Mitochondrial-DNA Phylogenetic Information and the Reconstruction of Human Population History: The South American Case

María Bárbara Postillone[1] *and S. Ivan Perez*[2]

ABSTRACT

Mitochondrial DNA (mtDNA) sequences are becoming increasingly important in the study of human population history. Here, we explore the differences in the amount of information of different mtDNA regions and their utility for the reconstruction of South American population history. We analyzed six data sets comprising 259 mtDNA sequences from South America: Complete mtDNA, Coding, Control, hypervariable region I (HVRI), Control plus cytochrome b (cyt*b*), and cyt*b* plus 12S plus 16S. The amount of information in each data set was estimated employing several site-by-site and haplotype-based statistics, distances among sequences, neighbor-joining trees, distances among the estimated trees, Bayesian skyline plots, and phylogenetic informativeness profiles. The different mtDNA data sets have different amounts of information to reconstruct demographic events and phylogenetic trees with confidence. Whereas HVRI is not suitable for phylogenetic reconstruction of ancient clades, this region, as well as the Control data set, displays information for the demographic reconstruction during the Holocene period, probably because of the high rate of mutation of these regions. As expected, the Complete mtDNA and Coding data sets, displaying slower rates of mutation, present suitable information to estimate the founding subhaplogroups that populated South America and for the reconstruction of ancient demographic events. Our results point out the importance of evaluating the utility of different DNA regions to respond to different questions and problems in the human population studies, mainly considering the time scale of the phenomenon and the informativeness of the molecular region in a particular geographical area.

Molecular data have become increasingly important during the last century for the study of human population histories. In recent years these studies have been revolutionized by the analysis of genome-scale data sets (Macaulay et al. 2005; Li and Durbin 2011; Rasmussen et al. 2011; Mallick et al. 2016). In particular, the analysis of mitochondrial DNA (mtDNA) genomes has had a profound impact on studies of the evolution of human populations at global and regional scales. In this context, our understanding of the early dispersion and demographic history of the humans who peopled the Americas, mainly South America, has been

[1]Departamento de Ciencias Naturales y Antropológicas, Universidad Maimonides, Centro de Estudios Biomédicos, Biotecnológicos, Ambientales, y de Diagnóstico (CEBBAD), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina.

[2]División Antropología, Facultad de Ciencias Naturales y Museo, Universidad Nacional de La Plata, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), La Plata, Argentina.

*Correspondence to: S. Ivan Perez, División Antropología, Facultad de Ciencias Naturales y Museo, Universidad Nacional de La Plata, 122 and 60, La Plata 1900, Argentina. E-mail: ivanperezmorea@gmail.com.

KEYWORDS: MOLECULAR DISTANCES, SUBSTITUTION RATES, BAYESIAN METHODS, POPULATION DYNAMICS.

Human Biology, Summer 2017, v. 89, no. 3, pp. 229–250. doi: 10.13110/humanbiology.89.3.05.

greatly impacted by these analyses (Fagundes et al. 2008b; Perego et al. 2010; Bodner et al. 2012; de Saint Pierre et al. 2012b; Llamas et al. 2016).

Despite many studies using genomic data, migratory and demographic topics are also addressed using mtDNA control region sequences and the nonrecombining region of the Y chromosome (Vigilant et al. 1991; Rubicz et al. 2007; Bisso-Machado et al. 2012; Wallace 2015; Poznik et al. 2016). In particular, several works employ sequences from hypervariable regions I and II (HVRI and HVRII) to explore hypotheses about the origin, migration route, and demographic history of aboriginal human populations that inhabited South America (e.g., Moraga et al. 2000; O'Rourke and Raff 2010; Kemp and Schurr 2010; de Saint Pierre et al. 2012a). More recently, studies exploring human evolution in the subcontinent have analyzed ancient DNA, sequencing mainly HVRI (e.g., Carnese et al. 2010; Fehren-Schmitz et al. 2011, 2014; Mendisco et al. 2014; Postillone et al. 2017). These studies have generated a large comparative database of HVRI and HVRII, as well as a more restricted sample of complete mtDNA genomic data. Therefore, a critical issue is to understand the utility of these different DNA regions or data sets for the study of past human populations in geographical regions as South America.

This is particularly important if we consider that different mtDNA regions display differences in their substitution rate (Endicott and Ho 2008; Soares et al. 2009) and therefore differences in the amount of phylogenetic information (Townsend 2007; Dornburg et al. 2014). The relevant question here is which mtDNA region is more informative about a process or event for a given time interval in the past. Whereas in interspecies studies researchers have long discussed the utility of different mitochondrial and nuclear genomic regions for phylogenetic inference (e.g., Goldman 1998; Graybeal 1998; Yang 1998; Cotton and Wilkinson 2008; Camargo et al. 2012; Dornburg et al. 2014), this problem has been scarcely explored in depth in human population studies (see Non et al. 2007). This is striking because, for example, the different human mtDNA regions are known to differ in their substitution rate and level of saturation (Soares et al. 2009). At present, there is no robust theory predicting the power of DNA regions for a particular time in the past, although the amount of information of a molecular sequence can be empirically quantified (Townsend 2007).

Here, we empirically explored differences in the amount of information of different mtDNA regions and their utility for the reconstruction of human population history. A previous study explored the utility of different mitochondrial genomic regions for phylogenetic inference in the Old World (Non et al. 2007), suggesting that the mtDNA control region has limited utility for the study of the human population history. We statistically address this problem employing samples from South America, a model system that displays a recent human peopling (ca. 15,000–20,000 years ago; Fagundes et al. 2008b; Llamas et al. 2016). Although we explored American sequences, the objective of our work was methodological, having as its goal to understand the performance of different data sets and not the examination of the peopling of America. Specifically, in this work we addressed the impact of variability and informativeness of different mtDNA regions, some of them previously employed in South American studies, on genealogical (i.e., individual phylogeny) and demographic reconstructions. Unlike previous studies, we explicitly explored the impact of differences in the amount of information of different mtDNA regions in the population dynamics estimation at a continental scale, in a model system with a more recent human peopling than previously used (Non et al. 2007). Therefore, we expected our results to differ from previous works. We explored this problem using complete mtDNA genomes, multivariate statistics, and Bayesian methods. We also discuss the best way to employ mtDNA data sets in human population studies. We expect our study to allow for a better use of the large quantity of mtDNA data and to improve genealogical and demographic inferences.

Materials and Methods

We obtained 259 DNA sequences comprising the whole mitochondrial genome (complete mtDNA) from GenBank and previous publications (Hartmann et al. 2009; Perego et al. 2009, 2010; Bodner et al. 2012; de Saint Pierre et al. 2012b; Llamas et al. 2016). Accession numbers for the mitochondrial sequences from South America are shown in

Supplementary Table S1. These sequences come from modern and prehistoric humans of South America, particularly from south-central Andes (n = 123; Hartmann et al. 2009; Perego et al. 2009, 2010; Bodner et al. 2012; de Saint Pierre et al. 2012b; Llamas et al. 2016), northeastern South America (n = 17; Hartmann et al. 2009; Perego et al. 2009, 2010; Bodner et al. 2012; Llamas et al. 2016), and Pampa-Patagonia (n = 119; Perego et al. 2009; 2010; Bodner et al. 2012; de Saint Pierre et al. 2012b; Llamas et al. 2016).

The sequences of mitochondrial genomes were aligned in the software MAFFT (version 7.012b; Katoh and Standley 2013), using the algorithm FFT-NS-2, and manually verified with the software BioEdit (version 7.0.0; Hall 2004). After alignment, the complete mtDNA data set was divided in two new data sets, the Control and Coding region data sets (Figure 1), which previously were used in studies about South American human evolution (e.g., de Saint Pierre et al. 2012a; Fagundes et al. 2008b). The Control data set comprises 1,122 base pairs (bp), including positions 1–576 and 16024–16569; the Coding data set comprises 15,391 bp, between the positions 577 and 16023 (with two intercalated noncoding segments of 25 and 31 bp; Andrews et al. 1999). We also analyzed only the 343 bp of the HVRI (Figure 1), from position 16024 to position 16365. Finally, we generated two novel data sets. The first novel data set, Control–cyt*b*, includes the Control region and the 1,140 bp from the cytochrome b (cyt*b*) gene (Figure 1; between positions 14747 and 15887). The second novel data set, cyt*b*–12S–16S, includes cyt*b* plus the 12S rRNA (between positions 648 and 1601) and 16S rRNA (between positions 1671 and 3229; Figure 1). We chose these data sets for two reasons: (a) the control, HVRI, and coding regions are widely used in studies of human population and evolution in South America (de Saint Pierre et al. 2012a; Fagundes et al. 2008b; Perez et al. 2016) and represent regions with marked differences in substitution rate and relative number of mutations (Soares et al. 2009); (b) conversely, the cyt*b*, 12S rRNA, and 16S rRNA regions, as well as the control region plus cyt*b* gene, are widely used in phylogenetic and biogeographic studies, mainly in analyses involving closely related species and subspecies (e.g., Marín et al. 2008; Lynch Alfaro et al. 2012; Metcalf et al. 2016), and therefore could be useful for human population studies. All data sets were used in the subsequent analyses.

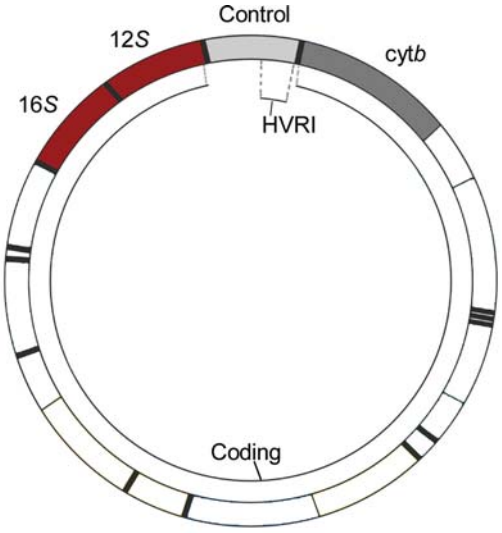


FIGURE 1. Map of the mitochondrial genome highlighting the subregions studied: Coding, Control, hypervariable region I (HVRI), cytochrome b (cyt*b*), 12S, and 16S.

We first explored the differences among the molecular mtDNA data sets by estimating the similarity between distance matrices and genealogical trees obtained using each data set. We estimated the maximum composite likelihood distance and employed the neighbor-joining method, implemented in MEGA (version 7.0.20; Kumar et al. 2016), to reconstruct the phylogenies. We employed the neighbor-joining method because it is more efficient for large data sets than the alternatives (Tamura et al. 2004). We then compared the matrices of distances and the genealogical trees obtained by calculating the matrix correlation (Smouse et al. 1986; Legendre and Legendre 1998) for the former and the Robinson-Foulds (RF) distance (Robinson and Foulds 1981; Kuhner and Felsenstein 1994) for the latter. To estimate the matrix correlation, the matrices were unfolded into vectors and a simple correlation was calculated. We also estimated dispersion plots, similar to empirical saturation plots (Graybeal 1994), to graphically evaluate the utility of the mtDNA regions to reconstruct a phylogeny. The RF distance or topological congruence between trees was calculated as the number of internal branches observed in one phylogeny but not in the other. The matrix of RF distances was analyzed using two multivariate analyses: the unweighted pair group method with arithmetic mean (UPGMA) cluster and the nonmetric multidimensional scaling (nmMDS). RF distances were calculated in the Tree distance 3.695 program of the PHYLIP package (Felsenstein 2005), whereas

the matrix correlations and multivariate analyzes were performed in the software PAST (version 3.0; Hammer et al. 2001).

We further described the differences among the molecular mtDNA data sets by calculating several site-by-site and haplotype-based statistics (Rozas 2009). For the site-by-site analyses, we calculated the number of variable positions or segregating sites (S) in each data set. Because this statistic is sensitive to the number of sites in the sequences (N_s; Rozas 2009), we calculated the number of segregating sites on the total number of sites in the sequence (S/N_s). We also estimated the mean number of nucleotide differences between sequences (k) and the average number of nucleotide differences per site or the nucleotide diversity (π), defined as k divided by the number of sites in the sequence, excluding the sites with alignment gaps (Rozas 2009). Finally, we estimated the number of different DNA sequences or haplotypes (h) and the mean of the haplotype proportions or haplotype diversity (H; Rozas 2009). Because the definition of haplotypes is related to the number of sites in the sequence (N_s), we estimated the number of haplotypes divided by the number of site in the sequences (h/N_s).

