A Reevaluation of the Native American MtDNA Genome Diversity and Its Bearing on the Models of Early Colonization of Beringia

Nelson J. R. Fagundes1,2*, Ricardo Kanitz1*, Sandro L. Bonatto1*
1 Faculdade de Biociências, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil, 2 Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil

Abstract

The Americas were the last continents to be populated by humans, and their colonization represents a very interesting chapter in our species’ evolution in which important issues are still contentious or largely unknown. One difficult topic concerns the details of the early peopling of Beringia, such as for how long it was colonized before people moved into the Americas and the demography of this occupation. A recent work using mitochondrial genome (mtDNA) data presented evidence for a so called “three-stage model” consisting of a very early expansion into Beringia followed by ~20,000 years of population stability before the final entry into the Americas. However, these results are in disagreement with other recent studies using similar data and methods. Here, we reanalyze their data to check the robustness of this model and test the ability of Native American mtDNA to discriminate details of the early colonization of Beringia. We apply the Bayesian Skyline Plot approach to recover the past demographic dynamic underpinning these events using different mtDNA data sets. Our results refute the specific details of the “three-stage model”, since the early stage of expansion into Beringia followed by a long period of stasis could not be reproduced in any mtDNA data set cleaned from non-Native American haplotypes. Nevertheless, they are consistent with a moderate population bottleneck in Beringia associated with the Last Glacial Maximum followed by a strong population growth around 18,000 years ago as suggested by other recent studies. We suggest that this bottleneck erased the signals of ancient demographic history from recent Native American mtDNA pool, and conclude that the proposed early expansion and occupation of Beringia is an artifact caused by the misincorporation of non-Native American haplotypes.

Introduction

The Americas were the last continents to be settled by modern humans, most probably from northeast Asia through Beringia, the landmass that connected Asia and the Americas during periods of low sea-level [1]. Archeological data suggest that the continent was colonized in the late Pleistocene after the Last Glacial maximum (LGM). The oldest sites for North and South America are about 14.5 ky old [1], possibly suggesting a fast southward movement of the initial settlers. However, the scarceness of late Pleistocene human remains in northeast Asia make it difficult to evaluate the details of the population processes in Beringia that ultimately led to the peopling of the New World.

As an alternative to the study of archeological data, molecular data have been extensively used to infer when and how modern humans colonized the world reviewed in [2,3]. Since the pioneering study by Cann et al. [4], the mitochondrial DNA (mtDNA) has become the most widely used genetic marker to study recent human evolution [5]. In Native Americans, early studies of mtDNA variation found that these populations have five distinct major mtDNA haplogroups (A, B, C, D and X) [6,7], all of Asian origin. Although most of these studies seemed to converge on a model suggesting a single pre-Clovis migration [8–10], no consensus emerged for details such as the timing and pace of the putative occupation event under this scenario. These controversies notwithstanding, the divergence between Native American and Asian sequences for each mtDNA haplogroup led some authors to suggest that the Native American founder population stayed isolated in Beringia from the remaining Asian populations prior to their entry in the Americas [9], the so-called “out of Beringia” model. Under this scenario, Beringia played a key-role in the differentiation of the mtDNA haplogroups presently found in Native Americans.

The study of complete mtDNA sequences from Native Americans [11–14] has allowed investigators to examine mtDNA variation in the New World with much greater resolution. The first systematic survey of coding-region mtDNA sequences including individuals from Native American ancestry was carried by Herrnstadt et al. [11], who studied individuals sampled from an urban population and relied on the screening of an incomplete set of HVS-I markers to identify mtDNA haplotypes as being of Native American origin [15]. Afterwards, a thorough revision of these sequences partially changed the original classification of the
Materials and Methods

