3-5-2015

Identification of Whole Mitochondrial Genomes from Venezuela and Implications on Regional Phylogenies in South America

Esther J. Lee
Texas Tech University, esther.lee@ttu.edu

D. Andrew Merriwether
Binghamton University, andym@binghamton.edu

Identification of Whole Mitochondrial Genomes from Venezuela and Implications on Regional Phylogenies in South America

Esther J. Lee[1,2*] and D. Andrew Merriwether[1]

[1]Department of Anthropology, Binghamton University, PO Box 6000, Binghamton, NY 13902-6000, USA

[2]Department of Sociology, Anthropology, and Social Work, Box 41012, Texas Tech University, Lubbock, TX 79409-1012, USA

*Correspondence to: Esther J. Lee, Box 41012, Texas Tech University, Lubbock, TX 79409-1012, Fax: 806-742-1088. E-mail: esther.lee@ttu.edu

Abstract

Recent studies have expanded and refined the founding haplogroups of the Americas using whole mitochondrial (mtDNA) genome analysis. In addition to pan-American lineages, a number of studies have identified specific variants that show higher frequencies in restricted geographical areas. In order to further characterize Native American maternal lineages and specifically examine local patterns within South America, we analyzed twelve maternally unrelated Yekuana whole mtDNA genomes from one village (Sharamaña) that include the four major Native American haplogroups A2, B2, C1, and D1. Our study proposes a reconfiguration of one subhaplogroup A2 (A2aa) that is specific to South America and identifies other singleton

branches across the four haplogroups. Furthermore, we show nucleotide diversity values that increase from north to south for haplogroups C1 and D1. The results from our work add to the growing mitogenomic data that highlight local phylogenies and support the rapid genetic differentiation of South American populations, which has been correlated with the linguistic diversity in the region by previous studies.

Mitochondrial DNA (mtDNA) studies continue to inform us regarding questions about the peopling of the Americas as well as regional patterns of genetic diversity and phylogenies. Aided by the growing data from whole mitogenomes, the five Native American mtDNA haplogroups initially proposed to be "founder" lineages, A-D and X, have been expanded to at least eight maternal monophyletic haplogroups distributed throughout the Americas: A2, B2, C1b, C1c, C1d, C1d1, D1, and D4h3a (Bandelt et al. 2003; de Saint-Pierre et al. 2012; Fagundes et al. 2008; Perego et al. 2009; Tamm et al. 2007; Torroni et al. 1992). Whole mitogenome studies have further refined the phylogenies of these founding haplogroups and have identified unique lineages within them, among which some appear specific to certain geographical regions (Bodner et al. 2012; Hooshiar Kashani et al. 2012; Perego et al. 2012). Regional patterns have been characterized in tandem with reconstructing the emerging complexity of movements from Beringia/Siberia into the American continent. These findings are now informed by genome-wide analyses that are contributing to the ongoing discussion had been based on a large body of uniparental genetic data from present and past populations as well as archaeological and climatological data (Goebel et al. 2008; Misarti et al. 2012; Raff and Bolnick 2014; Reich et al. 2012). While details of the migration patterns into North America are still not completely resolved, the initial population movement into South America remains further elusive.

A rapid migration along the coast from North to South America may account for the early human occupation evidence at Monte Verde (Dillehay 1997; Dillehay et al. 2008), but overall the dispersal patterns of the first peoples in South America remain unclear. Two main scenarios have been proposed to explain the migration history: 1) a single migration along the coastal side of the Andes from which later groups spread out across the continent (Bennett and Bird 1964; Salzano and Callegari-Jacques 1988), and 2) an earlier split of the source population, possibly in the

northern area of South America that resulted in separate coastal and inland/continental populations (Bodner et al. 2012; Hubbe et al. 2010; Neves et al. 2003). Early genetic studies supported the second of these scenarios, two distinct populations based on the genetic patterns between Andean and Amazonian groups (Luiselli et al. 2000; Monsalve et al. 1994; Rothhammer and Silva 1992). However, it is important to note that the presence of genetic structuring does not imply separate migration events (Fuselli et al. 2003; Lewis et al. 2007; Wang et al. 2007) and studies have pointed out the possible underestimation of the impact from European contact that depopulated the region and likely led to a drastic genetic bottleneck (Lewis and Long 2008). Furthermore, the spread from coastal Andean regions to Amazonia has underpinnings in the early notion that Amazonian groups devolved from complex Andean cultures, a notion which has been firmly refuted in the last couple of decades (e.g. Heckenberger et al. 2008; Neves 1999; Stahl 2002).

Amazonia refers to not only the Amazon River basin but also the Orinoco River basin that serves as the homeland of the Yekuana. A Carib-speaking people, Yekuana translates to "water-log people" in their language, but they are also known as "Makiritare" that was used by the Spanish who adopted the term from their Arawak-speaking guides (Guss 1989). As the name suggests, the Yekuana have been known for their skills in the waterways that has been integral to their identity and history (Butt 1973; Guss 1989). Within a broader framework to examine the Yekuana as being shaped by and interacting with the dynamic sociocultural and physical environment, the genetic histories of the Yekuana were investigated utilizing uniparental and biparental genetic markers (Lee 2009). Here, we focus on the findings based on whole mtDNA genome sequences to investigate regional phylogenies in South America.

