all suggest that the Americas were founded by a small number of migrants from East Asia. In addition, global genetic surveys suggest that Native American populations exhibit greater amounts of inter-population genetic differentiation relative to populations in other continental regions (Rosenberg et al., 2002). This large amount of inter-population genetic differentiation combined with the paucity of sampling in many regions in the Americas, makes it likely that significant undocumented genetic structure still exists in the Americas. In this study we demonstrate the existence of this undocumented genetic structure with the discovery of two individuals from China Lake, British Columbia (Fig. 1) that exhibit a mitochondrial DNA (mtDNA) haplogroup never before reported in any prehistoric or living indigenous population in the Americas. Both were found in the same burial, dated to 4950 ± 170 14C years before present (ybp), and were believed to be related due to similar morphological characteristics (Cybulski personal communication). Both individuals belong to haplogroup M, which is widely distributed throughout Asia (Kivisild et al., 2002). These individuals account for two of three samples dating to approximately 5000 ybp studied from the Northern Plateau region in Northwestern North America that were compared to 3658 sequences from Native Americans widely distributed throughout the Americas. The
third sample dating to 4975 ± 40 14C ybp from Big Bar Lake, British Columbia (Fig. 1), exhibits a haplogroup A haplotype that is shared with contemporary indigenous individuals. The discovery of a new mitochondrial haplogroup in the Americas conflicts with the presumed five-founder model, which implies that all Native American mtDNA derives from only five lineages, the founding haplotypes of haplogroups A, B, C, D, and X (Eshleman et al., 2003). Our discovery demonstrates that a more genetically diverse group of migrants colonized the Americas than previously thought and supports the hypothesis that significant undocumented genetic diversity likely still remains in the Americas. While the documentation of additional founding lineages will provide a finer resolution understanding of the peopling process of the Americas, additional discoveries are contingent upon being able to distinguish authentic founding lineages from contamination in the genetic analysis of prehistoric remains. Here, we describe a strategy to help distinguish authentic results from contamination in prehistoric remains.

2. Materials and methods

2.1. Contamination control

Because aDNA occurs in low copy number and is highly degraded (Lindahl, 1993; Pääbo, 1990), its extraction and analysis is highly susceptible to contamination originating from modern sources. First, the surfaces of archaeological specimens can become inadvertently contaminated by anyone who has handled the material. To control for this potential source of contamination, the portions of the bone and teeth used for DNA extraction were decontaminated with highly concentrated bleach. Contaminating DNA can be removed from the surfaces of bones and teeth with a treatment of highly concentration bleach because aDNA is more resistant to the oxidant than is contaminating DNA (Kemp and Smith, 2005; Salamon et al., 2005; Watt, 2005). Second, false positives can arise in the aDNA laboratory from contaminated reagents or lab-ware, PCR carryover, and/or DNA originating from lab personnel. DNA-free reagents and lab-ware were used, when available, and negative extraction and PCR controls served as monitors for potential contamination generated during this study. Databases of mitochondrial DNA sequences of all researchers who enter the ancient DNA facilities are kept at both UC Davis and Trace Genetics to compare with ancient DNA results. This study was performed in laboratories dedicated to the study of aDNA at the University of California, Davis, and Trace Genetics Inc. (Richmond, CA), wherein precautions to minimize contamination are routinely practiced.