Second, we used the six mtDNA data set to estimate the demographic trajectories of the South American populations employing the Bayesian skyline plot (BSP) method (Drummond et al. 2005) implemented in the software BEAST (version 1.6.1; Drummond and Rambaut 2007). The BSP method uses the shape of a genealogy estimated with molecular data to reconstruct the demographic dynamics of a population in the past (Drummond et al. 2005; Ho and Shapiro 2011). The method simultaneously estimates genealogy, coalescence time, and population size through time using a Markov chain Monte Carlo sampling procedure (Drummond et al. 2005). We used the BEAUti program to set the parameters of the analyses for the data sets. The models of substitution for each mtDNA region or data set were inferred with the Akaike information criterion with correction for sample size implemented in the software jModelTest (version 2.1.10; Guindon and Gascuel 2003; Darriba et al. 2012). The HKY+I+G model of mtDNA sequence substitution displayed the best fit for all data sets. The sequences were analyzed under an uncorrelated lognormal relaxed molecular clock model, and we set the tree priors as a coalescent Bayesian skyline. The number of generations was established at 50,000,000 and the sample frequency at 5,000 for the Markov chain Monte Carlo sampling. We employed two widely used substitution rates in South American studies (Fagundes et al. 2008b; de Saint Pierre et al. 2012a; Perez et al. 2016): 3.02E–7 substitutions per site per year (Endicott and Ho 2008) for the control, HVRI, and control plus cyt*b* regions, and 1.26E-08 substitutions per site per year (Fagundes et al. 2008b) for complete mtDNA, cyt*b* plus 12S plus 16S, and coding region. The use of ancient DNA can influence our demographic estimations, so we alternatively included and excluded the ancient sequences in the analyses. The BSPs were reconstructed using the estimated genealogies in the software Tracer (version 1.5; Rambaut and Drummond 2007). Tracer also was used to test for the convergence in the parameters of the Bayesian analyses.

Finally, we explicitly tested for the phylogenetic informativeness of the different data sets. We employed the online application PhyDesign (López-Giráldez and Townsend 2011), which implements the Townsend (2007) phylogenetic informativeness profile. This method provides a quantitative measurement of the utility of a molecular region or data set to reconstruct a phylogeny at different times in the past (Townsend 2007; López-Giráldez and Townsend 2011). Phylogenetic informativeness relates nucleotide saturation in a molecular region to the estimated divergence time (Townsend 2007; Dornburg et al. 2014). Specifically, the informativeness profile is estimated based on the ratio of the observed rate of substitution to the optimal rate of substitution for genealogical inference at different times in the past (Dornburg et al. 2014).

Results

The relationships among mtDNA lineages estimated by the neighbor-joining trees are similar for the Complete mtDNA, Control, Control–cyt*b*, and Coding data sets (Figure 2). The complete mtDNA tree shows that haplogroup B, including the subhaplogroups B2 and B2i, is related to subhaplogroup A2. Haplogroup D displays large variability and shows four well-defined clades or subhaplogroups (note that we use the terms

haplogroup, *subhaplogroup*, and *haplotype* only as a convenience in this work; for a broader discussion of the terminology, see Kemp and Schurr 2010), all of them monophyletic (D1, D4h3, D1g, and D1j). The tree also displays large diversity in haplogroup C, showing the monophyly of subhaplogroups C1b, C1c, and C1d. The trees based on the Control and Control–cyt*b* data sets display slight differences from the Complete mtDNA tree. In particular, in these data sets some subhaplogroups, such as D1 and D4h3, are not monophyletic. Conversely, the trees based on HVRI and cyt*b*–12S–16S data sets display large differences. In particular, the HVRI shows that almost all the subhaplogroups are not monophyletic, displaying different relationships among them. Finally, the tree based on the Coding data set displays some similarities with the Complete mtDNA tree, showing only subhaplogroups D1g and B2 as polyphyletic.

The pattern of similarity between the data sets can be better observed in the nmMDS (stress = 0) and UPGMA results (Figure 3). These results confirm that the Control and Control–cyt*b* data sets generate trees more similar to the Complete mtDNA tree, whereas the HVRI and cyt*b*–12S–16S trees display large differences compared to each other and to the Complete mtDNA tree. We also explored the variation in a data set displaying sequences from all America, including the 259 South American sequences and 174 sequences from North America (Supplementary Table S2; Just et al. 2008; Hartmann et al. 2009; Perego et al. 2009, 2010; Achilli et al. 2013; Llamas et al. 2016), to compare with the South American case. The results show a pattern of differences among the mtDNA data sets for the American sample (Supplementary Figure S1) that is similar to the pattern observed for the South American sample, suggesting that our results could be generalized to the entire peopling of America. When we observed the correlations between distance matrices (Table 1), the Control, Control–cyt*b*, and Coding data sets show correlations between 0.90 and 0.98, whereas the cyt*b*–12S–16S and HVRI data sets display correlation values of 0.84 and 0.87, respectively. Despite the global similarities, the dispersion plot showed in Figure 4 graphically suggests that the HVRI and Control data sets change at a much higher rate than do the Complete mtDNA, Coding, and cyt*b*–12S–16S data sets.

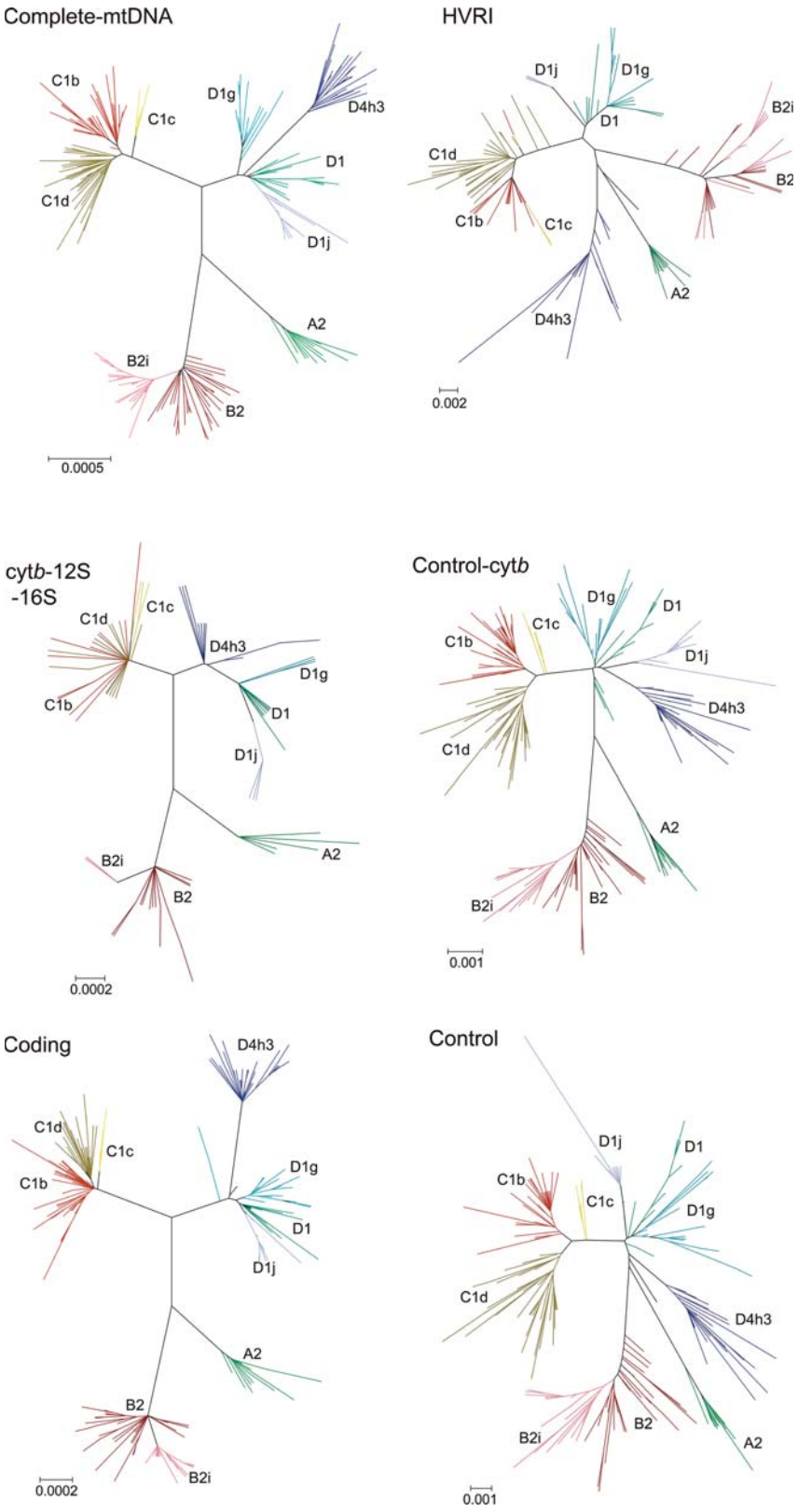


FIGURE 2. Phylogenetic trees estimated employing the neighbor-joining method and the maximum composite likelihood distance for each data set.

Table 1. Correlation and Significance Values between mtDNA Regions for All the South American Haplogroups

	Complete mtDNA	Coding	Control	HVRI	Control-cyt*b*	cyt*b*-12S-16S
Complete-mtDNA	—	0.000	0.000	0.000	0.000	0.000
Coding	0.982	—	0.000	0.000	0.000	0.000
Control	0.895	0.798	—	0.000	0.000	0.000
HVRI	0.875	0.812	0.897	—	0.000	0.000
Control-cyt*b*	0.923	0.852	0.967	0.874	—	0.000
cyt*b*-12S-16S	0.841	0.841	0.717	0.665	0.838	—

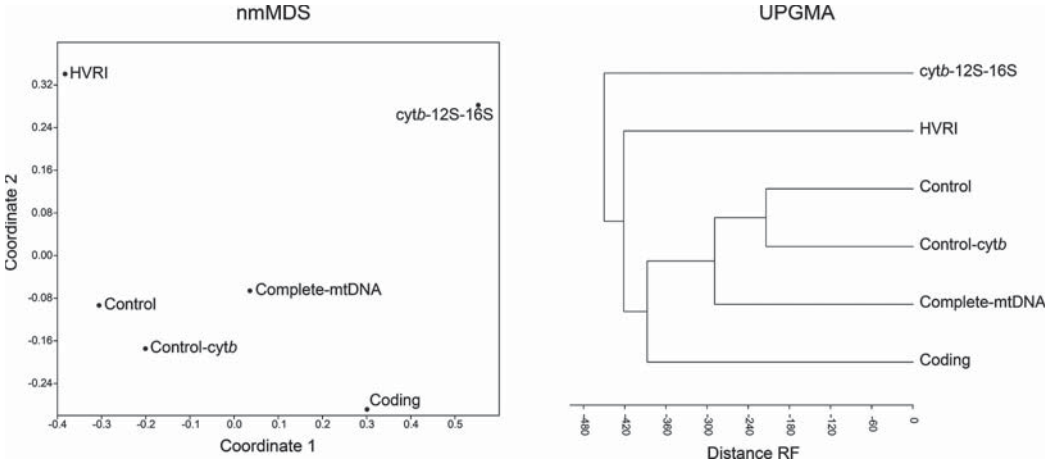


FIGURE 3. Nonmetric multidimensional scaling (nmMDS) and unweighted pair group method with arithmetic mean (UPGMA) cluster analyses of topological differences among the phylogenetic trees depicted in Figure 2. The analyses used the Robinson-Foulds (RF) distance to estimate the similarity between the data sets.