Recent mtDNA studies from South America have focused on characterizing specific regional phylogeographies. Mitochondrial DNA analysis in the Southern Cone region suggested subgroups such as D1g and B2i that occur in high frequencies in the region probably arose with the first migration wave into the continent but went through different phylogenetic trajectories (Bodner et al. 2012; de Saint Pierre et al. 2012). New lineages within haplogroups C1 and D1 have also been linked to local origins and continuity with prehistoric populations (Garcia et al. 2012; Sans et al. 2012). The Yekuana haplotypes add to the growing number of South American mitogenomes, particularly by expanding South American subhaplogroups (e.g. A2aa), and support the diversity of haplogroups C1 and D1 in the region.

Materials and Methods

Population samples

The Yekuana inhabit the area encompassing the northern bank of the upper Rio Orinoco to the upper regions of Rio Caura and Rio Ventuari within the states of Amazonas and Bolivar in Venezuela (Fig. 1). Blood samples were collected from thirteen Yekuana villages by James V. Neel and Richard Ward during the late 1960s and early 1970s in compliance with institutional review boards. Experimental procedures for this study were carried out at Binghamton University and were approved by the Human Subjects Research Review Committee (Protocol Number 597-07). Detailed genealogical information was recorded such that maternally unrelated individuals were selected for mtDNA analysis (Gershowitz et al. 1970). Control region mtDNA typing of 114 maternally unrelated individuals showed that haplotypes were assigned to the four Native Americans haplogroups A2, B2, C1, and D1. Preliminary results showed a higher frequency of haplogroup C1 among the Yekuana (61%) that resembles other Northern

Amazonian groups and a lower frequency of haplogroup B2 (5%), which is consistent with the broader genetic pattern in Amazonia as well as in South America (Easton et al. 1996; Lee 2009; Rodriguez-Delfin et al. 2001; Vona et al. 2005). All four haplogroups were identified in the village of Sharamaña and in order to further examine the four haplogroups in the broader context of South American phylogenies, all twelve haplotypes from the village were selected for whole mtDNA genome analysis.

Experimental procedures

DNA was extracted using the column based Qiagen DNA Extraction kit following manufacturer's instructions (Qiagen Inc, Valencia, CA). Twenty-four overlapping primers spanning the entire ~16,569 bp genome were used for whole mitochondrial genome PCR amplification and sequencing (Rieder et al. 1998). Amplification was verified by agarose gel electrophoresis and successful amplicons were purified on a 96-well Multiscreen HTS™ plate (Millipore, Billerica, MA) to remove single strand DNA and primers. Amplicons were prepared for sequencing using the Big Dye Terminator v.3.1 kit (Applied Biosystems, Foster City, CA). After standard alcohol precipitation, sequencing was carried out on an ABI PRISM™ 377XL DNA Sequencer (Applied Biosystems, Foster City, CA) following manufacturer's instructions.

Mutations were identified in comparison to the revised Cambridge Reference Sequence (Andrews et al. 1999) using Sequencher (GeneCodes Inc., Ann Arbor, MI). Native South American whole mtDNA genomes sequences were compiled from the published literature and database for statistical analysis (Bodner et al. 2012; Cardoso et al. 2012; de Saint Pierre et al. 2012; Fagundes et al. 2008; Gómez-Carballa et al. 2012; Ingman et al. 2000; Kivisild et al. 2006; Perego et al. 2010; Tamm et al. 2007). Sequences were assembled and aligned using Mega 5.2.2

6

(Tamura et al. 2011). A median-joining network for mitogenomes was constructed for each haplogroup (A2, B2, C1, and D1) using Network 4.6 (www.fluxus-engineering.com; Bandelt et al. 1999) and subhaplogroups were assigned and identified using the nomenclature from the most recent tree build, Build 16 (http://www.phylotree.org; van Oven and Kayser 2009). New phylogenies were constructed by hand following a maximum parsimony approach taking into the account the median-joining networks. Population statistical analyses were carried out with coding region sequences (np 577-16023) using Arlequin 3.5 (Excoffier and Lischer 2010) to estimate nucleotide diversity values and to assess demographic events (Tajima's and Fu's tests).