Data sets

Initially, the original Kitchen et al. [20] data set (n = 77) was used to replicate their original findings using the evolutionary model specified in that report and another one (see below). Based on information detailed in Bandelt et al. [15] and Achilli et al. [14], we removed seven individuals who are very likely of Asian ancestry. These include three individuals assigned to haplogroup E in Herrnstadt et al. [Figure 2 in [11]], as well as four individuals belonging to the Asian sub-haplogroups B1 or B4c [15]. Another individual (Hrn352; see [11]) was removed, since it actually belongs to the West European haplogroup H. This individual was probably incorporated into the data set by mistake, as individual 532 (from Native American haplogroup A2) is absent from their data set. Finally, another individual (Kiv2070) was removed, since it had all the diagnostic coding-region mutations for West European sub-haplogroup X2b (8393, 13708, 15927), and none of the diagnostic coding-region mutations for Native American sub-haplogroup X2a (8913, 12397, 14502), and thus it is likely of recent European ancestry. That modifications resulted in a new, corrected, data set of 68 sequences of likely Native American origin, and, in a third data set with only the 9 sequences which have been misincorporated in the original data set of 77 sequences analyzed by Kitchen et al. [20]. Finally, to better test whether the anomalous early expansion seen in Kitchen et al. [20] BSP results could be explained by the non-Native American mtDNA genomes, we created a set of 10 alignments with nine genomes each randomly selected from the mtDNA genomes from the macrohaplogroups M and N [24].

Bayesian Skyline Plot

BSPs [18] have been constructed in the program Beast 1.4.7 (http://beast.bio.ed.ac.uk/). For all analyses, Markov Chain Monte Carlo (MCMC) samples were based on 100,000,000 generations, logging every 2,500 steps, with the first 10,000,000 generations discarded as the burn-in. All analyses were run multiple times to check for convergence. Following Fagundes et al. [13], we used the HKY+Γ evolutionary model, a log-normal relaxed molecular clock with a mean substitution rate of 1.26×10^{-8} mutations/site/year [17] for the complete coding sequence. The scaled effective population size was converted to the effective female population size N_e, assuming a generation time of 25 years. Importantly, assumptions about the mutation rate and the generation time will only affect the scale of the BSP, but not its shape.

Results

Our analysis of the BSP for the original Kitchen et al. data set [20] reproduced their original results, which was two periods of population growth (Figure 1A) with a long stasis between them. However, the corrected data set provided evidence for a single, post-LGM, population growth (Figure 1C), in close accordance to the other BSP analyses using only mtDNA of Native American origin [13,19,21]. That is, there was only a long tail of roughly constant population size between the time for most recent common ancestor (TMRCA) of each Native American haplogroup around the LGM and the sample TMRCA.

More strikingly, when we considered only those individuals which were removed from the original data set, the signal for the earlier expansion reappeared, despite the very low sample size in this data set (n = 9) (Figure 1B). Using our substitution rate, this expansion began ~60 kya. These results clearly show that the signal for a population expansion that they detected [20] in Native Americans mtDNAs much before LGM is an artifact caused by the incorporation of non-Native American haplotypes into the analysis. Since these nine haplotypes seem to be of Asian and European origin, from macrohaplogroups M and N, we conjecture that this signal of expansion may be related to the demographic
expansion out-of-Africa that gave rise to Eurasians. The BSPs of the ten data sets from genomes selected from these two macrohaplogroups (Figure 2) showed a very similar expansion pattern, corroborating this hypothesis.

Discussion

Our results strongly suggest that the demographic expansion putatively associated with the geographical expansion out of Central Asia and the initial peopling of Beringia as well as the estimation of ~20 ky of occupation of Beringia by human groups before they entered the Americas [20] is merely a database artifact caused by the incorporation of mtDNA genomes of non-Native American ancestry in the analysis. Because the mtDNA haplogroups and sub-haplogroups have a strong and extensively studied geographic association e.g., [5], it is possible to identify almost unambiguously the ancestry of most haplotypes [15]. While this may be also true for the Y-chromosome [25], it is certainly not for most other nuclear markers, which typically display low levels of interpopulation differentiation and extensive haplotype sharing among populations e.g., [26,27]. Possible applications of the BSP to autosomal or sex-linked haplotypes must carefully select the sampled populations to avoid incorporating into the analysis those recently introduced by admixture. Our analyses suggest that even a relatively small proportion of 12% (9/77) of “admixed” (or misassigned) haplotypes may significantly bias the overall result.