A similar pattern of differences among the data sets arises from the analyses of site-by-site and haplotype variation (Table 2). The number of segregating sites was relatively large in the Complete mtDNA and Coding data sets, intermediate in the Control and HVRI data sets, and low in Control–cyt*b* and cyt*b*–12S–16S data sets. However, the Control and HVRI data sets display the largest values of proportional segregating sites (S/N_s). We observed the same pattern when we calculated the mean number of nucleotide differences between sequences (k) and the nucleotide diversity or the proportion of k on N_s (Table 2). When we consider the haplotype analyses, we also observed a pattern similar to the one we detected in the site-by-site analyses: whereas the Complete mtDNA and Coding data sets display the largest number of haplotypes (h), the Control and HVRI data sets display the largest values of proportional number of haplotypes (h/N_s).

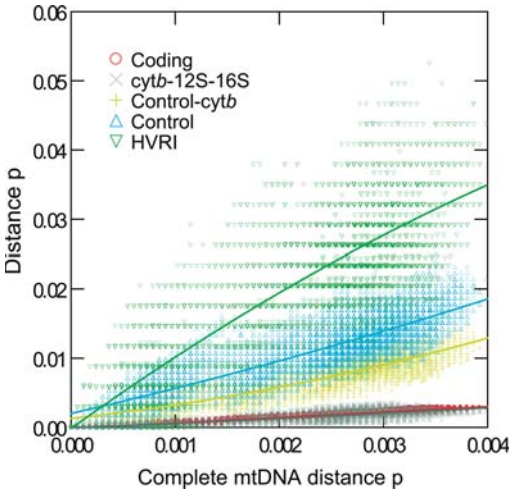


FIGURE 4. Relationships between the Complete mtDNA and other mtDNA data sets.

In concordance with the pattern of similarities in sequence variability and rate of substitution, the BSP result displays a clear pattern of differences between data sets. In particular, the noncoding data sets (Control and HVRI) show that the female effective population size was constant from the initial peopling until ca. 7,500 years ago and increased between 7,500 and 4,000 years ago (Figure 5). The Control–cyt*b* data set result displays a similar pattern of demographic change. Conversely, the Complete mtDNA and Coding data sets show more complex demographic dynamics, with an additional increase in the population size ca. 17,000 years BP (Figure 5). These data sets also display a later increase in population size ca. 5,500 years BP, suggesting that the largest data sets present more

Table 2. Analysis of Sequence Polymorphism for the mtDNA Data Sets of All South American Haplogroups

Datasets	Sample size		Nucleotide statistics				Haplotype statistics			
	N	N_s	*S*	S/N_s	*k*	π	*h*	*h/h-max*	h/N_s	*H*
Complete-mtDNA	259	16593	643	0.039	32.066	0.002	233.000	1.000	0.014	0.999
Coding	259	15450	500	0.032	23.562	0.002	212.000	0.910	0.014	0.998
Control	259	1144	143	0.125	8.504	0.009	169.000	0.725	0.148	0.994
HVRI	259	348	74	0.213	4.688	0.016	108.000	0.464	0.310	0.964
Control-cyt*b*	259	2280	198	0.087	11.641	0.006	189.000	0.811	0.083	0.996
cyt*b*-12S-16S	259	3648	97	0.027	5.157	0.001	90.000	0.386	0.025	0.960

N: number of sequences; N_s: number of sites; *S*: number of segregating sites or number of polymorphic sites; S/N_s: the proportion of *S* on N_s; *k*: mean number of nucleotide differences between sequences; π: nucleotide diversity or the proportion of *k* on *Ns*; *h*: number of haplotypes or the different DNA sequences; *h/h-max*: the proportion of *h* on the maximum value of *h*; h/N_s: the proportion of *h* on the number of sites; *H*: haplotype diversity.

information for demographic studies. The results shown that the exclusion of ancient DNA in the BSP analyses does not influence the demographic estimations (Supplementary Figure S2), although they represent more than the 30% of the samples used in the database. Perhaps this may relate to the fact that most of them correspond to the late Holocene (300–700 years BP; Llamas et al. 2016).

Figure 6 displays the results of the informativeness profiles for each data set. The Complete mtDNA and Coding data sets show the largest values of net informativeness in the more distant

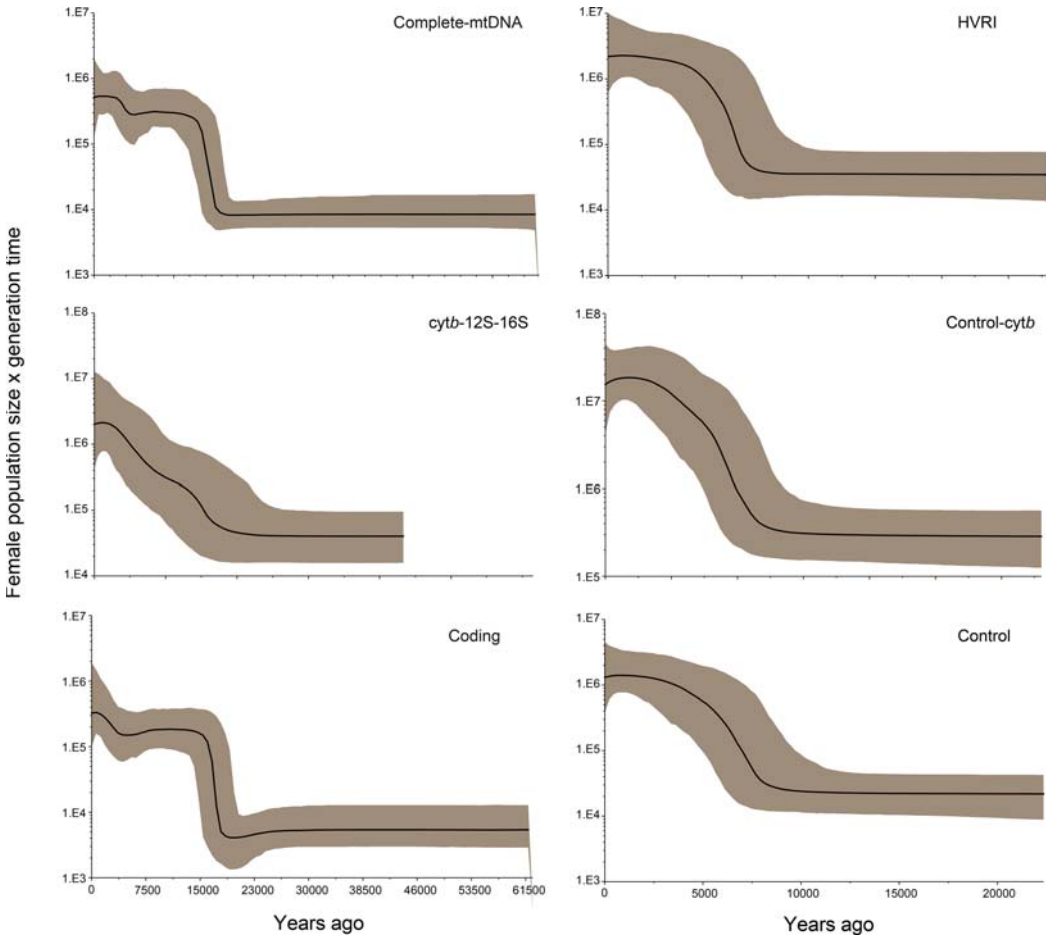


FIGURE 5. Demographic changes in South America estimated based on the molecular sequences of the six data sets studied. The times are scaled.

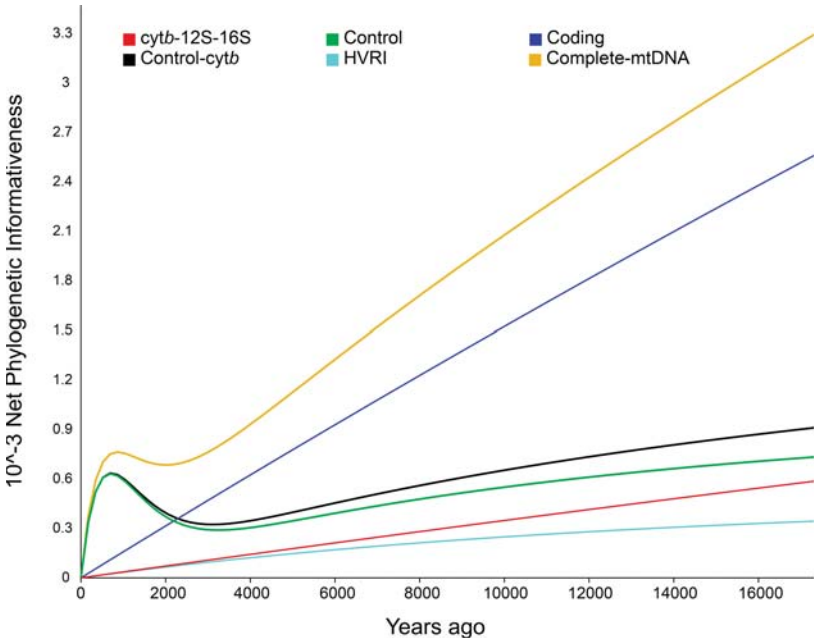


FIGURE 6. Net phylogenetic informativeness profiles of the six data sets studied.

past, ca. 17,000 years BP, decaying quickly from that time until the present. The Control and Control–cyt*b* data sets, together with the Complete mtDNA, display the largest informativeness during the last 2,000 years (Figure 6). The pattern of differences among the data sets observed in the informativeness profiles look similar to that observed in BSP results and in the other analyses, suggesting the importance of considering informativeness estimations in reconstructing population dynamics at different times in the past.

Discussion

Previous work in the Old World has shown that phylogenetic trees based on the mtDNA control region are poorly resolved (Non et al. 2007). In the same way, the works performed in America have suggested that data sets such as the mtDNA control region—including the HVRI—have limited utility for studying human population history in the continent (Tamm et al. 2007; Fagundes et al. 2008b; de Saint Pierre et al. 2012b). In particular, these studies have pointed out that the lineage genealogy or phylogeny of the mtDNA variants in America could not be reconstructed using the HVRI or control region, making it difficult to define the founding clades or subhaplogroups that arrived on the continent during the early peopling (Fagundes et al. 2008a; de Saint Pierre et al. 2012b). More generally, it was suggested that the time and mode of the early American peopling cannot be reconstructed with confidence when the HVRI or control region is used because they both display limited information—mainly due to the small size of these sequences—and high frequency of recurrent mutations (Tamm et al. 2007; Non et al. 2007; Fagundes et al. 2008b). Conversely, previous studies pointed out that the complete coding region and the complete mtDNA should be more appropriate for genealogical reconstruction and the study of population history in America (Tamm et al. 2007; Fagundes et al. 2008a, 2008b; de Saint Pierre et al. 2012b). Many of these studies assume that these large sequences are most useful than the noncoding ones because they present changes that are unique and irreversible (but see Non et al. 2007). This assumption is in line with more general discussions in studies of phylogenetic inference about the utility of unique and irreversible changes versus characters that present recurrent states (Townsend 2007; Yang 1998). To the best of our knowledge, although many statistics have been generated to measure the utility of different sequence data in phylogenetic inference, there have been few systematic explorations of this problem in anthropological studies (Non et al. 2007), and there has been no systematic research at small temporal scales, such as the observed in the South American peopling.

Contrary to suggestions of previous studies, our results display a complex scenario, where different mtDNA regions have variable usefulness for the study of different problems. In particular, the results suggest that sequence length is not the relevant dimension to discuss the utility of the different data sets for the study of human population history in South America. We demonstrate that data sets with different sequence lengths display very similar distance matrices and almost identical phylogenetic trees (Table 1, Figures 2 and 3). The Control and Control–cyt*b* data sets generated trees more similar to the Complete mtDNA data set tree. When we compare the distance matrices, we can observe that the Coding, Control, and Control–cyt*b* data sets display a pattern of differences among cases similar to that of the Complete mtDNA data set. These results suggest that the control region has sufficient information to reconstruct the lineage

genealogy or phylogeny of the mtDNA variants in South America. This result contrasts with the most general expectation that the length of sequence is important in experimental design (Goldman 1998). Conversely, the tree and distance matrix obtained with the HVRI data set display important differences with the other data sets. Therefore, we argue that this mtDNA region does not provide sufficient information to reconstruct a reliable phylogenetic tree and define the founding subhaplogroups that peopled South America, as has been previously suggested (Fagundes et al. 2008a; de Saint Pierre et al. 2012b).