Results

Haplogroup characterization

The twelve mitogenomes from our study did not match any previously published genome sequence. Among the twelve mitogenome sequences, four were assigned to haplogroup A2, two to B2, five to C1, and one to D1. Sequences can be found under Genbank accession numbers KM656453-KM656464. While more than twenty branches of haplogroup A2 have been identified, characterization of new lineages has become complicated by reoccurring mutations (van Oven and Kayser 2009). One Yekuana sequence (F530) was assigned to A2h, which has been observed in North, Central, and South America (Fig. 2). In particular, A2h includes haplotypes from the Yanomamö, a neighboring group of the Yekuana from northern Brazil, and Kogui from northern Colombia (Fagundes et al. 2008). Classification of the other three haplotypes (F525, 528, 535) revealed possible reconfiguration of some branches. The current mtDNA tree build defines A2m by mutations at np 8947, 9039, 15172, and 16240, which includes two haplotypes that likely originated in Mexico (Kumar et al. 2011). Because the

mutation at np 16240 was observed in one Yekuana haplotype (F535), we suggest revising A2m so that it is defined by the mutation at np 16240 so that it would incorporate the Yekuana haplotype (Fig. 2). Subhaplogroup A2aa was characterized by mutations at np 16239 and 16266 and includes haplotypes from Brazil and Guayana (Fagundes et al. 2008). Two Yekuana haplotypes (F525, 528) harbored the mutation at np 16266, and we reconfigured this branch so that np 16266 defines A2aa and a new subhaplogroup A2aa1 is defined by np 16239 (Fig. 2).

Haplogroup B2 has been subdivided into subhaplogroups B2a - B2y and South American haplotypes are identified in B2b, B2d, B2e, and B2h. The two Yekuana haplotypes were identical and harbored the mutation at np 16311, which occurs multiple times in the phylogeny. Other haplotypes with the recurrent mutation 16311 included those from Colombia and southern Brazil that are currently unassigned to any specific subclade (Fagundes et al. 2008). Though perhaps tenuous, based on the shared presence among several South American haplotypes, we characterized the branch as subhaplogroup B2z, defined by the mutation at np 16311 (Fig. 2).

Native American lineages within haplogroup C1 are identified as C1b, C1c, and C1d. Three Yekuana haplotypes were assigned to C1b (F532, 533, 544), among which two did not share any of the characterized 12 subbranches of C1b. One Yekuana C1b haplotype (F532) had a mutation at np 16362 but lacked the mutation at np 15613 that defines C1b8. When excluding np 15613, another South American haplotype (ARL58) shared the mutation at np 16362 (Fagundes et al. 2008), which may then characterize a subbranch that also includes the Yekuana haplotype (Fig. 2). One C1c haplotype was observed in the Yekuana (F516), but did not fall into any of the known eight subclades. C1d is defined by a mutation at np 16051 and includes three subhaplogroups, C1d1-C1d3. One Yekuana haplotype (F521) had the back mutation at np 194 that characterizes C1d1a, which is observed among Mexican-American individuals, but did not

share the other mutations observed in those haplotypes (Kumar et al. 2011). Therefore, we left the haplotype as a new subbranch of C1d1.

Native American haplogroup D1 is currently divided into subhaplogroups D1a through D1n. The one Yekuana D1 haplotype (F539) could not be assigned to any of the characterized subhaplogroups (Fig. 2).

Population statistics

Nucleotide diversity values were calculated to characterize the variability based on haplogroup and continent (North, Central, South America). Tajima's and Fu's tests were conducted to infer demographic events including population bottlenecks and expansions (Schneider et al. 2000). Nucleotide diversity estimates for the South America continent (0.001741) showed similar values to North and Central America from a previous study (Kumar et al. 2011). Furthermore, our results were consistent with the significantly negative values for Tajima's D (-2.295) and Fu's Fs (-23.605) that would suggest population expansion (Table 1). There was no significant difference in nucleotide diversity estimates between each haplogroup in South America (0.00048 - 0.000523) and all haplogroups showed significantly negative Tajima's D and Fu's Fs values.

Discussion

Mitochondrial genomic data has provided higher resolution of Native American lineages, identifying pan-American as well as region-specific haplogroups. The twelve Yekuana haplotypes from our study further refines Native American mtDNA lineages and reveals interesting phylogenetic patterns. Subhaplogroup A2h appears to be pan-America and includes the branch A2h1 that is observed in North and Central America and haplotypes from eastern

South America (Brazil, Colombia, Venezuela), including the one from our study (F530). While the A2af branch has been defined as specific to Central America (Perego et al. 2012), no other branches of haplogroup A2 had been previously found to be exclusive to South America. Including the two Yekuana haplotypes, the subhaplogroup A2aa consists of only South American haplotypes, in particular from the northeast region (Brazil, Guayana, Venezuela). While this subhaplogroup may have origins in the first migrants that came into the continent directly moving from the Atlantic coast into the Amazon (Callegari-Jacques et al. 1994; Salzano and Callegari-Jacques 1988), further data from other South American groups will reveal whether it remains restricted to the region.