2.2. DNA extraction (UC Davis)

DNA was extracted from an upper left third molar and partial femur, representing two individuals from the China Lake site (EiRm 7:1 and EiRm 7:2-6), respectively. DNA was also extracted from a lower right third molar of an individual
from the Big Bar site (EhRk 4:1). Portions of the samples (Table 1) were carefully removed from the whole and submerged in 6% sodium hypochlorite (full strength Clorox bleach) for 15 min to remove any surface contamination (Kemp and Smith, 2005). The bleach was poured off and the samples rinsed with DNA-free ddH2O (Gibco) to remove any remaining bleach. Negative controls, to which no bone or tooth were added, accompanied the extractions and were subjected to all of the following steps. The samples were transferred to 15 mL conical tubes and demineralized by gentle rocking in 2 mL molecular grade 0.5 M EDTA, pH 8.0 (Gibco), for > 2 days at room temperature. Three milligrams of Proteinase K were added to the samples, followed by incubation at 65 °C for 4.5 h. DNA was first extracted by adding an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) to the EDTA, which was then vortexed briefly and centrifuged at 3100 rpm for 5 min. The aqueous phase was removed and subsequently extracted using phenol:chloroform:isoamyl alcohol (25:24:1), as just described. A third extraction was performed using an equal volume of chloroform:isoamyl alcohol (24:1), which was then vortexed briefly and centrifuged at 3100 rpm for 3 min. DNA was precipitated from the solution by adding one half volume of room temperature 5 M ammonium acetate and, to this combined volume, one volume of room temperature absolute isopropanol (Hänni et al., 1995), then storing the solution overnight at room temperature. The DNA was pelleted by centrifuging the tube for 30 min at 3100 rpm. The isopropanol was discarded and the samples air-dried for 15 min. The DNA was washed with 1 mL of 80% ethanol by vortexing, pelleted again by centrifuging for 30 min at 3100 rpm, and air-dried for 15 min after decanting the ethanol. The DNA was re-suspended in 300 µL of DNA-free ddH2O and silica extracted (Höss and Pääbo, 1993) using the Wizard PCR Preps DNA Purification System (Promega), following the manufacturer’s instructions except that: (1) the “Direct Purification Buffer” was not added and (2) DNA was finally eluted with 100 µL DNA-free ddH2O.

2.3. Haplogroup determination

The extracts were screened for the mutations that define Native American mitochondrial haplogroups A, B, C, and D (Forster et al., 1996; Schurr et al., 1990). PCR conditions and primer coordinates for each marker tested are provided in Table 2. Approximately 5 µL of each amplicon were separated on 6% polyacrylamide gels. The gels were stained with ethidium bromide and visualized under UV light to confirm successful PCR amplification for later restriction enzyme digestion or to identify the 9 bp deletion, definitive of haplogroup B. The remainder of each amplicon was digested for ~ 3 h at 37 °C with 1 U of the appropriate restriction enzyme. The digested products were separated and visualized as described above to identify the presence or absence of the restriction sites that characterize haplogroups A, C, and D. As both China Lake samples (EiRm 7:1 and EiRm 7:2-6) were determined to lack the markers definitive of haplogroups A, B, C, or D, they were subsequently screened for the AluI

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Sample</th>
<th>Extract</th>
<th>Material processed (g)</th>
<th>Haplogroup</th>
<th>HVRI a sequence</th>
<th>HVRII b sequence</th>
<th>M7 region sequence</th>
<th>M8 region sequence</th>
<th>M9 region sequence</th>
<th>M10 region sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>China Lake, EiRm 7:1, upper left</td>
<td>UC Davis</td>
<td>0.18</td>
<td>M</td>
<td>16093, 16213, 16223</td>
<td>N/A</td>
<td>CRS</td>
<td>004769</td>
<td>CRS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>China Lake, EiRm 7:1, lower left</td>
<td>Trace Genetics</td>
<td>0.25</td>
<td>M</td>
<td>16093, 16213, 16223</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>China Lake, EiRm 7:2-6</td>
<td>UC Davis</td>
<td>0.66</td>
<td>M</td>
<td>16093, 16213, 16223</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>China Lake, partial femur</td>
<td>UC Davis</td>
<td>0.52</td>
<td>M</td>
<td>16093, 16213, 16223</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Big Bar Lake, EhRk 4:1, lower</td>
<td>UC Davis</td>
<td>0.26</td>
<td>A</td>
<td>16111, 16223, 16278, 16290, 16319, 16362</td>
<td>00064, 00073, 00146, 00235, 00263, 00309.1 (C), 00309.2 (C), 00315.1 (C)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>China Lake, EiRm 7:1, upper left</td>
<td>UC Davis</td>
<td>0.33</td>
<td>M</td>
<td>16093, 16213, 16223</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

a Sequence read from nps 16011--16382.
b Sequence read from nps 00169--00330.
c Relative to the Cambridge Reference Sequence (Anderson et al., 1981; Andrews et al., 1999).
d Sequence read from nps 00135--00136.
e Sequence read from nps 004787.
site gain at np 10397, definitive of macro-haplogroup M, using the PCR conditions and primers described in Table 1. After confirmation that both of these individuals belong to haplogroup M, portions of their mitochondrial genome were screened for polymorphisms defining sub-haplogroups M7, M8, and M9 (Table 3), derived forms of haplogroup M found frequently in East Asian and Siberian populations (Starikovskaya et al., 2005; Yao et al., 2002). PCR conditions and primer coordinates for screening each of these markers are provided in Table 3a. The amplicons were directly sequenced as described below to screen the markers, except that they were sequenced only in the forward direction.