However, the problem discussed here is not simple. As we point out above, previous studies also suggested that the high mutation and/or substitution rate of the noncoding HVRI and full control region could be a problem for the study of ancestral population dynamics in America, particularly because of the high frequency of recurrent mutations (Tamm et al. 2007; Fagundes et al. 2008b). Nevertheless, our results suggest that HVRI and the control region display a convenient rate of mutation for studying details of the Holocene demographic change in South America (Figures 4 and 5). In particular, when we analyze these molecular data sets, and when we explore the Control–cyt*b* data set, the BSP analyses show a quick population increase ca. 7,000–6,000 years ago, which is similar to the results shown for South America by previous works employing molecular and archaeological data (Marquet et al. 2012; Goldberg et al. 2016; Perez et al. 2016). The results also show the impact of the European colonization during the last 500 years on the mtDNA molecular variation. Conversely, the Coding and Complete mtDNA data sets show less detail for the dynamics of population increase after ca. 7,000–6,000 years ago. The analyses of these data sets also suggest that this event was more recent, ca. 5,000 years ago, than the interpretation based on the BSP analysis of the noncoding sequences. Although this temporal difference in the time of the Holocene population increase could be related to uncertainties in the estimation of the substitution rate employed (Llamas et al. 2016), the BSP based on Coding and Complete mtDNA data sets shows another important difference: an additional population increase event ca. 19,000 years ago. This population increase has been described previously in DNA studies employing different data (Fagundes et al. 2008a, 2008b; Kitchen et al. 2008; Llamas et al. 2016; Poznik et al. 2016). However, it is important to remark that the estimated mean time of this event varies between 15,000 and 25,000 years ago, depending on the substitution rate employed by each study (Llamas et al. 2016).

All the results obtained in this work make sense if we consider the time scale of South American peopling and the values of sequence informativeness. The informativeness profiles show that only three data sets, Complete mtDNA, Control, and Control–cyt*b*, display relatively high values of net informativeness during the last 2,000 years (Figure 6). In particular, the Control and Control–cyt*b* data sets display relatively high values of informativeness mainly in times close to the present, whereas the Complete mtDNA data displays high values in this period and during the initial divergence time ca. 17,000 years ago. The Coding data set also displays high values of informativeness for the earliest times analyzed. Both the Coding and Complete mtDNA data sets are very useful to study relatively ancient processes and events in America, but this large quantity of information about ancient events could cause problems in the estimation of more recent events, as suggested by our BSP results (Figure 5). These results, together with our other results, do not support the opinion that the mtDNA control region presents limited information for genealogical reconstruction and for the study of population history in America (Tamm et al. 2007; Fagundes et al. 2008a, 2008b). Conversely, the HVRI and the cyt*b*–12S–16S data sets show low values of informativeness over time, corroborating its previously hypothesized limited value for the American human population studies. Our results also suggest that mtDNA regions widely used previously to explore phylogenetic relationships among closely related species or subspecies, such as cyt*b*, 12S, and 16S, do not display sufficient information to investigate the population processes in the time frame of South America human evolution, or in similar problems such as the peopling of the entire America, probably related to the fact that a relatively small number of mutations can be observed in this mtDNA region for our data set (Table 2, Figure 6).

In summary, we demonstrate the complex behavior of the different mtDNA data sets used in previous studies. They have different degrees

of information to reconstruct phylogenetic trees with reliability and infer without error the founding monophyletic clades or subhaplogroups of America. Our results indicate that HVRI is not suitable for phylogenetic reconstruction of ancient clades because this mtDNA region does not have sufficient information of events that occurred in the distant past, that is, during the Pleistocene. Moreover, the HVRI and Control data sets display information for the demographic reconstruction during the Holocene period, where the high rate of mutation seems to be a valuable characteristic. Conversely, the Complete mtDNA, Coding, Control, and Control–cyt*b* data sets display sufficient information for phylogenetic reconstruction in this geographic region during the time span of the human peopling. As expected, the complete mtDNA and coding region, displaying slower rates of mutation, present better information for the reconstruction of ancient demographic events.

Final Remarks

During the last decades it became obvious that inference of evolutionary patterns from molecular sequences is a statistical problem and requires the use of experimental design (Goldman 1998; Townsend 2007; Yang and Rannala 2012). Nevertheless, human population studies have paid little attention to this problem. Previous studies have followed the "empirical folklore" and a few world-scale studies (Non et al. 2007), such as occurred decades ago in the interspecies works (Goldman 1998), about the best molecular sequences to use and the sampling of individuals for population analyses. They have not systematically employed formal methods to explore the utility of different sequences or to sample individuals to test different population problems. Our results indicate the importance of evaluating the utility of different DNA regions to address different questions and problems in human population studies, mainly considering the time scale of the phenomenon and the informativeness of the molecular region in a particular geographical area.

In this study, we explored the utility of sequences for the reconstruction of human population history. However, the discussions about experimental design in molecular studies go beyond sequence informativeness (Goldman 1998; Graybeal 1998; Geuten et al. 2007). Therefore, future studies are needed to explore the impact of the use of multiple sequences and the sampling of individuals from different times and geographical areas in the reconstruction of human population history. This is particularly relevant if we consider that the mtDNA databases for regions such as HVRI or the control region are considerably larger than the sample of complete mtDNA genomes.

ACKNOWLEDGMENTS

We are sincerely grateful to Sergio Furtado dos Reis (UNICAMP, Brazil) for useful discussions about experimental design in phylogenetic studies. We also thank two anonymous reviewers for their comments that have greatly improved this article.

Received 17 July 2017; revision accepted for publication 5 February 2018.

LITERATURE CITED

Achilli, A., U. A. Perego, H. Lancioni et al. 2013. Reconciling migration models to the Americas with the variation of North American native mitogenomes. *Proc. Natl. Acad. Sci. U. S. A.* 110:14,308–14,313.

Andrews, R. M., I. Kubacka, P. F. Chinnery et al. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* 23:147.

Bisso-Machado, R., M. C. Bortolini, and F. M. Salzano. 2012. Uniparental genetic markers in South Amerindians. *Genet. Mol. Biol.* 35:365–387.

Bodner, M., U. A. Perego, G. Huber et al. 2012. Rapid coastal spread of first Americans: Novel insights from South America's Southern Cone mitochondrial genomes. *Genome Res.* 22:811–820.

Camargo, A., L. J. Avila, M. Morando et al. 2012. Accuracy and precision of species trees: Effects of locus, individual, and base pair sampling on inference of species trees in lizards of the *Liolaemus darwinii* group (Squamata, Liolaemidae). *Syst. Biol.* 61:272–288.

Carnese, F. R., F. Mendisco, C. Keyser et al. 2010. Paleogenetical study of pre-Columbian samples from Pampa Grande (Salta, Argentina). *Am. J. Phys. Anthropol.* 141:452–462.

Cotton, J. A., and M. Wilkinson. 2008. Quantifying the potential utility of phylogenetic characters. *Taxon* 57:131–136.

Darriba, D., G. L. Taboada, R. Doallo et al. 2012. jModelTest 2: More models, new heuristics and parallel computing. *Nat. Methods* 9:772.

de Saint Pierre, M., C. M. Bravi, J. M. Motti et al. 2012a. An alternative model for the early peopling of southern South America revealed by analyses of three mitochondrial DNA haplogroups. *PLoS One* 7:e43486.

de Saint Pierre, M., F. Gandini, U. A. Perego et al. 2012b. Arrival of Paleo-Indians to the Southern Cone of South America: New clues from mitogenomes. *PLoS One* 7:e51311.

Dornburg, A., J. P. Townsend, M. Friedman et al. 2014. Phylogenetic informativeness reconciles ray-finned fish molecular divergence times. *BMC Evol. Biol.* 14:1.

Drummond, A. J., and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:214.

Drummond, A. J., A. Rambaut, B. Shapiro et al. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* 22:1,185–1,192.

Endicott, P., and S. Y. W. Ho. 2008. A Bayesian evaluation of human mitochondrial substitution rates. *Am. J. Hum. Genet.* 82:895–902.

Fagundes, N. J., R. Kanitz, and S. L. Bonatto. 2008a. A reevaluation of the Native American mtDNA genome diversity and its bearing on the models of early colonization of Beringia. *PLoS One* 3:e3157.

Fagundes, N., R. Kanitz, R. Eckert et al. 2008b. Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. *Am. J. Hum. Genet.* 82:583–592.

Fehren-Schmitz, L. W. Haak, B. Mächtle et al. 2014. Climate change underlies global demographic, genetic, and cultural transitions in pre-Columbian southern Peru. *Proc. Natl. Acad. Sci. U. S. A.* 111:9,443–9,448.

Fehren-Schmitz, L., B. Llamas, E. Tomasto et al. 2011. El ADN antiguo y la historia del poblamiento temprano del Oeste de Sudamérica: Lo que hemos aprendido y hacia dónde vamos. *Bol. Arqueol. PUCP* 15:17–41.

Felsenstein, J. 2005. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.

Geuten, K., T. Massingham, P. Darius et al. 2007. Experimental design criteria in phylogenetics: Where to add taxa. *Syst. Biol.* 56:609–622.

Goldberg, A., A. M. Mychajliw, and E. A. Hadly. 2016. Post-invasion demography of prehistoric humans in South America. *Nature* 532:232–235.

Goldman, N. 1998. Phylogenetic information and experimental design in molecular systematics. *Proc. Biol. Sci.* 265:1,779–1,786.

Graybeal, A. 1994. Evaluating the phylogenetic utility of genes: A search for genes informative about deep divergences among vertebrates. *Syst. Biol.* 43:174–193.

Graybeal, A. 1998. Is it better to add taxa or characters to a difficult phylogenetic problem? *Syst. Biol.* 47:9–17.

Guindon, S., and O. Gascuel. 2003. A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52:696–704.

Hall, T. 2004. BioEdit 7.0.0. http://www.mbio.ncsu.edu/bioedit/bioedit.html.

Hammer, Ø., D. A. T. Harper, and P. D. Ryan. 2001. PAST: Paleontological statistics software package for education and data analysis. *Palaeontol. Electron.* 4:1–9.

Hartmann, A., M. Thieme, L. K. Nanduri et al. 2009. Validation of microarray-based resequencing of 93 worldwide mitochondrial genomes. *Hum. Mutat.* 30:115–122.

Ho, S. Y. W., and B. Shapiro. 2011. Skyline-plot methods for estimating demographic history from nucleotide sequences. *Mol. Ecol. Resour.* 11:423–434.

Just, R. S., T. M. Diegoli, J. L. Saunier. 2008. Complete mitochondrial genome sequences for 265 African American and US "Hispanic" individuals. *Forensic Sci. Int. Genet.* 2:e45–e48.

Katoh, K., and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* 30:772–780.

Kemp, B. M., and T. G. Schurr. 2010. Ancient and modern genetic variation in the Americas. In *Human Variation in the Americas: The Integration of Archaeology and Biological Anthropology,* B. M. Auerbach, ed. Carbondale: Southern Illinois University Board of Trustees, 12–50.

Kitchen, A., M. M. Miyamoto, and C. J. Mulligan. 2008. A three-stage colonization model for the peopling of the Americas. *PLoS One* 3:e1596.

Kuhner, M. K., and J. Felsenstein. 1994. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. *Mol. Biol. Evol.* 11:459–468.

Kumar, S., G. Stecher, and K. Tamura 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33:1,870–1,874.

Legendre, P., and L. Legendre. 1998. *Numerical Ecology*. Amsterdam: Elsevier.

Li, H., and Durbin, R. 2011. Inference of human population history from individual whole-genome sequences. *Nature* 475:493–496.

Llamas, B., L. Fehren-Schmitz, G. Valverde et al. 2016. Ancient mitochondrial DNA provides high-resolution time scale of the peopling of the Americas. *Sci. Adv.*

2:e1501385.