The population expansion that began 60–55 kya when non-Native American haplotypes are incorporated into the analysis most likely reflects, at least in part, the early expansion and diversification of macrohaplogroups M and N in Eurasia [28], which is unrelated to the specific process of the peopling of Beringia. As a consequence, Kitchen et al.’s estimation of a period of ~20 ky of population occupation in Beringia based on the time interval between the “two expansions” is meaningless in the context of the peopling of Beringia or the Americas. In addition, it is important to stress that, because the mtDNA haplogroups currently in America represent derivations of both macrohaplogroup M (C, D) and N (A, B, X) e.g., [14], their TMRCA reflects the TMRCA of macrohaplogroups M and N in Asia (~60 kya) [28]. Therefore, the >40 ky of constant population size found in the corrected data sets extending from the LGM bottleneck to the past to the TMRCA of all Native American
mtDNA haplogroups does not offer any detailed view of the demographic history of Native Americans before the bottleneck. The genetic bottleneck associated with human isolation in Beringia [29,30] may have eroded from the recent non-Beringian Native American mtDNA data most of the details of its pre-Beringian demographic history. In this regard, discerning the population size changes during this period would mostly require acquiring mtDNA information of ancient samples from this time.

Interestingly, an almost identical pattern of population size change was found with the Kitchen et al. corrected data set and our previous analyses of mostly Native South American mtDNAs [13]. These results, therefore, strongly corroborate the mtDNA scenario for the peopling of the Americas presented in Fagundes et al. [13] and the integrated model that we suggested elsewhere [31]. This model suggests that the ancestral population colonizing Beringia more than five thousand years before the LGM, remained isolated there during LGM, and likely experienced a population reduction and loss of genetic diversity by drift. The strong population expansion shown to have started around the end of LGM (~18 kya) probably reflects the fast migration south of the Laurentide and Cordilleran ice sheets. Taking into account that the ice-free corridor between the ice sheets had not opened yet, the ice-free corridor had not opened yet between the Laurentide and Cordilleran ice sheets. Taking into account that the ice-free corridor had not opened yet between the Laurentide and Cordilleran ice sheets.

Strong population expansion shown to have started around the end of LGM (~18 kya) probably reflects the fast migration south of the Laurentide and Cordilleran ice sheets. Taking into account that the ice-free corridor between the ice sheets had not opened yet, the ice-free corridor had not opened yet between the Laurentide and Cordilleran ice sheets. Taking into account that the ice-free corridor had not opened yet between the Laurentide and Cordilleran ice sheets.

We would like to thank an anonymous reviewer for improving an earlier version of the manuscript.

Author Contributions

Conceived and designed the experiments: NJRF RK SLB. Performed the experiments: NJRF RK SLB. Analyzed the data: NJRF RK SLB. Contributed reagents/materials/analysis tools: SLB. Wrote the paper: NJRF RK SLB.