Within haplogroup B2, studies have suggested that B2j and B2k are specific to Venezuela (Gómez-Carballa et al. 2012) but Yekuana haplotypes from our study do not fall into those subclades. Three C1 haplotypes from the Yekuana are assigned to the pan-American C1 subhaplogroups C1b, C1c and C1d. Within haplogroup D1, D1f is observed across North, Central, and South America (Achilli et al. 2008; Kumar et al. 2011; Tamm et al. 2007). Subhaplogroups D1a and D1e seem to be exclusive to South America while D1g and D1j seems exclusive to the Southern Cone region of Chile and Argentina (Bodner et al. 2012; Fagundes et al. 2008; Kivisild et al. 2006). The sole Yekuana D1 haplotype was not assigned to any of these known subhaplogroups.

The number of South American whole mitogenome sequences has accumulated to over 200, though much of the number is from large-scale regional or haplogroup-specific studies (e.g. de Saint Pierre et al. 2012; Perego et al. 2010). Still, it is notable that the nucleotide diversity estimates from our study match the values for North and Central America that was previously reported (Kumar et al. 2011). Significantly negative Tajima's D and Fu's Fs values are consistent

with earlier studies, supporting a simultaneous spread of the earliest lineages into the three regions with rapid population expansion. Haplogroup diversity in South America show similar levels between the four major groups (0.00048 - 0.000523), which is in contrast to North America that shows higher diversity levels for haplogroups A2 and B2 (Kumar et al. 2011). The higher diversity levels of A2 and B2 in North America may reflect the complex population movements involving the earliest Americans as well as Paleo-Eskimos and Aleuts (Achilli et al. 2013).

Comparison of diversity estimates from whole genome studies and haplogroup frequencies from the larger control region database across the Americas (North, Central, South), it has been suggested that haplogroups C1 and D1 show a increasing clinal distribution from North to South America, possibly reaching fixation at the extreme south of the continent (García et al. 2006; Kumar et al. 2011; Lorenz and Smith 1996; Moraga et al. 2000). While diversity levels are similar across the four haplogroups in South America, diversity estimates of C1 and D1 in South America are higher than in North and Central America (Table 1). C1b, C1c and C1d are suggested to be independent founding lineages that arose at least as early as the other haplogroups (A2, B2 and D1) and the higher diversity values in South America may be attributed to a rapid diversification in the continent unique to those lineages, accompanied by multiple population splits and limited gene flow (Reich et al. 2012; Schurr and Sherry 2004; Tamm et al. 2007). On the other hand, an early differentiation of lineages within the continent has been associated with the higher diversity of D1 in South America and regional specific subhaplogroups (Bodner et al. 2012).

Ongoing identification of unique mtDNA lineages among South American groups is consistent with early genetic studies that showed regional "variants" using protein markers (Neel

1973, 1978). Rapid genetic diversification has been an early hypothesis in genetic studies of South American populations (Neel 1973), and in more recent time may have been associated with the great linguistic diversity that is present today (Campbell 1997; Loukotka 1968; Ramallo et al. 2013). The lack of substantial mitogenome data from Andean populations still leaves unresolved questions regarding any significant differentiation between east and west South America. Further studies from the underrepresented areas will enhance the higher resolution offered by whole genome sequences and continue to refine Native American phylogenies.

Literature Cited

Achilli, A., U.A. Perego, C.M. Bravi et al. 2008. The phylogeny of the four pan-American MtDNA haplogroups: implications for evolutionary and disease studies. *PLoS ONE* 3(3):e1764.

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(35):14308–14313.

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(2):147.

Bandelt, H.J., P. Forster, and A. Rohl. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16(1):37–48.

Bandelt, H.J., C. Herrnstadt, Y.G. Yao et al. 2003. Identification of Native American founder mtDNAs through the analysis of complete mtDNA sequences: some caveats. *Ann Hum Genet* 67(Pt 6):512–524.

Bennett, W.C., and J.B. Bird. 1964. Andean culture history. Garden City: The Natural History Press.

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(5):811–820.

Butt, A.J. 1973. Intertribal trade in the Guiana highlands. *Antropológica* 34:2–70.

Callegari-Jacques, S.M., F.M. Salzano, T.A. Weimer et al. 1994. Further blood genetic studies on Amazonian diversity--data from four Indian groups. *Ann Hum Biol* 21(5):465–481.

Campbell, L. 1997. American Indian languages: the historical linguistics of Native America. New York: Oxford University Press.

Cardoso, S., M.A. Alfonso-Sanchez, L. Valverde et al. 2012. Genetic uniqueness of the Waorani tribe from the Ecuadorian Amazon. *Heredity (Edinb)* 108(6):609–615.

de Saint Pierre, M., F. Gandini, U.A. Perego et al. 2012. Arrival of Paleo-Indians to the southern cone of South America: new clues from mitogenomes. *PLoS ONE* 7(12):e51311.

Dillehay, T.D. 1997. Monte Verde: a late Pleistocene settlement in Chile. Vol 2: The Archaeological Context and Interpretation. Washington: Smithsonian Institution Press.

Dillehay, T.D., C. Ramirez, M. Pino et al. 2008. Monte Verde: seaweed, food, medicine, and the peopling of South America. *Science* 320(5877):784–786.

