Received: 20 February 2017 | Revised: 17 September 2017 | Accepted: 7 October 2017

DOI: 10.1002/ajpa.23345

RESEARCH ARTICLE

WILEY American Journal of PHYSICAL ANTHROPOLOGY

High-resolution mitochondrial DNA analysis sheds light on human diversity, cultural interactions, and population mobility in Northwestern Amazonia

Leonardo Arias[1,3] | Chiara Barbieri[2] | Guillermo Barreto[3] | Mark Stoneking[1] | Brigitte Pakendorf[4]

[1]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig D-04103, Germany

[2]Department of Linguistic and Cultural Evolution, Max Planck Institute for the Science of Human History, Jena D-07745, Germany

[3]Laboratorio de Genética Molecular Humana, Universidad del Valle, Cali, Colombia

[4]Dynamique du Langage, UMR5596, CNRS & Université de Lyon, Lyon Cedex 07 69363, France

Correspondence
Leonardo Arias, Department of Evolutionary Genetics, Max-Planck-Institute for Evolutionary Anthropology, Deutscher Platz 6, Leipzig, D-04103, Germany.
Email: leoarias2@gmail.com

Funding information
Departamento Administrativo de Ciencia, Tecnología e Innovación (COLCIENCIAS); Max Planck Society

Abstract

Objectives: Northwestern Amazonia (NWA) is a center of high linguistic and cultural diversity. Several language families and linguistic isolates occur in this region, as well as different subsistence patterns, with some groups being foragers and others agriculturalists. In addition, speakers of Eastern Tukanoan languages are known for practicing linguistic exogamy, a marriage system in which partners are taken from different language groups. In this study, we use high-resolution mitochondrial DNA sequencing to investigate the impact of this linguistic and cultural diversity on the genetic relationships and population structure of NWA groups.

Methods: We collected saliva samples from individuals representing 40 different NWA ethnolinguistic groups and sequenced 439 complete mitochondrial genomes to an average coverage of 1,030×.

Results: The mtDNA data revealed that NWA populations have high genetic diversity with extensive sharing of haplotypes among groups. Moreover, groups who practice linguistic exogamy have higher genetic diversity, while the foraging Nukak have lower genetic diversity. We also find that rivers play a more important role than either geography or language affiliation in structuring the genetic relationships of populations.

Discussion: Contrary to the view of NWA as a pristine area inhabited by small human populations living in isolation, our data support a view of high diversity and contact among different ethnolinguistic groups, with movement along rivers probably facilitating this contact. Additionally, we provide evidence for the impact of cultural practices, such as linguistic exogamy, on patterns of genetic variation. Overall, this study provides new data and insights into a remote and little-studied region of the world.

KEYWORDS
haplogroup, South America, language, exogamy, Amazonia

1 | INTRODUCTION

Northwestern Amazonia (NWA) contains tremendous biological, linguistic, and cultural diversity, which likely reflects the heterogeneity of the landscape, especially the complex and extensive network of rivers found in this area. The region (Figure 1) extends from the Andean foothills in the west to the area between the Orinoco River and the Rio Negro in the east, and extends south until the confluence between the Rio Negro and the Amazon River. The northern border is defined by the Eastern Andean Cordillera and the Colombian-Venezuelan llanos, and the southern by the full length of the Putumayo River (Eriksen, 2011).

In terms of linguistic diversity, NWA harbors ethnolinguistic groups belonging to the main South American language families accepted by