Acknowledgments

References

1. Goebel T, Waters MR, O’Rourke DH (2008) The Late Pleistocene Dispersal of Modern Humans in the Americas. Science 319: 1497–1502.
2. Schurr T (2004) The Peopling of the New World: Perspectives from Molecular Anthropology. Annu Rev Anthropol 33: 551–583.
3. Salzano FM (2007) The prehistoric colonization of the Americas. In: Crawford MH, ed. Anthropological Genetics: Theory, Methods, and Applications. Cambridge: Cambridge University Press. pp 433–455.
4. Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. Nature 325: 17–26.
5. Torroni A, Achilli A, Macaulay V, Richards M, Bandelt HJ (2006) Harvesting the Fruit of the Human MtDNA Tree. Trends in Genetics 22: 339–345.
6. Schurr TG, Ballinger SW, Gan YY, Hodge JA, Merriweather DA, et al. (1990) Mitochondrial Haplogroup X: an Ancient Link between Europe/Western Asia and North America? Am J Hum Genet 63: 1852–1861.
7. Forster P, Harding R, Torroni A, Bandelt HJ (1996) Origin and Evolution of Native American MtDNA Variation: a Reappraisal. Am J Hum Genet 59: 935–943.
8. Bonatto SI, Salzano FM (1997) A Single and Early Migration for the Peopling of the Americas Supported by Mitochondrial DNA Sequence Data. Proc Natl Acad Sci USA 94: 1066–1071.
9. Bonatto SI, Salzano FM (1997) Diversity and age of the four major mtDNA haplogroups and their implications for the peopling of the New World. Am J Hum Genet 61: 1415–1423.
10. Herrnstadt C, Elson JL, Fahy E, Preston G, Turnbull DM, et al. (2002) New Coalescence Times for the Major Native American mtDNA Haplogroups: Their Implications for the Peopling of the Americas. Am J Hum Genet 69: 152–153.
11. Tamam E, Kvisiad T, Reidl M, Metpolu M, Smith DG, et al. (2007) Rethinking the Standstill and Spread of Native American Founders. PLoS One 2: e1596.
12. Achiili A, Perego UA, Bravi CM, Coile KD, Kong QP, et al. (2008) The Phylogeography of the Four Pan-American MitDNA Haplogroups: Implications for Evolutionary and Disease Studies. PLoS One 3: e1764.
13. Bandelt HJ, Herrnstadt C, Yao YG, Kong QP, Rengo C, et al. (2003) Identification of Native American Founder mtDNAs Through the Analysis of Complete mtDNA Sequences: Some Caveats. Ann Hum Genet 67: 512–524.
14. Kvisiad T, Shen P, Wall DP, Do B, Sung R, et al. (2006) The Role of Selection in the Evolution of Human Mitochondrial Genomes. Genetics 172: 373–387.
15. Mishmar D, Ruiz-Pesini E, Golli P, Macaulay V, Clark AG, et al. (2003) Natural Selection Shaped Regional MtDNA Variation in Humans. Proc Natl Acad Sci USA 100: 171–176.
16. Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian Coalescent Inference of Past Population Dynamics from Molecular Sequences. Mol Biol Evol 22: 1185–1192.
17. Kitchen A, Miyamoto MM, Mulligan CJ (2008) Utility of DNA Viruses for Studying Human Host History: Case Study of JC Virus. Mol Phylogen Evol 46: 673–682.
18. Kitchen A, Miyamoto MM, Mulligan CJ (2008) A Three-Stage Colonization Model for the Peopling of the Americas. PLoS One 3: e1396.
19. Atkinson QD, Gray RD, Drummond AJ (2008) MtDNA Variation Predicts Population Size in Humans and Reveals a Major Southern Asian Chapter in Human Prehistory. Mol Biol Evol 25: 465–474.
20. Bandelt HJ, Kong QP, Richards M, Macaulay V (2006) Estimation of Mutation Rates and Coalescence Times: Some Caveats. In: Bandelt HJ, Macaulay V, Richards M, eds. Human Mitochondrial DNA and the Evolution of Homo sapiens. Berlin: Springer. pp 47–90.
21. Bandelt HJ, Yao YG, Salas A, Kvisiad T, Bravi CM (2007) High Penetration of Sequencing Errors and Interpretative Shortcomings in mtDNA sequence analysis of LHON patients. Biochem Biophys Res Commun 352: 283–291.
22. Ingman M, Kaessmann H, Paabo S, Gyllensten U (2000) Mitochondrial Genome Variation and the Origin of Modern Humans. Nature 408: 708–713.
23. The Y Chromosome Consortium (2002) A Numerical System for the Tree of Human Y-Chromosomal Binary Haplogroups. Genome Res 12: 339–348.
24. Fagundes NJR, Salzano FM, Batzer MA, Deininger PL, Bonatto SL (2005) Worldwide Genetic Variation at the 3′-UTR Region of the LDLR Gene: Possible Influence of Natural Selection. Am J Hum Genet 69: 309–400.
25. Battilana J, Cardoso-Silva L, Barrancas R, Hill K, Hurtado AM, et al. (2007) Molecular Variability of the 16p13.3 Region in Amerindians and its Anthropological Significance. Am J Hum Genet 71: 64–74.
26. Macaulay V, Hill C, Achiili A, Rengo C, Clarke D, et al. (2005) Single, Rapid Coalescent Genomics Supports a Single Pre-Clovis Origin with a Coastal Route for the Peopling of the Americas. Am J Hum Genet 76: 1152–1171.
27. Achiili A, Perego UA, Bravi CM, Coile KD, Kong QP, et al. (2008) The Phylogeography of the Four Pan-American MitDNA Haplogroups: Implications for Evolutionary and Disease Studies. PLoS One 3: e1396.
28. Macaulay V, Hill C, Achiili A, Rengo C, Clarke D, et al. (2005) Single, Rapid Coalescent Genomics Supports a Single Pre-Clovis Origin with a Coastal Route for the Peopling of the Americas. Am J Hum Genet 76: 1152–1171.
29. Fagundes NJR, Salzano FM, Batzer MA, Deininger PL, Bonatto SL (2005) Worldwide Genetic Variation at the 3′-UTR Region of the LDLR Gene: Possible Influence of Natural Selection. Am J Hum Genet 69: 309–400.
30. Battilana J, Cardoso-Silva L, Barrancas R, Hill K, Hurtado AM, et al. (2007) Molecular Variability of the 16p13.3 Region in Amerindians and its Anthropological Significance. Am J Hum Genet 71: 64–74.
31. Macaulay V, Hill C, Achiili A, Rengo C, Clarke D, et al. (2005) Single, Rapid Coalescent Genomics Supports a Single Pre-Clovis Origin with a Coastal Route for the Peopling of the Americas. Am J Hum Genet 76: 1152–1171.
32. Fagundes NJR, Battilana J, Barros NA, Achiili A, Rengo C, et al. (2006) Alu insertion polymorphisms in Native Americans and Related Asian Populations. Am J Hum Biol 18: 142–160.
33. Fagundes NJR, Ray N, Beaumont M, Neunschwalder S, Salzano FM, et al. (2007) Statistical Evaluation of Alternative Models of Human Evolution. Proc Natl Acad Sci USA 104: 17614–17619.
31. Gonzales-Jose´ R, Bortolini MC, Santos FR, Bonatto SL (2008) The Peopling of America: Craniofacial Shape Variation on a Continental Scale and its Interpretation From an Interdisciplinary View. Am J Phys Anthropol; In press: doi: 1002/ajpa.20854.