2.4. Hypervariable region sequencing

The first mitochondrial hypervariable region (HVRI) of the samples was sequenced from nps 16011—16382 in four small, overlapping fragments, using primers HVRI-1, 2, 3, 4 and Table 3a.

Table 3
Primers, with annealing temperatures, used (a) to screen markers definitive of subhaplogroups M7, M8, and M9 and (b) to sequence the hypervariable regions

<table>
<thead>
<tr>
<th>Target region</th>
<th>Defining marker</th>
<th>Primer</th>
<th>Primer coordinates</th>
<th>Annealing temperature (°C)</th>
<th>Primer citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Subhaplogroups M7, M8, and M9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mitochondrial sub-haplogroup M7</td>
<td>np 9824 (C)</td>
<td>9764F</td>
<td>09764—09785</td>
<td>66</td>
<td>This study</td>
</tr>
<tr>
<td>Mitochondrial sub-haplogroup M8</td>
<td>np 4715 (G)</td>
<td>4641F</td>
<td>04641—04661</td>
<td>62</td>
<td>This study</td>
</tr>
<tr>
<td>Mitochondrial sub-haplogroup M9</td>
<td>np 3394 (C)</td>
<td>3309F</td>
<td>3309—3329</td>
<td>64</td>
<td>This study</td>
</tr>
<tr>
<td>(b) Hypervariable regions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HVRI-1</td>
<td>15986F</td>
<td>15986—16010</td>
<td>62</td>
<td>Kemp et al., in press</td>
<td></td>
</tr>
<tr>
<td>HVRI-2</td>
<td>16513R</td>
<td>16152—16153</td>
<td>62</td>
<td>Kemp et al., in press</td>
<td></td>
</tr>
<tr>
<td>HVRI-3</td>
<td>16251R</td>
<td>16230—16251</td>
<td>58</td>
<td>Kemp et al., in press</td>
<td></td>
</tr>
<tr>
<td>HVRI-4</td>
<td>16355R</td>
<td>16331—16355</td>
<td>58</td>
<td>Kemp et al., in press</td>
<td></td>
</tr>
<tr>
<td>HVRII-1</td>
<td>00034F</td>
<td>00034—00058</td>
<td>62</td>
<td>This study</td>
<td></td>
</tr>
<tr>
<td>HVRII-2</td>
<td>00112F</td>
<td>00112—00135</td>
<td>62</td>
<td>This study</td>
<td></td>
</tr>
<tr>
<td>HVRII-3</td>
<td>00184F</td>
<td>00184—00208</td>
<td>62</td>
<td>This study</td>
<td></td>
</tr>
</tbody>
</table>

Thirty microliter PCR amplification reactions contained: 0.32 mM dNTPs, 1× PCR buffer, 1.5 mM MgCl₂, 2.4 mM primers, 0.3 U of platinum Taq (Invitrogen), and 3.0 μL of DNA template. PCR conditions were as follows: (1) 3 min denaturing at 94 °C, (2) 60 cycles of 15 s holds at 94 °C, the annealing temperature specified below, and 72 °C, and (3) a final 3 min extension period at 72 °C.

a Relative to the Cambridge Reference Sequence (Anderson et al., 1981; Andrews et al., 1999).