Easton, R.D., D.A. Merriwether, D.E. Crews et al. 1996. mtDNA variation in the Yanomami: evidence for additional New World founding lineages. *Am J Hum Genet* 59(1):213–225.

Excoffier, L., and H.E. Lischer. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10(3):564–567.

Fagundes, N.J., R. Kanitz, R. Eckert et al. 2008. Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. *Am J Hum Genet* 82(3):583–592.

Fuselli, S., E. Tarazona-Santos, I. Dupanloup et al. 2003. Mitochondrial DNA diversity in South America and the genetic history of Andean highlanders. *Mol Biol Evol* 20(10):1682–1691.

García, F., M. Moraga, S. Vera et al. 2006. mtDNA microevolution in Southern Chile's archipelagos. *Am J Phys Anthropol* 129(3):473–481.

Garcia, A., M. Pauro, R. Nores et al. 2012. Phylogeography of mitochondrial haplogroup D1: An early spread of subhaplogroup D1j from Central Argentina. *Am J Phys Anthropol* 149(4):583–590.

Gershowitz, H., M. Layrisse, Z. Layrisse et al. 1970. Gene frequencies and microdifferentiation among the Makiritare Indians. I. Eleven blood group systems and the ABH-Le secretor traits: a note on Rh gene frequency determinations. *Am J Hum Genet* 22(5):515–525.

Goebel, T., M.R. Waters, and D.H. O'Rourke. 2008. The late Pleistocene dispersal of modern humans in the Americas. *Science* 319(5869):1497–1502.

Gómez-Carballa, A., A. Ignacio-Veiga, V. Álvarez-Iglesias et al. 2012. A melting pot of multicontinental mtDNA lineages in admixed Venezuelans. *Am J Phys Anthropol* 147(1):78–87.

Guss, D.M. 1989. To weave and sing: art, symbol, and narrative in the South American rain forest. Berkeley: University of California Press.

Heckenberger, M.J., J.C. Russell, C. Fausto et al. 2008. Pre-Columbian urbanism, anthropogenic landscapes, and the future of the Amazon. *Science* 321(5893):1214–1217.

Hooshiar Kashani, B., U.A. Perego, A. Olivieri et al. 2012. Mitochondrial haplogroup C4c: a rare lineage entering America through the ice-free corridor? *Am J Phys Anthropol* 147(1):35–39.

Hubbe, M., W.A. Neves, and K. Harvati. 2010. Testing evolutionary and dispersion scenarios for the settlement of the new world. *PLoS ONE* 5(6):e11105.

Ingman, M., H. Kaessmann, S. Pääbo et al. 2000. Mitochondrial genome variation and the origin of modern humans. *Nature* 408(6813):708–713.

Kivisild, T., P. Shen, D.P. Wall et al. 2006. The role of selection in the evolution of human mitochondrial genomes. *Genetics* 172(1):373–387.

Kumar, S., C. Bellis, M. Zlojutro et al. 2011. Large scale mitochondrial sequencing in Mexican Americans suggests a reappraisal of Native American origins. *BMC Evol Biol* 11:293.

Lee, E.J. 2009. Genetic histories of the Yekuana from southern Venezuela: perspectives from mitochondrial DNA, Y-chromosome, and autosomal DNA. Binghamton: Binghamton University. 218 p.

Lewis, C.M., Jr., B. Lizarraga, R.Y. Tito et al. 2007. Mitochondrial DNA and the peopling of South America. *Hum Biol* 79(2):159–178.

Lewis, C.M., Jr., and J.C. Long. 2008. Native South American genetic structure and prehistory inferred from hierarchical modeling of mtDNA. *Mol Biol Evol* 25(3):478–486.

Lorenz, J.G., and D.G. Smith. 1996. Distribution of four founding mtDNA haplogroups among Native North Americans. *Am J Phys Anthropol* 101(3):307–323.

Loukotka, C. 1968. Classification of South American Indian languages. Los Angeles: Latin American Center, University of California, Los Angeles.

Luiselli, D., L. Simoni, E. Tarazona-Santos et al. 2000. Genetic structure of Quechua-speakers of the Central Andes and geographic patterns of gene frequencies in South Amerindian populations. *Am J Phys Anthropol* 113(1):5–17.

Misarti, N., B.P. Finney, J.W. Jordan et al. 2012. Early retreat of the Alaska Peninsula Glacier Complex and the implications for coastal migrations of First Americans. *Quat Sci Rev* 48(0):1–6.

Monsalve, M.V., H. Groot de Restrepo, A. Espinel et al. 1994. Evidence of mitochondrial DNA diversity in South American aboriginals. *Ann Hum Genet* 58(Pt 3):265–273.

Moraga, M.L., 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(1):19–29.

Neel, J.V. 1973. Diversity within and between South American Indian tribes. *Israel Journal of Medical Sciences* 9:1216–1224.

Neel, J.V. 1978. Rare variants, private polymorphisms, and locus heterozygosity in Amerindian populations. *Am J Hum Genet* 30(5):465–490.