32. Fladmark KR (1979) Routes: Alternate Migration Corridors for Early Man in North America. American Antiquity 44: 53–69.

33. Dixon EJ (1993) Quest for the Origins of the First Americans. Albuquerque: University of New Mexico Press. 160 p.

34. Wang S, Lewis CM, Ramachandran S, Ray N, et al. (2007) Genetic Variation and Population Structure in Native Americans. PLoS Genet 3: e185.

35. Gilbert MTP, Jenkins DL, Go ¨therstrom A, Naveran N, Sanchez JJ, et al. (2008) DNA from Pre-Clovis Human Coprolites in Oregon, North America. Science 320: 786–789.

36. Dillehay TD, Ramirez C, Pino M, Collins MB, Rosen J, et al. (2008) Monte Verde: Seaweed, Food, Medicine, and the Peopling of South America. Science 320: 784–786.

37. Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, et al. (2004) The Molecular Dissection of mtDNA Haplogroup H Confirms That the Franco-Cantabrian Glacial Refuge Was a Major Source for the European Gene Pool. Am J Hum Genet 75: 910–911.

38. Ho SYW, Phillips MJ, Cooper A, Drummond AJ (2005) Time Dependency of Molecular Rate Estimates and Systematic Overestimation of Recent Divergence Times. Mol Biol Evol 22: 1561–1568.

39. Ho SYW, Shapiro B, Phillips MJ, Cooper A, Drummond AJ (2007) Evidence for Time Dependency of Molecular Rate Estimates. Syst Biol 56: 515–522.

40. Bandelt HJ, Parson W (2008) Consistent treatment of length variants in the human mtDNA control region: a reappraisal. Int J Legal Med 122: 11–21.