b Touch-down PCR (Don et al., 1991) used, decreasing the annealing temperature 0.1 °C after each cycle.
PCR conditions described in Table 3b. The second hypervariable region (HVRII) of the Big Bar sample (EhRk 4:1) was sequenced from nps 00059—00330 in three overlapping fragments using primers HVRII-1, 2, and 3 and PCR conditions described in Table 3b. The HVRII of the China Lake samples were not sequenced, as effort was instead directed at screening coding region SNPs in these samples, described above. To confirm amplification, 3—4 μL of PCR product were visualized on a polyacrylamide gel as described above. Single stranded DNA (e.g. excess primers) was destroyed by adding the remaining PCR product to a 60 μL Exol digestion cocktail that contained 40 U ExoI (New England Bio Labs) and 0.33 ExoI buffer. This reaction was incubated at 37 °C for 90 min, followed by an 80 °C hold for 20 min to denature the ExoI. The ExoI digested DNA was filtered through a Millipore 96-well Montage PCR Microfine Plate and re-suspended in 25 μL ddH2O. This product was submitted for direct sequencing to the Division of Biological Sciences (DBS) Automated DNA Sequencing Facility at the University of California, Davis. Sequencing was performed in both directions and sequences were read off both strands.

2.5. Confirmation of results at UC Davis

A second portion of the samples was removed from the whole (Table 1) and extracted as described above. Confirmation of haplogroup assignment was made by screening the Big Bar sample for the HaeIII site gain at np 663, definitive of mitochondrial haplogroup A. Both of the China Lake samples were screened, as described above, for the mutations definitive of macro-haplogroup M, and sub-haplogroups M7, M8, and M9. The sequences generated from the first extractions were replicated from these second extracts, as described above.

2.6. Confirmation of results at Trace Genetics

DNA was extracted from a third portion of the China Lake partial femur (EiRm 7:2-6) that was analyzed at UC Davis and from a second tooth of the other China Lake individual (EiRm 7:1), that was sent directly to Trace Genetics for independent extraction and analysis (Table 1). Both samples were screened for the AluI site gain at np 10397 and the HVRI sequence was generated for both individuals, as described above.

3. Results

The DNA sequences generated from PCR product exhibited a “clean” signal with no indication of double peaks due to contamination, DNA damage, or heteroplasmy. None of the DNA sequences generated in this study match those of individual researchers working in the ancient DNA laboratories. DNA sequences from multiple amplicons from multiple extractions performed in two independent laboratories from the two China Lake individuals exhibit the exact same mutations. Therefore it is unlikely that the results of this study are due to either DNA damage or contamination. Both China Lake individuals belong to haplogroup M, exhibiting the AluI site gain at np 10397. They do not, however, belong to haplogroup C, D, or sub-haplogroup M7, M8, or M9, all representing derived forms of haplogroup M. The individuals likely do not belong to haplogroup G as they lack HVRI mutations specific to its sub-clades. Both individuals exhibited identical HVRI sequences (Table 1). The Big Bar individual exhibited the HaeIII site gain at np 663 and, therefore, belongs to haplogroup A. The HVRI sequence exhibited by this individual confirms the haplogroup assignment and matches one Nuu-Chah-Nulth individual (Ward et al., 1991).

4. Discussion

Genetic studies of living populations have profoundly shaped the view of the peopling of the Americas. The present consensus from mtDNA, Y chromosome, and autosomal studies is that the Americas were first settled by a small number of migrants (Hey, 2005) likely from a single source population (Kolman et al., 1996; Merriwether et al., 1995a; Zegura et al., 2004). Many of these studies have contended that the Americas were colonized by a single haplotype from each of the accepted five founding haplogroups and as a result, the timing of the peopling of the Americas can be ascertained by estimating the amount of diversity within a haplogroup, given an assumed rate of change (Bonatto and Salzano, 1997; Forster et al., 1996; Torroni et al., 1993). DNA studies of prehistoric populations in the Americas reveal that a majority of the individuals belonged to one of the undisputed founding haplogroups in the Americas (Carlyle et al., 2000; Kaestle and Smith, 2001; Kemp et al., 2005; Malhi et al., 2002; Parr et al., 1996; Stone and Stoneking, 1998). In addition, the few early Holocene remains in the Americas whose mtDNA has been studied belong to the founding lineages of haplogroups B and C (Kaestle and Smith, 2001; Stone and Stoneking, 1996). In sum, these studies have reinforced the five-founder model.

However, some studies based on DNA analysis from contemporary populations suggest that more than one founding haplotype per haplogroup exists in the Americas (Derbeneva et al., 2002; Malhi et al., 2002; Rickards et al., 1999). In addition, analysis of DNA from an early-Holocene skeleton in Alaska (Kemp et al., 2006) convincingly demonstrates the existence of an additional founding lineage for haplogroup D in the Americas. In combination with the evidence from the present study, a clear trend is evident; as additional samples from contemporary and ancient populations are analyzed, assumptions made in earlier studies are being called into question, particularly, the five-founder model of the peopling of the Americas. Based on the studies mentioned above, we can no longer assume that a single haplotype founded each founding haplogroup in the Americas and as a result, genetic estimates for the timing of the peopling of the Americas based on this assumption are inaccurate.