López-Giráldez, F., and J. P. Townsend. 2011. PhyDesign: An online application for profiling phylogenetic informativeness. *BMC Evol. Biol.* 11:1.

Lynch Alfaro, J. W., J. P. Boubli, L. E. Olson et al. 2012. Explosive Pleistocene range expansion leads to widespread Amazonian sympatry between robust and gracile capuchin monkeys. *J. Biogeogr.* 39:272–288.

Macaulay, V., C. Hill, A. Achilli et al. 2005. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* 308:1,034–1,036.

Mallick, S., H. Li, M. Lipson et al. 2016. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* 538:201–206.

Marín, J. C., A. E. Spotorno, B. A. González et al. 2008. Mitochondrial DNA variation and systematics of the guanaco (*Lama guanicoe*, Artiodactyla: Camelidae). *J. Mammal.* 89:269–281.

Marquet, P. A., C. M. Santoro, C. Latorre et al. 2012. Emergence of social complexity among coastal hunter-gatherers in the Atacama Desert of northern Chile. *Proc. Natl. Acad. Sci. U. S. A.* 109:14,754–14,760.

Mendisco, F., C. Keyser, V. Seldes et al. 2014. Genetic diversity of a late prehispanic group of the Quebrada de Humahuaca, northwestern Argentina. *Ann. Hum. Genet.* 78:367–380.

Metcalf, J. L., C. Turney, R. Barnett et al. 2016. Synergistic roles of climate warming and human occupation in Patagonian megafaunal extinctions during the last deglaciation. *Sci. Adv.* 2:e1501682.

Moraga, M., P. Rocco, J. F. Miquel et al. 2000. Mitochondrial DNA polymorphisms in Chilean aboriginal populations: Implications for the peopling of the Southern Cone of the continent. *Am. J. Phys. Anthropol.* 113:19–29.

Non, A. L., A. Kitchen, and C. J. Mulligan. 2007. Identification of the most informative regions of the mitochondrial genome for phylogenetic and coalescent analyses. *Mol. Phylogenet. Evol.* 44:1,164–1,171.

O'Rourke, D. H., and J. A. Raff. 2010. The human genetic history of the Americas: The final frontier. *Curr. Biol.* 20:R202–R207.

Perego, U. A., A. Achilli, N. Angerhofer et al. 2009. Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. *Curr. Biol.* 19:1–8.

Perego, U. A., N. Angerhofer, M. Pala et al. 2010. The initial peopling of the Americas: A growing number of founding mitochondrial genomes from Beringia. *Genome Res.* 20:1,174–1,179.

Perez, S. I., M. B. Postillone, D. Rindel et al. 2016. Peopling time, spatial occupation and demography of late Pleistocene-Holocene human population from Patagonia. *Quat. Int.* 425:214–223.

Postillone, M. B., M. L. Fuchs, C. M. Crespo et al. 2017. Linajes maternos en muestras antiguas de la Puna Jujeña: Comparación con estudios de la región Centro-Sur Andina. *Rev. Argent. Antropol. Biol.* 19:7–22.

Poznik, G. D., Y. Xue, F. L. Mendez et al. 2016. Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. *Nat. Genet.* 48:593–599.

Rambaut, A., and A. J. Drummond. 2007. Tracer v1.5. http://tree.bio.ed.ac.uk/software/tracer/.

Rasmussen, M., X. Guo, Y. Wang et al. 2011. An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science* 334:94–98.

Robinson, D. F., and L. R. Foulds. 1981. Comparison of phylogenetic trees. *Math. Biosci.* 53:131–147.

Rozas, J. 2009. DNA sequence polymorphism analysis using DnaSP. In *Bioinformatics for DNA Sequence Analysis: Methods in Molecular Biology Series,* Vol. 537, D. Posada, ed. New York: Humana Press, 337–350.

Rubicz, R. C., P. E. Melton, and M. H. Crawford. 2007. Molecular markers in anthropological genetic studies. In *Anthropological Genetics: Theory, Methods and Applications,* M. H. Crawford, ed. Cambridge: Cambridge University Press, 141–186.

Smouse, P. E., J. C. Long, and R. R. Sokal. 1986. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Syst. Zool.* 35:627–632.

Soares, P., L. Ermini, N. Thomson et al. 2009. Correcting for purifying selection: An improved human mitochondrial molecular clock. *Am. J. Hum. Genet.* 84:740–759.

Tamm, E., T. Kivisild, M. Reidla et al. 2007. Beringian standstill and spread of Native American founders. *PLoS One* 2:829.

Tamura, K., M. Nei, and S. Kumar. 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc. Natl. Acad. Sci. U. S. A.* 101:11,030–11,035.

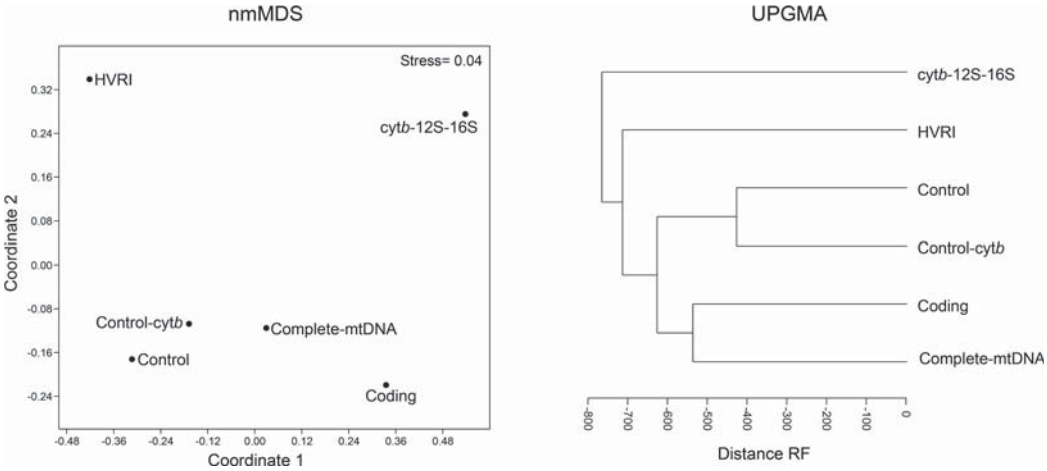
Townsend, J. P. 2007. Profiling phylogenetic informativeness. *Syst. Biol.* 56:222–231.

Vigilant, L., M. Stoneking, H. Harpending et al. 1991. African populations and the evolution of human mitochondrial DNA. *Science* 253:1,503–1,507.

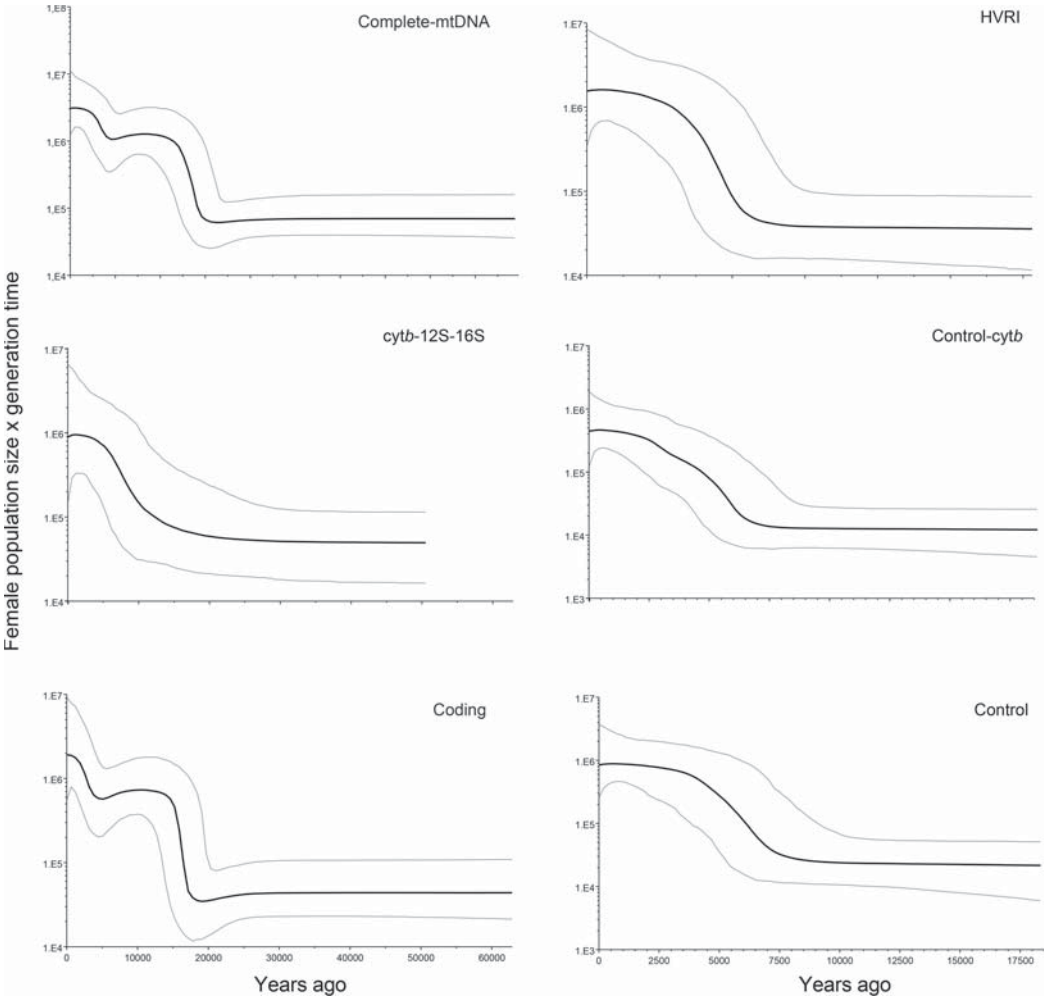
Wallace, D. C. 2015. Mitochondrial DNA variation in human radiation and disease. *Cell* 163:33–38.

Yang, Z. 1998. On the best evolutionary rate for phylogenetic analysis. *Syst. Biol.* 47:125–133.

Yang, Z., and B. Rannala. 2012. Molecular phylogenetics: Principles and practice. *Nat. Rev. Genet.* 13:303–314.



SUPPLEMENTARY FIGURE S1. Nonmetric multidimensional scaling (nmMDS) and unweighted pair group method with arithmetic mean (UPGMA) cluster analyses of topological differences among the phylogenetic trees for all American sequences. The analyses used the Robinson-Foulds (RF) distance to estimate the similarity between the data sets.



SUPPLEMENTARY FIGURE S2. Demographic changes in South America estimated based on the modern molecular sequences of the six data sets studied. The times are scaled.