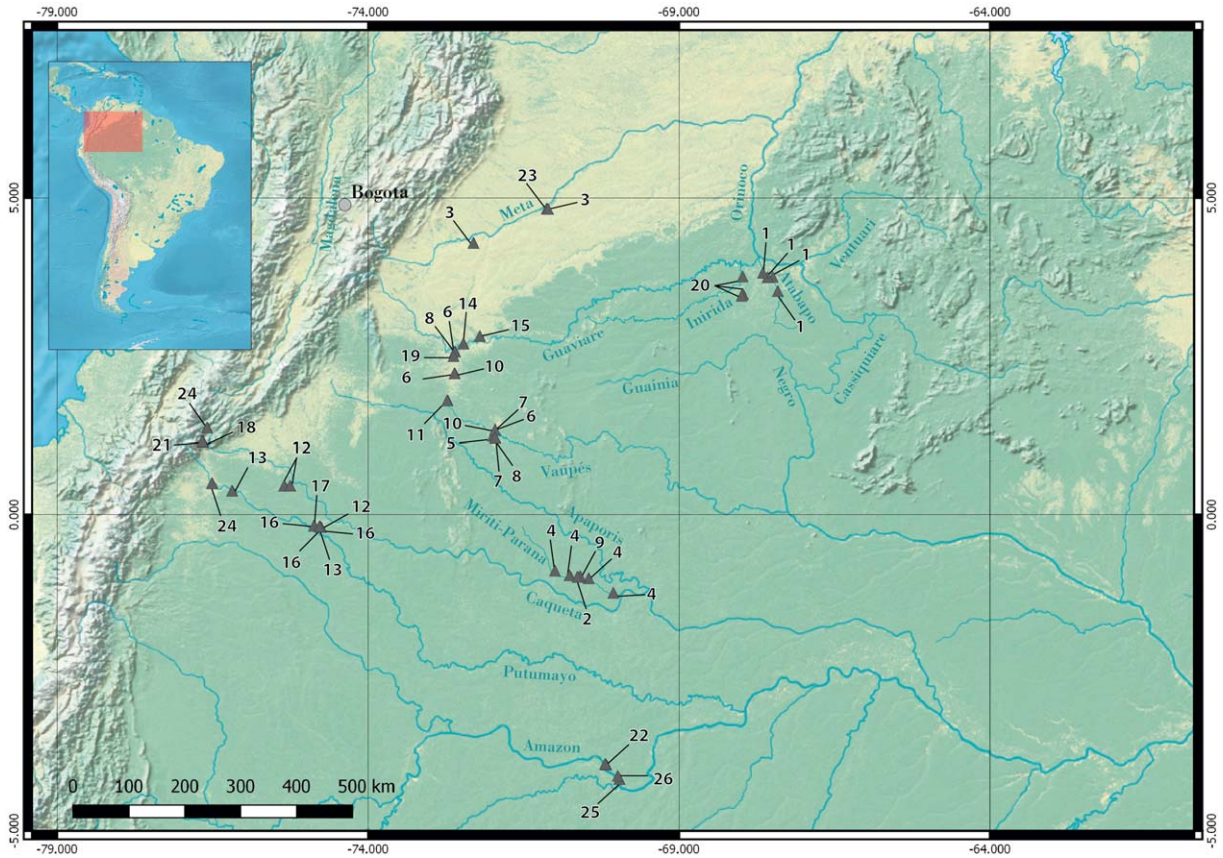

FIGURE 1 Geographic location of the sampling sites. Every triangle corresponds to a single community, which may contain more than one ethnolinguistic group. 1. Curripaco and Bare, 2. Matapi, 3. Ach-Piapoco, 4. Yucuna, 5. Carijona, 6. Desano, Yuruti, Pisamira, and Karapana, 7. Pira-Wanano, 8. Siriano, 9. Tanimuka, 10. Tukano, 11. Tuyuca and Tatuyo, 12. Coreguaje, 13. Siona, 14. Guayabero, 15. Sikuani, 16. Murui, 17.Uitoto, 18. Kamentsa, 19. Nukak, 20. Puinave, 21. Pasto, 22. Yagua, 23. Saliba, 24. Inga, 25. Tikuna, 26. Cocama

most linguists (Campbell, 1997; Chacon, 2014; Dixon and Aikhenvald, 1999), namely Arawakan, Carib, Tupi, and Quechua. Additionally, several local language families are also present, such as Tukanoan, Guahiban, Huitotoan, Boran, Peba-Yaguan, Piaroa-Saliban, and Maku-Puinave (see glottolog.org for a classification of the language families), as well as various isolate languages like Tikuna, Cofan, and Kamentsa (Landaburu, 2000). Furthermore, several indigenous groups live in voluntary isolation, and almost nothing is known about their linguistic affiliation (Franco, 2002). The area has been proposed as the place of origin of the Arawakan family, since it contains the highest linguistic diversity within the family (Aikhenvald, 1999; Heckenberger, 2002; Zucchi, 2002). In addition, all 20 languages of the Tukanoan family are found in the area. These are classified into two branches: the Western Tukanoan branch distributed along the Putumayo, Caquetá, and Napo Rivers and the Eastern Tukanoan branch along the Vaupés, Rio Negro, and Apaporis Rivers and their tributaries (Chacon, 2014). The Carib, Tupi, and Quechua language families are probably recent immigrants to NWA, since only one language per family is present in the area, while the majority of languages within these families are found elsewhere. In addition, the Tupi language Nheengatú or Lingua Geral is found in the region; however, this is a very recent introduction spread by missionaries during the 17th and 18th centuries and by traders during the rubber boom in the 19th century, when it was used as a trade language (Sorensen, 1967; Stenzel, 2005).