The discovery of haplogroup M in the Americas is consistent with the hypothesis of a single colonization for the Americas since this haplogroup is found in Southern Siberia, the
presumed homeland of the ancestors of the Native Americans (Bonatto and Salzano, 1997; Merriwether et al., 1995a). However it also demonstrates the limitations of using genetic data solely from contemporary populations to infer the events and early population history of the Americas. Using genetic data from contemporary populations to infer early prehistoric demographic events is even less accurate when population history has been variable over time. Evidence from genetic, archaeological, and paleo-climatic studies indicate that the population history in Northern North America has in fact been unstable; for example, population bottlenecks caused by climatic events such as the Younger Dryas have occurred that were followed by population expansions from within Northern North America and/or from Siberia (Forster et al., 1996). Therefore data based on living populations in Northwestern North America might bias interpretations of population prehistory in the Americas.

This discovery broadens existing views of the colonization of the Americas. Researchers studying mtDNA of contemporary populations may have ignored evidence of additional founding haplogroups because it did not fit the prevailing five-founder model for the peopling of the Americas. In addition, studies of ancient DNA in the Americas may have misidentified authentic evidence of additional founder lineages as contamination and, as such, failed to report the results. We contend and demonstrate that with the proper protocols (Gilbert et al., 2005) and an awareness of the potential for the discovery of additional novel haplogroups, ancient DNA studies can provide a more accurate estimate of prehistoric genetic diversity and subsequently a more complete view of the peopling of the Americas than studies of contemporary populations alone.

Specifically, we suggest the following three-step strategy to help identify authentic novel founding haplotypes in the Americas. Step 1. If a prehistoric sample is identified as not belonging to haplogroups A, B, C, D, X or M, it should be determined to what haplogroup the sample belongs by screening diagnostic portions of the coding region. Step 2. Once such a sample is identified, the genetic characterization should be repeated in an independent laboratory. If possible this is best performed with an independent sample that does not pass through the first lab (e.g. a second tooth). Step 3. Once the reliability of a haplotype designation of a sample has been established an assessment to determine the authenticity of the haplotype is needed. We suggest the following to help determine the authenticity of the results. In accord with the recommendations of Gilbert and colleagues (Gilbert et al., 2005), the results should make phylogenetic sense, in this case with the prevailing theories for the peopling of the Americas. Specifically, if a haplogroup is found that has not been identified in East Asia, greater caution should be taken and additional proof will be needed to confirm such results. In addition, if the haplotype of the sample is similar to haplotypes exhibited by researchers working on the sample, additional confirmation of the authenticity of the results will be needed. To help provide this additional proof we suggest using a third independent laboratory to analyze the sample and again confirm the results. We have identified the existence of samples identified as “others” or not belonging to haplogroups A, B, C, D, or X in the literature (González-Oliver et al., 2001; Kaestle and Smith, 2001; Lalueza et al., 1997; Parr et al., 1996; Shimada et al., 2004; Stone and Stoneking, 1998). We suggest that the three-step strategy outlined above begin with these samples if possible.

With the maturity of the ancient DNA field, an appreciation for the prevalence of contamination and biochemical processes that can plague ancient DNA results has emerged (Gilbert et al., 2005). The diversity in size, scope, and resources available to various ancient DNA laboratories dictates how each research group will efficiently and creatively overcome these problems and prove the authenticity of their results. Such diversity in laboratories precludes the enforcement of authoritarian regulations for ancient DNA research. Therefore, the above suggestions are intended as guidelines to assist ancient DNA researchers gain confidence in and help prove the authenticity of their results.

Acknowledgments

We are indebted to the members of the Canoe Creek, Soda Creek, and Dog Creek Bands who allowed the DNA testing of their ancestors. We would also like to thank Toomas Kivi-sild for consultation on what genetic markers to screen and Graciela Cabana for valuable comments on this manuscript.

References