Neves, E. 1999. Changing perspectives in Amazonian archaeology. In: *Archaeology in Latin America,* G.G. Politis, and B. Alberti, eds. London: Routledge. p 216–243.

Neves, W.A., A. Prous, R. Gonzalez-Jose et al. 2003. Early Holocene human skeletal remains from Santana do Riacho, Brazil: implications for the settlement of the New World. *J Hum Evol* 45(1):19–42.

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):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(9):1174–1179.

Perego, U.A., H. Lancioni, M. Tribaldos et al. 2012. Decrypting the mitochondrial gene pool of modern Panamanians. *PLoS ONE* 7(6):e38337.

Raff, J.A., and D.A. Bolnick. 2014. Palaeogenomics: genetic roots of the first Americans. *Nature* 506(7487):162–163.

Ramallo, V., R. Bisso-Machado, C. Bravi et al. 2013. Demographic expansions in South America: enlightening a complex scenario with genetic and linguistic data. *Am J Phys Anthropol* 150(3):453–463.

Reich, D., N. Patterson, D. Campbell et al. 2012. Reconstructing Native American population history. *Nature* 488(7411):370–374.

Rieder, M.J., S.L. Taylor, V.O. Tobe et al. 1998. Automating the identification of DNA variations using quality-based fluorescence re-sequencing: analysis of the human mitochondrial genome. *Nucleic Acids Res* 26(4):967–973.

Rodriguez-Delfin, L.A., V.E. Rubin-de-Celis, and M.A. Zago. 2001. Genetic diversity in an Andean population from Peru and regional migration patterns of Amerindians in South

America: data from Y chromosome and mitochondrial DNA. *Hum Hered* 51(1–2):97–106.

Rothhammer, F., and C. Silva. 1992. Gene geography of South America: testing models of population displacement based on archeological evidence. *Am J Phys Anthropol* 89(4):441–446.

Salzano, F.M., and S.M. Callegari-Jacques. 1988. South American Indians: a case study in evolution. Oxford: Clarendon Press.

Sans, M., G. Figueiro, and P.C. Hidalgo. 2012. A new mitochondrial C1 lineage from the prehistory of Uruguay: population genocide, ethnocide, and continuity. *Hum Biol* 84(3):287–305.

Schneider, S., D. Roessli, and L. Excoffier. 2000. Arlequin: a software for population genetics data analysis. 2.00 ed: Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva.

Schurr, T.G., and S.T. Sherry. 2004. Mitochondrial DNA and Y chromosome diversity and the peopling of the Americas: evolutionary and demographic evidence. *Am J Hum Biol* 16(4):420–439.

Stahl, P.W. 2002. Paradigms in Paradise: revising standard Amazonian prehistory. *The Review of Archaeology* 23(2):39–50.

Tamm, E., T. Kivisild, M. Reidla et al. 2007. Beringian standstill and spread of Native American founders. *PLoS ONE* 2(9):e829.

Tamura, K., D. Peterson, N. Peterson et al. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28(10):2731–2739.

Torroni, A., T.G. Schurr, C.C. Yang et al. 1992. Native American mitochondrial DNA analysis

indicates that the Amerind and the Nadene populations were founded by two independent

migrations. *Genetics* 130(1):153–162.

van Oven, M., and M. Kayser. 2009. Updated comprehensive phylogenetic tree of global human

mitochondrial DNA variation. *Hum Mutat* 30(2):E386–394.

Vona, G., A. Falchi, P. Moral et al. 2005. Mitochondrial sequence variation in the Guahibo

Amerindian population from Venezuela. *Am J Phys Anthropol* 127(3):361–369.

Wang, S., C.M. Lewis, M. Jakobsson et al. 2007. Genetic variation and population structure in

native Americans. *PLoS Genet* 3(11):e185.

Region	N	Nucleotide Diversity	Tajima's D	Fu's Fs	Reference
Diversity within entire geographic region					
North America	73	0.001747 (0.000858)	-2.042	-24.041	Kumar et al. 2011
Central America	139	0.001735 (0.000846)	-2.195	-23.787	Kumar et al. 2011
South America	243	0.001742 (0.000847)	-2.295	-23.605	this study
Within haplogroup diversity					
Haplogroup A2					
North America	29	0.000615 (0.000323)	-2.569	-21.684	Kumar et al. 2011
Central America	86	0.000495 (0.000258)	-2.816	-24.954	Kumar et al. 2011
South America	35	0.000506 (0.000268)	-2.57	-8.961	this study
Haplogroup B2					
North America	18	0.000668 (0.000358)	-2.471	-10.079	Kumar et al. 2011
Central America	55	0.000539 (0.000281)	-2.598	-24.826	Kumar et al. 2011
South America	53	0.000480 (0.000253)	-2.504	-24.981	this study
Haplogroup C1					
North America	14	0.000487 (0.000271)	-1.994	-5.519	Kumar et al. 2011
Central America	61	0.000501 (0.000262)	-2.51	-24.908	Kumar et al. 2011
South America	89	0.000523 (0.000272)	-2.711	-24.770	this study
Haplogroup D1					
North America	12	0.000363 (0.000211)	-2.125	-7.565	Kumar et al. 2011
Central America	17	0.000425 (0.000237)	-2.235	-12.457	Kumar et al. 2011

South America	66	0.000515 (0.000269)	-2.344	-23.605	this study

Table 1. Diversity indices and neutrality tests for Native American haplogroups within North, Central, and South America based on mitochondrial DNA coding region (np 577-16023) sequences.