Supplementary Table S1. Data Set of 259 Complete Mitochondrial DNA (mtDNA) Sequences from Modern and Ancient Humans of South America (South-Central Andes, *n* = 123; Northeast, *n* = 17; Pampa-Patagonia, *n* = 119) Used in This Study

Genebank Accession #	This Study ID	Haplogroup/Haplotype	Geographic Region	Paper
JN253392.1	1.D1	D1	Argentina, Río Negro	Bodner et al. 2012
JN253393.1	2.D1g	D1g	Chile, Los Lagos	Bodner et al. 2012
JN253394.1	3.D1g1a	D1g1a	Argentina, Buenos Aires	Bodner et al. 2012
JN253395.1	4.D1g1a	D1g1a	Argentina, Río Negro	Bodner et al. 2012
JN253396.1	5.D1g1a	D1g1a	Argentina, Río Negro	Bodner et al. 2012
JN253397.1	6.D1g1a	D1g1a	Argentina, Río Negro	Bodner et al. 2012
JN253398.1	7.D1g1b	D1g1b	Chile, Tarapacá	Bodner et al. 2012
JN253399.1	8.D1g1b	D1g1b	Argentina, Chubut	Bodner et al. 2012
JN253400.1	9.D1g1b	D1g1b	Argentina, Río Negro	Bodner et al. 2012
JN253401.1	10.D1g2	D1g2	Chile, Biobio	Bodner et al. 2012
JN253402.1	11.D1g2	D1g2	Argentina, Buenos Aires	Bodner et al. 2012
JN253403.1	12.D1g2	D1g2	Argentina, Río Negro	Bodner et al. 2012
JN253404.1	13.D1g2a	D1g2a	Argentina, Río Negro	Bodner et al. 2012
JN253405.1	14.D1g2a	D1g2a	Argentina, Río Negro	Bodner et al. 2012
JN253406.1	15.D1g2a	D1g2a	Argentina, Río Negro	Bodner et al. 2012
JN253407.1	16.D1g3	D1g3	Chile, Maule	Bodner et al. 2012
JN253408.1	17.D1g3	D1g3	Chile, Valparaíso	Bodner et al. 2012
JN253409.1	18.D1g3	D1g3	Argentina, Río Negro	Bodner et al. 2012
JN253410.1	19.D1g4a	D1g4a	Chile, Valparaíso	Bodner et al. 2012
JN253411.1	20.D1g4a	D1g4a	Argentina, Río Negro	Bodner et al. 2012
JN253412.1	21.D1g4	D1g4	Chile, Atacama	Bodner et al. 2012
JN253413.1	22.D1g5	D1g5	Chile, Aisen	Bodner et al. 2012
JN253414.1	23.D1g5	D1g5	Argentina, Buenos Aires	Bodner et al. 2012
JN253415.1	24.D1g5	D1g5	Argentina, Río Negro	Bodner et al. 2012
JN253416.1	25.D1g5	D1g5	Argentina, Río Negro	Bodner et al. 2012
JN253417.1	26.D1g6	D1g6	Chile, Araucania	Bodner et al. 2012
JN253418.1	27.D1g6	D1g6	Argentina, Neuquén	Bodner et al. 2012
JN253419.1	28.D1j	D1j	Chile, Biobio	Bodner et al. 2012
JN253420.1	29.D1j	D1j	Brazil, Río Grande do Sul	Bodner et al. 2012
JN253421.1	30.D1j1	D1j1	Argentina, Buenos Aires	Bodner et al. 2012
JN253422.1	31.D1j1	D1j1	Argentina, Catamarca	Bodner et al. 2012
JN253423.1	32.D1j1a	D1j1a	Argentina, Córdoba	Bodner et al. 2012
JN253424.1	33.D1j1a	D1j1a	Chile, Maule	Bodner et al. 2012
JN253425.1	34.D1j1a	D1j1a	Chile, Valparaíso	Bodner et al. 2012
JN253426.1	35.D1j1a	D1j1a	Brazil, São Paulo	Bodner et al. 2012
JN253427.1	36.D1j1a	D1j1a	Argentina, Catamarca	Bodner et al. 2012
JN253428.1	37.D1j1a	D1j1a	Argentina, Catamarca	Bodner et al. 2012
JN253429.1	38.D1j1a1	D1j1a1	Argentina, Tucumán	Bodner et al. 2012
JN253430.1	39.D1j1a1	D1j1a1	Argentina, Buenos Aires	Bodner et al. 2012
JN253431.1	40.D1j1a1	D1j1a1	Argentina, Buenos Aires	Bodner et al. 2012
JN253432.1	41.D1j1a1	D1j1a1	Argentina, Buenos Aires	Bodner et al. 2012
JN253433.1	42.D1j1a1	D1j1a1	Argentina, Salta	Bodner et al. 2012
JN253434.1	43.D1j1a2	D1j1a2	Argentina, Buenos Aires	Bodner et al. 2012
JN253435.1	44.D1j1a2	D1j1a2	Argentina, Corrientes	Bodner et al. 2012
JN253391.1	45.D1	D1	Argentina, Buenos Aires	Bodner et al. 2012
JX413011.1	46.B2i2a	B2i2a	Argentina, Río Negro	de Saint Pierre et al. 2012
JX413012.1	47.B2i2a	B2i2a	Argentina, Neuquén	de Saint Pierre et al. 2012
JX413013.1	48.B2i2a	B2i2a	Argentina, Neuquén	de Saint Pierre et al. 2012
JX413014.1	49.B2i2a	B2i2a	Chile, Isla de Chiloé	de Saint Pierre et al. 2012
JX413015.1	50.B2i2a	B2i2a	Chile, Isla de Chiloé	de Saint Pierre et al. 2012

Genebank Accession #	This Study ID	Haplogroup/Haplotype	Geographic Region	Paper
JX413016.1	51.B2i2a	B2i2a	Chile, Isla de Chiloé	de Saint Pierre et al. 2012
JX413017.1	52.B2i2a	B2i2a	Chile, Aconcagua	de Saint Pierre et al. 2012
JX413018.1	53.B2i2a	B2i2a	Chile	de Saint Pierre et al. 2012
JX413019.1	54.B2i2a	B2i2a	Argentina, Río Negro	de Saint Pierre et al. 2012
JX413020.1	55.B2i2a	B2i2a	Chile	de Saint Pierre et al. 2012
JX413021.1	56.B2i2a	B2i2a	Chile, Aconcagua	de Saint Pierre et al. 2012
JX413022.1	57.B2i2a	B2i2a	Chile	de Saint Pierre et al. 2012
JX413023.1	58.B2i2a	B2i2a	Chile, Aconcagua	de Saint Pierre et al. 2012
JX413024.1	59.B2i2b	B2i2b	Chile, Aconcagua	de Saint Pierre et al. 2012
JX413025.1	60.B2i2b	B2i2b	Chile, Aconcagua	de Saint Pierre et al. 2012
JX413026.1	61.B2i2b	B2i2b	Chile, Isla de Chiloé	de Saint Pierre et al. 2012
JX413027.1	62.B2i2b	B2i2b	Chile, Valparaíso	de Saint Pierre et al. 2012
JX413028.1	63.B2i2b	B2i2b	Chile	de Saint Pierre et al. 2012
JX413029.1	64.B2i2b	B2i2b	Chile, Aconcagua	de Saint Pierre et al. 2012
JX413030.1	65.B2i2b	B2i2b	Chile, Isla de Chiloé	de Saint Pierre et al. 2012
JX413031.1	66.B2i2b	B2i2b	Argentina, Río Negro	de Saint Pierre et al. 2012
JX413032.1	67.B2i2b	B2i2b	Chile, Aconcagua	de Saint Pierre et al. 2012
JX413033.1	68.B2i2b	B2i2b	Chile, Aconcagua	de Saint Pierre et al. 2012
JX413034.1	69.B2i2b	B2i2b	Chile, San Juan de la Costa	de Saint Pierre et al. 2012
JX413035.1	70.B2i2	B2i2	Chile, Santiago	de Saint Pierre et al. 2012
JX413036.1	71.C1b13a	C1b13a	Chile, San Juan de la Costa	de Saint Pierre et al. 2012
JX413037.1	72.C1b13a	C1b13a	Chile, Aconcagua	de Saint Pierre et al. 2012
JX413038.1	73.C1b13a	C1b13a	Chile, Aconcagua	de Saint Pierre et al. 2012
JX413039.1	74.C1b13a	C1b13a	Chile	de Saint Pierre et al. 2012
JX413040.1	75.C1b13a	C1b13a	Chile, Aconcagua	de Saint Pierre et al. 2012
JX413041.1	76.C1b13b	C1b13b	Chile, San Juan de la Costa	de Saint Pierre et al. 2012
JX413042.1	77.C1b13b	C1b13b	Chile, isla de Chiloé	de Saint Pierre et al. 2012
JX413043.1	78.C1b13b	C1b13b	Chile, Talagante	de Saint Pierre et al. 2012
JX413044.1	79.C1b13c	C1b13c	Argentina, Neuquén	de Saint Pierre et al. 2012
JX413045.1	80.C1b13c	C1b13c	Chile, Isla de Chiloé	de Saint Pierre et al. 2012
JX413046.1	81.C1b13c	C1b13c	Chile	de Saint Pierre et al. 2012
JX413047.1	82.C1b13c	C1b13c	Chile	de Saint Pierre et al. 2012
JX413048.1	83.C1b13c	C1b13c	Chile, Trapa Trapa	de Saint Pierre et al. 2012
JX413049.1	84.C1b13d	C1b13d	Chile, Aconcagua	de Saint Pierre et al. 2012
JX413050.1	85.C1b13d	C1b13d	Chile	de Saint Pierre et al. 2012
JX413051.1	86.C1b13e	C1b13e	Chile, isla de Chiloé	de Saint Pierre et al. 2012
JX413052.1	87.C1b13e	C1b13e	Chile	de Saint Pierre et al. 2012
JX413053.1	88.C1b13e	C1b13e	Chile, isla de Chiloé	de Saint Pierre et al. 2012
JX413054.1	89.C1b13e	C1b13e	Chile, Aconcagua	de Saint Pierre et al. 2012
JX413055.1	90.C1b13	C1b13	Argentina, Salta	de Saint Pierre et al. 2012
JX413056.1	91.C1b13	C1b13	Chile, Isla de Chiloé	de Saint Pierre et al. 2012
EU597486.1	92.A2	A2	Colombia	Hartmann et al. 2009
EU597498.1	93.D1	D1	Brazil, Karitiana	Hartmann et al. 2009
EU597510.1	94.D1	D1	Brazil, Karitiana	Hartmann et al. 2009
EU597534.1	95.B2	B2	Andes	Hartmann et al. 2009
EU597546.1	96.D4h	D4h	Brazil, Guyana, Venezuela	Hartmann et al. 2009
EU597569.1	97.B2	B2	Colombia	Hartmann et al. 2009
EU597580.1	98.B2	B2	Colombia	Hartmann et al. 2009
FJ168713.1	99.D4h3a1	D4h3a1	Chile, Atacama	Perego et al. 2009
FJ168714.1	100.D4h3a1	D4h3a1	Chile, Coquimbo	Perego et al. 2009
FJ168715.1	101.D4h3a1	D4h3a1	Chile, O'Higgins	Perego et al. 2009
FJ168716.1	102.D4h3a1	D4h3a1	Chile, Biobio	Perego et al. 2009
FJ168717.1	103.D4h3a1	D4h3a1	Chile, Biobio	Perego et al. 2009
FJ168718.1	104.D4h3a1	D4h3a1	Chile, Coquimbo	Perego et al. 2009