In terms of cultural diversity, while NWA has often been viewed as a pristine area inhabited only by small, isolated, seminomadic tribes with an economy based on hunting and gathering (Denevan, 1992; Meggers, 1954), in fact there is considerable variation in subsistence and marriage practices. Although some groups are traditional foragers, others engage in agriculture, and instead of being isolated, archeological and anthropological evidence now shows that NWA was indeed part of a continent-wide network of exchange and trade. Complex societies organized in chiefdoms and multiethnic confederations arose in the region, and multilingualism and extensive interactions among ethnolinguistic groups were the norm (Heckenberger, 2002; Hornborg, 2005; Santos-Granero, 2002; Vidal, 1997).

In particular, the groups speaking Eastern Tukanoan languages and some of their Arawakan neighbors living in the basin of the Vaupés River and Rio Negro engage in an exceptional marital practice known as linguistic exogamy (Aikhenvald, 1996; Chacon and Cayón, 2013; Sorensen, 1967; Stenzel, 2005). According to this cultural norm, individuals are required to marry someone from a different language group, with each individual's linguistic affiliation being determined by the language of the father. Linguistic exogamy thus creates a situation of

multilingualism and movement of people (particularly women, since it is accompanied by patrilocality and patrilineality) among the groups participating in the system (Sorensen, 1967).

Historical linguistics, cultural anthropology, and archeology are the main disciplines that have traditionally addressed questions regarding the origins, pre-history, and genetic relationships of NWA ethnolinguistic groups (Campbell, 1997; Chacon, 2014; Heckenberger, 2008; Lathrap, 1970; Meggers, 1948; Nettle, 1999). However, due to the incomplete archeological record, the limitations of linguistic methods based on lexical cognates to establish deep time relationships (Dediu and Levinson, 2012; Hock and Joseph, 2009), and the insufficiency of documentation and description of a large number of the NWA societies, many of these questions remain to be fully answered. The oldest archeological evidence of human occupation in NWA comes from a single site on the Middle Caquetá River, which has been dated between 9250 and 8100 BP. It contains a great variety of stone artifacts, carbonized seeds and other botanical remains from different palm species, as well as phytoliths of bottle gourd, leren, and pumpkin (Aceituno, Loaiza, Delgado-Burbano, & Barrientos, 2013; Gnecco and Mora, 1997), indicating that these early human groups relied on vegetable resources that are still being exploited by contemporary societies in NWA.

One hypothesis about the peopling of NWA was proposed by Nimuendajú (1950), who suggested that the region was first inhabited by hunter-gatherer populations (HGPs), perhaps the ancestors of the Maku-Puinave groups, most of whom still practice a foraging lifestyle. Proto-Arawakan groups then started expanding into the region from their place of origin located between the Orinoco River and the Rio Negro (Heckenberger, 2002; Lathrap, 1970). Finally, the Tukanoans are assumed to have arrived in the area and displaced peoples speaking Arawakan and Maku-Puinave languages from the Vaupés (the Tukanoans probably came from the Napo-Putumayo, where Western Tukanoans still live). However, this scenario does not account for the presence of groups belonging to the Carib, Guahiban, Huitotoan, and Boran language families and the various language isolates in the region.

Genetic studies can provide insights into population history, and indeed studies of mitochondrial DNA (mtDNA) genetic variation in Native American populations have contributed greatly to our knowledge about the peopling of the Americas. Early studies using restriction fragment length polymorphisms (RFLP) and sequencing of the hypervariable region one (HVS-I) identified five founder lineages or haplogroups, designated as A–D and X (Bailliet, Rothhammer, Carnese, Bravi, & Bianchi, 1994; Barbieri, Heggarty, Castri, Luiselli, & Pettener, 2011; Gaya-Vidal et al., 2011; Keyeux, Rodas, Gelvez, & Carter, 2002; Lewis et al., 2007; Schurr, 2004; Torroni et al., 1993). Whereas haplogroups A–D are widely distributed in the Americas, haplogroup X is restricted to North America (Bolnick and Smith, 2003; Malhi, Schultz, & Smith, 2001). The analysis of HVS-I in several Native American populations showed that haplogroups A-D exhibit similar levels of diversity (Bonatto and Salzano, 1997), supporting the hypothesis of a single origin of all Native American populations from a Northeast Asian source. Additionally, HVS-I data have been used to determine the genetic relationships among indigenous populations in South America and to test hypotheses concerning how genetic variation is structured at the regional and continental levels (Barbieri et al., 2011; Gaya-Vidal et al., 2011; Lewis et al., 2007; Marrero et al., 2007; Melton et al., 2007). These studies revealed that Andean (or western) populations show higher levels of diversity and low genetic distances in contrast to the Eastern populations, who show the opposite pattern. However, in previous studies, NWA populations have been generally underrepresented, and hence inferences about the genetic structure of the entire Amazonian region are based on data from a small number of populations.