Figure 1. Map of Yekuana territory in southern Venezuela. The star indicates the location of the
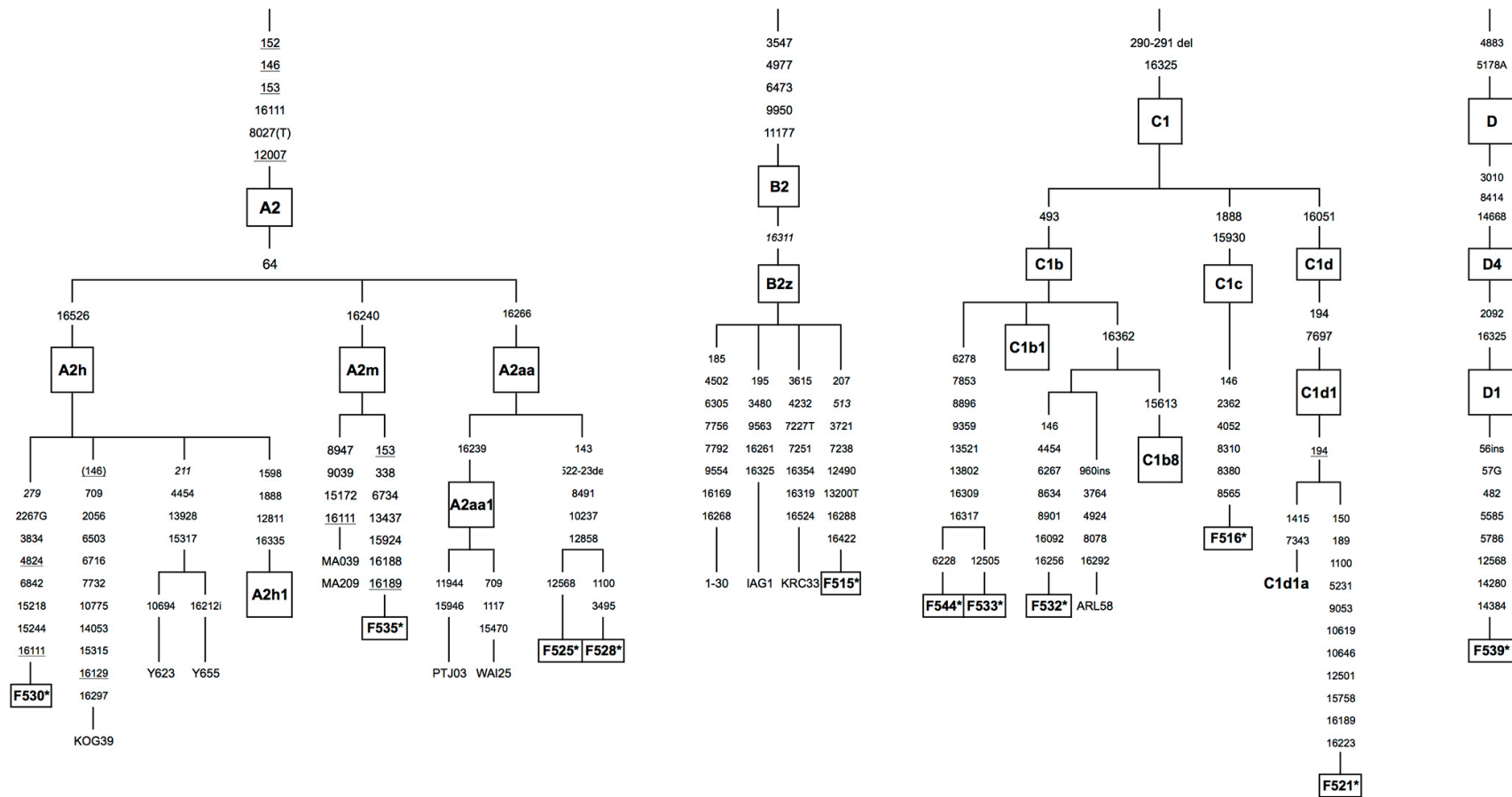
Yekuana village Sharamaña.
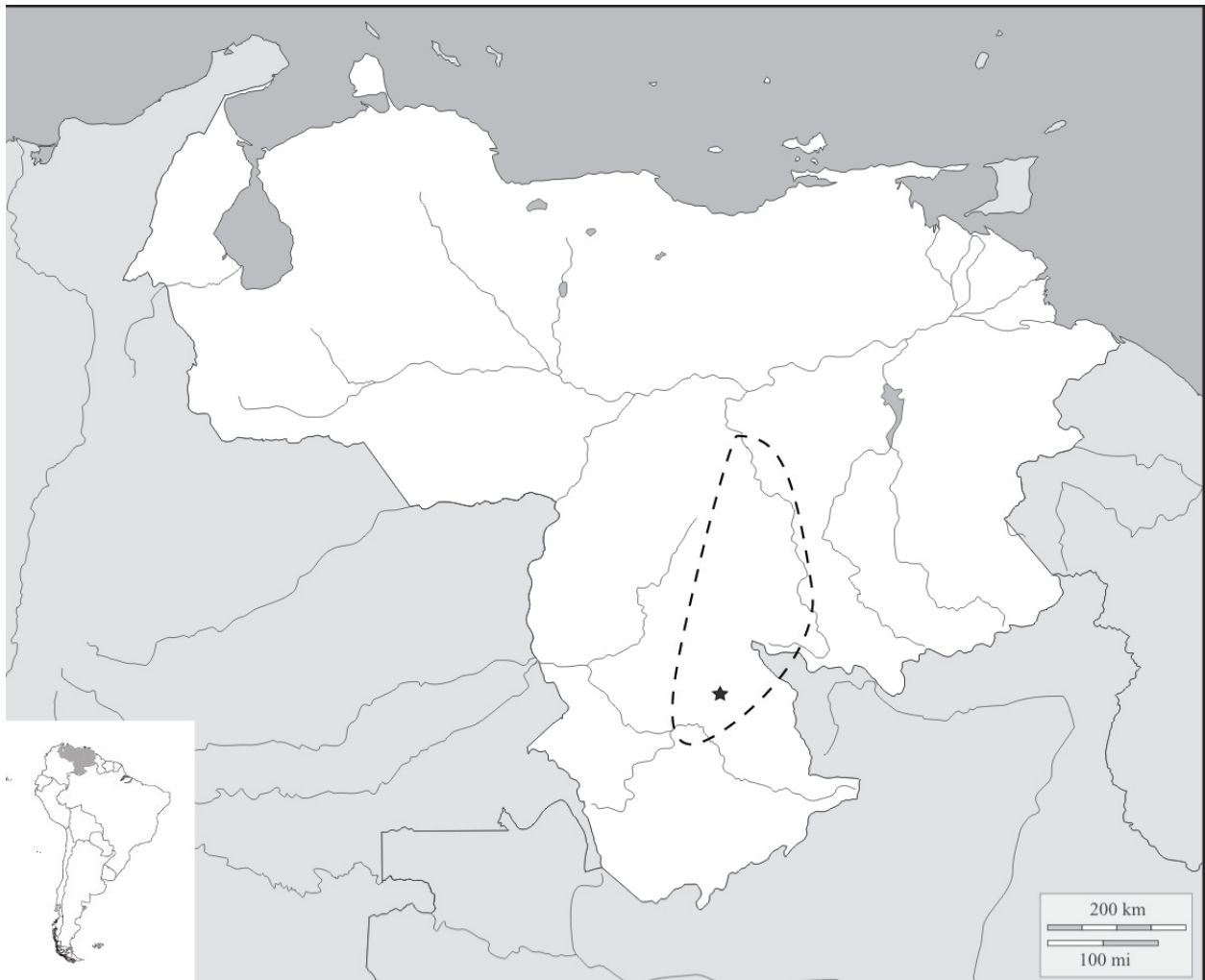
Figure 2. Schematic tree of mitochondrial haplogroups A2, B2, C1 and D1 with Yekuana mitogenomes from this study. Yekuana haplotypes are boxed and indicated by the * that follows the ID (F515, F516, F521, F525, F528, F530, F532, F533, F535, F539, and F544). The nucleotide substitutions are listed relative to the revised Cambridge Reference Sequence. Suffixes A, C, G, and T indicate transversions, "del" signifies a deletion and "ins" indicates an insertion. Back mutations are underlined and recurrent mutations are

italicized. The A/C stretch length polymorphism in regions 16180–16193 and 303–315, 522–523 and mutation 16519, all known to be hypervariable, were disregarded for tree reconstruction. Information on the other haplotypes included are: KOG39 (EU095545, Tamm et al. 2007), Y623 (EU095202, Fagundes et al. 2008), Y655 (EU095200, Fagundes et al. 2008), MA039 (HQ012062, Kumar et al. 2011), MA209 (HQ012131, Kumar et al. 2011), PTJ03 (EU095201, Fagundes et al. 2008), and WAI25 (EU095196, Fagundes et al. 2008), 1-30 (EU095535, Tamm et al. 2007), IAG1 (EU431084, Achilli et al. 2008), KRC33 (EU095219, Fagundes et al. 2008), and ARL58 (EU095227, Fagundes et al. 2008).