Genebank Accession #	This Study ID	Haplogroup/Haplotype	Geographic Region	Paper
FJ168719.1	105.D4h3a1	D4h3a1	Chile, Talcahuano	Perego et al. 2009
FJ168720.1	106.D4h3a1	D4h3a1	Chile, Santiago	Perego et al. 2009
FJ168721.1	107.D4h3a1	D4h3a1	Chile, Los Lagos	Perego et al. 2009
FJ168722.1	108.D4h3a1	D4h3a1	Chile, Biobio	Perego et al. 2009
FJ168723.1	109.D4h3a1	D4h3a1	Chile, Biobio	Perego et al. 2009
FJ168724.1	110.D4h3a2	D4h3a2	Chile, Biobio	Perego et al. 2009
FJ168725.1	111.D4h3a2	D4h3a2	Chile, Biobio	Perego et al. 2009
FJ168726.1	112.D4h3a2	D4h3a2	Chile, Coquimbo	Perego et al. 2009
FJ168727.1	113.D4h3a2	D4h3a2	Chile, Biobio	Perego et al. 2009
FJ168728.1	114.D4h3a3	D4h3a3	Chile, Tarapacá	Perego et al. 2009
FJ168735.1	115.D4h3a4	D4h3a4	Perú, Suyu	Perego et al. 2009
FJ168736.1	116.D4h3a4	D4h3a4	Perú, Arequipa	Perego et al. 2009
FJ168737.1	117.D4h3a4	D4h3a4	Perú, Arequipa	Perego et al. 2009
FJ168738.1	118.D4h3a5	D4h3a5	Perú, La Libertad	Perego et al. 2009
FJ168739.1	119.D4h3a5	D4h3a5	Chile, Maule	Perego et al. 2009
FJ168740.1	120.D4h3a5	D4h3a5	Chile, Los Lagos	Perego et al. 2009
FJ168741.1	121.D4h3a5	D4h3a5	Chile, Santiago	Perego et al. 2009
FJ168743.1	122.D4h3a3	D4h3a3	Perú, Arequipa	Perego et al. 2009
FJ168744.1	123.D4h3a	D4h3a	Perú, Arequipa	Perego et al. 2009
FJ168747.1	124.D4h3a	D4h3a	Perú, Loreto	Perego et al. 2009
FJ168748.1	125.D4h3a	D4h3a	Perú, Loreto	Perego et al. 2009
FJ168749.1	126.D4h3a	D4h3a	Perú, Piura	Perego et al. 2009
FJ168750.1	127.D4h3a	D4h3a	Bolivia, La Paz	Perego et al. 2009
FJ168751.1	128.D4h3a	D4h3a	Perú, Apurimac	Perego et al. 2009
FJ168752.1	129.D4h3a	D4h3a	Perú, Ancash	Perego et al. 2009
FJ168754.1	130.D4h3a	D4h3a	Brazil, Maranhâo	Perego et al. 2009
HM107309.1	131.C1d	C1d	Argentina, Salta	Perego et al. 2010
HM107310.1	132.C1d	C1d	Argentina, Salta	Perego et al. 2010
HM107311.1	133.C1d2a	C1d2a	Colombia	Perego et al. 2010
HM107312.1	134.C1d2a	C1d2a	Colombia	Perego et al. 2010
HM107313.1	135.C1d2a	C1d2a	Colombia	Perego et al. 2010
HM107314.1	136.C1d2a	C1d2a	Colombia	Perego et al. 2010
HM107315.1	137.C1d2	C1d2	Colombia	Perego et al. 2010
HM107316.1	138.C1d	C1d	Argentina, Buenos Aires	Perego et al. 2010
HM107317.1	139.C1d	C1d	Colombia	Perego et al. 2010
HM107323.1	140.C1d1b	C1d1b	Argentina, Salta	Perego et al. 2010
HM107324.1	141.C1d1b	C1d1b	Argentina, Catamarca	Perego et al. 2010
HM107325.1	142.C1d1b	C1d1b	Argentina, Salta	Perego et al. 2010
HM107326.1	143.C1d1b1	C1d1b1	Argentina, Catamarca	Perego et al. 2010
HM107327.1	144.C1d1b1	C1d1b1	Argentina, Buenos Aires	Perego et al. 2010
HM107328.1	145.C1d1b1	C1d1b1	Argentina, Río Negro	Perego et al. 2010
HM107329.1	146.C1d1b1	C1d1b1	Argentina, Buenos Aires	Perego et al. 2010
HM107330.1	147.C1d1b1	C1d1b1	Argentina, Corrientes	Perego et al. 2010
HM107331.1	148.C1d1b1	C1d1b1	Uruguay, Flores	Perego et al. 2010
HM107332.1	149.C1d1b1	C1d1b1	Argentina, Buenos Aires	Perego et al. 2010
HM107333.1	150.C1d1b1	C1d1b1	Argentina, Salta	Perego et al. 2010
HM107338.1	151.C1d1	C1d1	Brazil, Río Grande do Sul	Perego et al. 2010
HM107339.1	152.C1d1	C1d1	Perú, Lima	Perego et al. 2010
HM107340.1	153.C1d1	C1d1	Argentina, Buenos Aires	Perego et al. 2010
HM107341.1	154.C1d1	C1d1	Perú, Loreto	Perego et al. 2010
HM107342.1	155.C1d1	C1d1	Perú, Loreto	Perego et al. 2010
HM107343.1	156.C1d1	C1d1	Ecuador, Imbabura	Perego et al. 2010
HM107344.1	157.C1d1	C1d1	Colombia	Perego et al. 2010
HM107345.1	158.C1d1	C1d1	Colombia	Perego et al. 2010

Genebank Accession #	This Study ID	Haplogroup/Haplotype	Geographic Region	Paper
HM107346.1	159.C1d1d	C1d1d	Argentina, Buenos Aires	Perego et al. 2010
HM107347.1	160.C1d1d	C1d1d	Brazil, Río Grande do Sul	Perego et al. 2010
HM107348.1	161.C1d1d	C1d1d	Uruguay	Perego et al. 2010
HM107349.1	162.C1d1	C1d1	Brazil, Minas Gerais	Perego et al. 2010
HM107350.1	163.C1d1e	C1d1e	Chile, Biobio	Perego et al. 2010
HM107351.1	164.C1d1e	C1d1e	Argentina, Río Negro	Perego et al. 2010
HM107352.1	165.C1d1	C1d1	Perú, Cajamarca	Perego et al. 2010
HM107353.1	166.C1d1	C1d1	Perú, Huanucu	Perego et al. 2010
HM107354.1	167.C1d1	C1d1	Perú, Puca Puca	Perego et al. 2010
HM107355.1	168.C1d1	C1d1	Argentina, Buenos Aires	Perego et al. 2010
HM107356.1	169.C1d1	C1d1	Brazil, Mato grosso do Sul	Perego et al. 2010
HM107357.1	170.C1d1	C1d1	Paraguay	Perego et al. 2010
HM107358.1	171.C1d1	C1d1	Argentina, Salta	Perego et al. 2010
HM107359.1	172.C1d1	C1d1	Perú, Piura	Perego et al. 2010
HM107360.1	173.C1d1	C1d1	Perú, Huancavelica	Perego et al. 2010
HM107361.1	174.C1d1	C1d1	Argentina, Corrientes	Perego et al. 2010
HM107362.1	175.C1d1	C1d1	Chile, Los Lagos	Perego et al. 2010
HM107363.1	176.C1d1	C1d1	Chile, Los Lagos	Perego et al. 2010
EU095222.1	177.C1d1	C1d1	Brazil	Perego et al. 2010
KU523264.1	178.A2	A2	Perú, Wari	Llamas et al. 2016
KU523265.1	179.A2	A2	Perú, Wari	Llamas et al. 2016
KU523266.1	180.A2	A2	Perú, Lima	Llamas et al. 2016
KU523267.1	181.A2	A2	Perú	Llamas et al. 2016
KU523268.1	182.A2	A2	Perú, Tiwanaku	Llamas et al. 2016
KU523269.1	183.A2	A2	Argentina, Arroyo Seco	Llamas et al. 2016
KU523270.1	184.A2	A2	Chile	Llamas et al. 2016
KU523271.1	185.A2	A2	Perú	Llamas et al. 2016
KU523272.1	186.A2	A2	Chile	Llamas et al. 2016
KU523273.1	187.A2	A2	Chile	Llamas et al. 2016
KU523274.1	188.A2	A2	Perú	Llamas et al. 2016
KU523275.1	189.B2b	B2b	Perú	Llamas et al. 2016
KU523276.1	190.B2	B2	Perú	Llamas et al. 2016
KU523277.1	191.B2b	B2b	Perú	Llamas et al. 2016
KU523278.1	192.B2b	B2b	Perú	Llamas et al. 2016
KU523279.1	193.B2	B2	Perú	Llamas et al. 2016
KU523280.1	194.B2b	B2b	Perú, Wari	Llamas et al. 2016
KU523281.1	195.B2b	B2b	Perú, Wari	Llamas et al. 2016
KU523282.1	196.B2b	B2b	Perú, Wari	Llamas et al. 2016
KU523283.1	197.B2b	B2b	Perú, Lima	Llamas et al. 2016
KU523284.1	198.B2	B2	Perú	Llamas et al. 2016
KU523285.1	199.B2b	B2b	Perú	Llamas et al. 2016
KU523286.1	200.B2	B2	Perú, Lima	Llamas et al. 2016
KU523287.1	201.B2	B2	Perú, Lima	Llamas et al. 2016
KU523288.1	202.B2	B2	Perú, Lima	Llamas et al. 2016
KU523289.1	203.B2	B2	Chile	Llamas et al. 2016
KU523290.1	204.B2	B2	Chile	Llamas et al. 2016
KU523291.1	205.B2	B2	Chile	Llamas et al. 2016
KU523292.1	206.B2	B2	Perú	Llamas et al. 2016
KU523293.1	207.B2	B2	Perú	Llamas et al. 2016
KU523294.1	208.B2	B2	Perú	Llamas et al. 2016
KU523295.1	209.B2b	B2b	Perú	Llamas et al. 2016
KU523296.1	210.B2b	B2b	Perú	Llamas et al. 2016
KU523297.1	211.B2	B2	Perú	Llamas et al. 2016
KU523298.1	212.B2b	B2b	Perú	Llamas et al. 2016

Genebank Accession #	This Study ID	Haplogroup/Haplotype	Geographic Region	Paper
KU523299.1	213.B2	B2	Bolivia	Llamas et al. 2016
KU523300.1	214.B2	B2	Bolivia	Llamas et al. 2016
KU523301.1	215.B2	B2	Bolivia	Llamas et al. 2016
KU523302.1	216.B2	B2	Bolivia	Llamas et al. 2016
KU523303.1	217.B2	B2	Bolivia	Llamas et al. 2016
KU523308.1	218.B2	B2	Perú	Llamas et al. 2016
KU523309.1	219.B2	B2	Perú	Llamas et al. 2016
KU523310.1	220.B2	B2	Perú	Llamas et al. 2016
KU523311.1	221.B2	B2	Perú	Llamas et al. 2016
KU523312.1	222.C1b	C1b	Argentina, Llullaillaco	Llamas et al. 2016
KU523313.1	223.C1b	C1b	Perú	Llamas et al. 2016
KU523314.1	224.C1b	C1b	Perú	Llamas et al. 2016
KU523315.1	225.C1b	C1b	Perú	Llamas et al. 2016
KU523316.1	226.C1b	C1b	Perú	Llamas et al. 2016
KU523317.1	227.C1b	C1b	Perú	Llamas et al. 2016
KU523318.1	228.C1b	C1b	Perú	Llamas et al. 2016
KU523319.1	229.C1b	C1b	Perú	Llamas et al. 2016
KU523320.1	230.C1b	C1b	Perú	Llamas et al. 2016
KU523321.1	231.C1b	C1b	Perú	Llamas et al. 2016
KU523322.1	232.C1c	C1c	Perú	Llamas et al. 2016
KU523323.1	233.C1b	C1b	Perú	Llamas et al. 2016
KU523324.1	234.C1b	C1b	Perú	Llamas et al. 2016
KU523325.1	235.C1b	C1b	Perú	Llamas et al. 2016
KU523326.1	236.C1b	C1b	Perú	Llamas et al. 2016
KU523327.1	237.C1b	C1b	Perú	Llamas et al. 2016
KU523328.1	238.C1b	C1b	Perú	Llamas et al. 2016
KU523329.1	239.C1c	C1c	Perú	Llamas et al. 2016
KU523330.1	240.C1b	C1b	Perú	Llamas et al. 2016
KU523331.1	241.C1c	C1c	Perú	Llamas et al. 2016
KU523332.1	242.C1b	C1b	Perú	Llamas et al. 2016
KU523333.1	243.C1b	C1b	Perú	Llamas et al. 2016
KU523334.1	244.C1c	C1c	Perú	Llamas et al. 2016
KU523335.1	245.C1c	C1c	Bolivia	Llamas et al. 2016
KU523336.1	246.C1	C1	Bolivia	Llamas et al. 2016
KU523337.1	247.C1b	C1b	Bolivia	Llamas et al. 2016
KU523339.1	248.C1d	C1d	Perú	Llamas et al. 2016
KU523340.1	249.C1d	C1d	Perú	Llamas et al. 2016
KU523341.1	250.D1	D1	Argentina, Llullaillaco	Llamas et al. 2016
KU523342.1	251.D1	D1	Perú	Llamas et al. 2016
KU523343.1	252.D	D	Perú	Llamas et al. 2016
KU523344.1	253.D1	D1	Perú	Llamas et al. 2016
KU523345.1	254.D	D	Perú	Llamas et al. 2016
KU523346.1	255.D1	D1	Perú	Llamas et al. 2016
KU523347.1	256.D1	D1	Perú	Llamas et al. 2016
KU523348.1	257.D1	D1	Perú	Llamas et al. 2016
KU523349.1	258.D1	D1	Argentina, Arroyo Seco	Llamas et al. 2016
KU523350.1	259.D1	D1	Perú	Llamas et al. 2016

Supplementary Table S2. Data Set of 174 Complete Mitochondrial DNA (mtDNA) Sequences from Modern Humans of North America Used in This Study

Genbank Accession #	Haplogroup/Haplotype	Geographic Region	Paper
KC710999	A2a	Alaska	Achilli et al. 2013
KC711000	A2a3	Groenlandia	Achilli et al. 2013
KC711001	A2a4	Nuevo México	Achilli et al. 2013
KC711002	A2a4	Nuevo México	Achilli et al. 2013
KC711003	A2a4	Arizona	Achilli et al. 2013
KC711004	A2a4	Nuevo México	Achilli et al. 2013
KC711005	A2a4	Chihuhua, México	Achilli et al. 2013
KC711006	A2a4	Chihuhua, México	Achilli et al. 2013
KC711007	A2a4	Arizona/ Nuevo México	Achilli et al. 2013
KC711008	A2a	noroeste de Canadá	Achilli et al. 2013
KC711009	A2a5	Nuevo México	Achilli et al. 2013
KC711010	A2a5	Texas	Achilli et al. 2013
KC711011	A2a5	California	Achilli et al. 2013
KC711012	A2a5	Arizona	Achilli et al. 2013
KC711013	A2a5	Arizona	Achilli et al. 2013
KC711014	A2a5	California	Achilli et al. 2013
KC711015	A2a5	Arizona	Achilli et al. 2013
KC711016	A2a5	Nuevo México	Achilli et al. 2013
KC711017	A2a5	Canadá	Achilli et al. 2013
KC711018	A2a5	noroeste de Canadá	Achilli et al. 2013
KC711019	A2a5	Nuevo México	Achilli et al. 2013
KC711020	A2a5	Nuevo México	Achilli et al. 2013
KC711021	B2a	noroeste de Canadá	Achilli et al. 2013
KC711022	B2a	Chihuhua, México	Achilli et al. 2013
KC711023	B2a1a	México	Achilli et al. 2013
KC711024	B2a1a	Chihuhua, México	Achilli et al. 2013
KC711025	B2a1b	Chihuhua, México	Achilli et al. 2013
KC711026	B2a1	Nuevo México	Achilli et al. 2013
KC711027	B2a1	México	Achilli et al. 2013
KC711028	B2a2	Nuevo México	Achilli et al. 2013
KC711029	B2a2	Colorado	Achilli et al. 2013
KC711030	B2a2	Colorado	Achilli et al. 2013
KC711031	B2a3	Chihuhua, México	Achilli et al. 2013
KC711032	B2a3	Durango, México	Achilli et al. 2013
KC711033	B2a4a	Sinaloa, México	Achilli et al. 2013
KC711034	B2a4a1	Chihuhua, México	Achilli et al. 2013
KC711035	B2a4a1	Jalisco, México	Achilli et al. 2013
KC711036	B2a4a1	Durango, México	Achilli et al. 2013
KC711037	B2a5	Arizona	Achilli et al. 2013
KC711038	B2a5	Utah	Achilli et al. 2013
KC711039	B2a5	Arizona	Achilli et al. 2013
EU597533	C1c	Pima, México	Hartmann et al. 2008
EU597545	C1b	Pima, México	Hartmann et al. 2008
EU597557	C1b	Pima, México	Hartmann et al. 2008
FJ168729	D4h3a3	Nuevo León, México	Perego et al. 2009
FJ168730	D4h3a3	California, Estados Unidos	Perego et al. 2009
FJ168731	D4h3a3	Chihuhua, México	Perego et al. 2009
FJ168732	D4h3a3	Chihuhua, México	Perego et al. 2009
FJ168733	D4h3a3	Chihuhua, México	Perego et al. 2009
FJ168734	D4h3a3	Tarahumara, México	Perego et al. 2009
FJ168742	D4h3a	Veracruz, México	Perego et al. 2009

Genbank Accession #	Haplogroup/Haplotype	Geographic Region	Paper
FJ168745	D4h3a	Sonora, México	Perego et al. 2009
FJ168746	D4h3a	Chihuhua, México	Perego et al. 2009
FJ168753	D4h3a	San Luis Potosi, México	Perego et al. 2009
FJ168755	D4h3a	California, Estados Unidos	Perego et al. 2009
HM107306	C1d	Tamaulipas, México	Perego et al. 2010
HM107307	C1d	Guanajuato, México	Perego et al. 2010
HM107308	C1d	Chihuahua, México	Perego et al. 2010
HM107318	C1d1a1	Oklahoma, Estados Unidos	Perego et al. 2010
HM107319	C1d1a1	Montana, Estados Unidos	Perego et al. 2010
HM107320	C1d1a1	Quebec, Canadá	Perego et al. 2010
HM107321	C1d1a1	Zacatecas, México	Perego et al. 2010
HM107322	C1d1a	Sonora, México	Perego et al. 2010
HM107334	C1d1c	Oaxaca, México	Perego et al. 2010
HM107335	C1d1c1	Texas, Estados Unidos	Perego et al. 2010
HM107336	C1d1c	Texas, Estados Unidos	Perego et al. 2010
HM107337	C1d1c	Michigan, Estados Unidos	Perego et al. 2010
HM107364	C1d1	Oklahoma, Estados Unidos	Perego et al. 2010
DQ282387	A2	Estados Unidos	Just et al. 2008
DQ282388	A2	Estados Unidos	Just et al. 2008
DQ282389	A2	Estados Unidos	Just et al. 2008
DQ282390	A2	Estados Unidos	Just et al. 2008
DQ282391	A2	Estados Unidos	Just et al. 2008
DQ282392	A2	Estados Unidos	Just et al. 2008
DQ282393	A2	Estados Unidos	Just et al. 2008
DQ282394	A2	Estados Unidos	Just et al. 2008
DQ282395	A2	Estados Unidos	Just et al. 2008
DQ282396	A2	Estados Unidos	Just et al. 2008
DQ282397	A2	Estados Unidos	Just et al. 2008
DQ282398	A2	Estados Unidos	Just et al. 2008
DQ282399	A2	Estados Unidos	Just et al. 2008
DQ282400	A2	Estados Unidos	Just et al. 2008
DQ282401	A2	Estados Unidos	Just et al. 2008
DQ282402	A2	Estados Unidos	Just et al. 2008
DQ282403	A2	Estados Unidos	Just et al. 2008
DQ282404	A2	Estados Unidos	Just et al. 2008
DQ282405	A2	Estados Unidos	Just et al. 2008
DQ282406	A2	Estados Unidos	Just et al. 2008
DQ282407	A2	Estados Unidos	Just et al. 2008
DQ282408	A2	Estados Unidos	Just et al. 2008
DQ282409	A2	Estados Unidos	Just et al. 2008
DQ282410	A2	Estados Unidos	Just et al. 2008
DQ282411	A2	Estados Unidos	Just et al. 2008
DQ282412	A2	Estados Unidos	Just et al. 2008
DQ282413	A2	Estados Unidos	Just et al. 2008
DQ282414	A2	Estados Unidos	Just et al. 2008
DQ282415	A2	Estados Unidos	Just et al. 2008
DQ282416	A2	Estados Unidos	Just et al. 2008
DQ282417	A2	Estados Unidos	Just et al. 2008
DQ282418	A2	Estados Unidos	Just et al. 2008
DQ282419	A2	Estados Unidos	Just et al. 2008
DQ282420	A2	Estados Unidos	Just et al. 2008
DQ282421	A2	Estados Unidos	Just et al. 2008
DQ282422	A2	Estados Unidos	Just et al. 2008
DQ282423	A2	Estados Unidos	Just et al. 2008

Genbank Accession #	Haplogroup/Haplotype	Geographic Region	Paper
DQ282424	A2	Estados Unidos	Just et al. 2008
DQ282425	A2	Estados Unidos	Just et al. 2008
DQ282426	A2	Estados Unidos	Just et al. 2008
DQ282427	A2	Estados Unidos	Just et al. 2008
DQ282428	A2	Estados Unidos	Just et al. 2008
DQ282429	A2	Estados Unidos	Just et al. 2008
DQ282430	A2	Estados Unidos	Just et al. 2008
DQ282431	A2	Estados Unidos	Just et al. 2008
DQ282432	A2	Estados Unidos	Just et al. 2008
DQ282433	A2	Estados Unidos	Just et al. 2008
DQ282434	B2	Estados Unidos	Just et al. 2008
DQ282435	B2	Estados Unidos	Just et al. 2008
DQ282436	B2	Estados Unidos	Just et al. 2008
DQ282437	B2	Estados Unidos	Just et al. 2008
DQ282438	B2	Estados Unidos	Just et al. 2008
DQ282439	B2	Estados Unidos	Just et al. 2008
DQ282440	B2	Estados Unidos	Just et al. 2008
DQ282441	B2	Estados Unidos	Just et al. 2008
DQ282442	B2	Estados Unidos	Just et al. 2008
DQ282443	B2	Estados Unidos	Just et al. 2008
DQ282444	B2	Estados Unidos	Just et al. 2008
DQ282445	B2	Estados Unidos	Just et al. 2008
DQ282446	B2	Estados Unidos	Just et al. 2008
DQ282447	C1	Estados Unidos	Just et al. 2008
DQ282448	C1	Estados Unidos	Just et al. 2008
DQ282449	C1	Estados Unidos	Just et al. 2008
DQ282450	C1	Estados Unidos	Just et al. 2008
DQ282451	C1	Estados Unidos	Just et al. 2008
DQ282452	C1	Estados Unidos	Just et al. 2008
DQ282453	C1	Estados Unidos	Just et al. 2008
DQ282454	C1	Estados Unidos	Just et al. 2008
DQ282455	C1	Estados Unidos	Just et al. 2008
DQ282456	C1	Estados Unidos	Just et al. 2008
DQ282457	C1	Estados Unidos	Just et al. 2008
DQ282458	C1	Estados Unidos	Just et al. 2008
DQ282459	C1	Estados Unidos	Just et al. 2008
DQ282460	C1	Estados Unidos	Just et al. 2008
DQ282461	C1	Estados Unidos	Just et al. 2008
DQ282462	C1	Estados Unidos	Just et al. 2008
DQ282463	C1	Estados Unidos	Just et al. 2008
DQ282464	C1	Estados Unidos	Just et al. 2008
DQ282465	C1	Estados Unidos	Just et al. 2008
DQ282466	C1	Estados Unidos	Just et al. 2008
DQ282467	C1	Estados Unidos	Just et al. 2008
DQ282468	C1	Estados Unidos	Just et al. 2008
DQ282469	C1	Estados Unidos	Just et al. 2008
DQ282470	C1	Estados Unidos	Just et al. 2008
DQ282471	C1	Estados Unidos	Just et al. 2008
DQ282472	C1	Estados Unidos	Just et al. 2008
DQ282473	C1	Estados Unidos	Just et al. 2008
DQ282474	C1	Estados Unidos	Just et al. 2008
DQ282475	C1	Estados Unidos	Just et al. 2008
DQ282476	C1	Estados Unidos	Just et al. 2008
DQ282477	D1	Estados Unidos	Just et al. 2008

Genbank Accession #	Haplogroup/Haplotype	Geographic Region	Paper
DQ282478	D1	Estados Unidos	Just et al. 2008
DQ282479	D1	Estados Unidos	Just et al. 2008
DQ282480	D1	Estados Unidos	Just et al. 2008
DQ282481	D1	Estados Unidos	Just et al. 2008
DQ282482	D1	Estados Unidos	Just et al. 2008
DQ282483	D1	Estados Unidos	Just et al. 2008
DQ282484	D1	Estados Unidos	Just et al. 2008
DQ282485	D1	Estados Unidos	Just et al. 2008
DQ282486	D1	Estados Unidos	Just et al. 2008
DQ282487	D1	Estados Unidos	Just et al. 2008
KU523304	B2	Estados Unidos	Just et al. 2008
KU523305	B2	Estados Unidos	Just et al. 2008
KU523306	B2	Estados Unidos	Just et al. 2008
KU523307	B2	Estados Unidos	Just et al. 2008
KU523338	C1b	Estados Unidos	Just et al. 2008