Recent developments in sequencing technology allow the determination of complete mtDNA genomes at the population level and thus enable unbiased insights into the maternal history of human populations (Delfin et al., 2014; Gunnarsdottir, Li, Bauchet, Finstermeier, & Stoneking, 2011; Kivisild, 2015). At present, no such studies are reported for South American indigenous populations. Available studies of complete mtDNA genomes from Native Americans have been restricted to a limited number of individuals carrying particular haplogroups, usually selected based on their HVS-I sequences (Achilli et al., 2013; Bodner et al., 2012; de Saint Pierre et al., 2012; Fagundes et al., 2008; Lee and Merriwether, 2015; Perego et al., 2009, 2010), or to archeological remains from different time periods (Fehren-Schmitz et al., 2015; Llamas et al., 2016). These studies have primarily focused on inferences about the peopling of the continent, the number of migrations, the divergence times of haplogroups and changes in the effective population size through time.

Nevertheless, several problems and biases are associated with this sampling strategy. First, the overall diversity might be underestimated, since individuals carrying the same HVS-I sequence can exhibit considerable variation in the coding region (Gunnarsdottir et al., 2011). Second, the reconstruction of demographic trends can be skewed, since the estimation of effective population sizes through time using Bayesian coalescent methods (i.e., Bayesian skyline plots [BSPs] in BEAST) can generate spurious signals of population growth when based on samples selected by haplogroup (Gunnarsdottir et al., 2011). Last, the histories and origins of specific populations cannot be investigated, since the coalescent age of a particular lineage does not correspond to the age of the population, especially when the diversity within each lineage is unknown (Schurr, 2004).

In this study, we use complete mtDNA sequencing in a large and representative sample of populations covering the extant ethnolinguistic diversity from NWA to reconstruct their maternal history, as well as to determine their genetic diversity and to make inferences about the origins of this diversity. Finally, we aim to investigate the impact of pre-historic population dynamics and cultural interactions on the structure of the genetic variation observed among present-day NWA populations.

2 | MATERIALS AND METHODS

2.1 | Sample collection

Samples from unrelated individuals belonging to 40 ethnolinguistic groups were collected during several expeditions carried out by one of

TABLE 1 Sampled ethnolinguistic groups with information on merged groups (see "Population Samples" Section) given below the compound names.

Population	Label in Figure 1	*n*	Census size[a]	Language family	Subsistence strategy[b]	River/place of residence
Yucu-Matapi		39		Arawakan	AG	Miritíi-Paraná
Yucuna	4	31	550	Arawakan	AG	Miritíi-Paraná
Matapi	2	8	220	Arawakan	AG	Miritíi-Paraná
Curripaco		17		Arawakan	AG	Atabapo
Curripaco	1	16	7,827	Arawakan	AG	Atabapo
Bare	1	1	NA[c]	Arawakan	AG	Atabapo
Ach-Piapoco		24		Arawakan	AG	Meta
Achagua	3	6	283	Arawakan	AG	Meta
Piapoco	3	18	4,926	Arawakan	AG	Meta
Cabiyari[d]		1	311	Arawakan	AG	Miritíi-Paraná
Carijona	5	8	307	Carib	AG	Upper-Vaupés
Cofan		6	877	Cofan	AG	Guamúez
Barasano		4	2,008	Eastern Tukanoan	AG	Upper-Vaupés
Desano	6	17	2,457	Eastern Tukanoan	AG	Upper-Vaupés
Kubeo		5	6,647	Eastern Tukanoan	AG	Upper-Vaupés
Other-ET		10		Eastern Tukanoan	AG	Upper-Vaupés
Tuyuca	11	7	642	Eastern Tukanoan	AG	Upper-Vaupés
Yuruti	6	1	687	Eastern Tukanoan	AG	Upper-Vaupés
Pisamira	6	1	61	Eastern Tukanoan	AG	Upper-Vaupés
Karapana	6	1	464	Eastern Tukanoan	AG	Upper-Vaupés
Pira-Wanano		13		Eastern Tukanoan	AG	Upper-Vaupés
Piratapuyo	7	8	697	Eastern Tukanoan	AG	Upper-Vaupés
Wanano	7	5	1,395	Eastern Tukanoan	AG	Upper-Vaupés
Siriano	8	10	749	Eastern Tukanoan	AG	Upper-Vaupés
Tanimuka	9	10	1,247	Eastern Tukanoan	AG	Miritíi-Paraná
Tuka-Tatuyo		10		Eastern Tukanoan	AG	Upper-Vaupés
Tukano	10	8	6,996	Eastern Tukanoan	AG	Upper-Vaupés
Tatuyo	11	2	331	Eastern Tukanoan	AG	Upper-Vaupés
Siona	13	17	734	Western Tukanoan	AG	Putumayo
Coreguaje	12	19	2,212	Western Tukanoan	AG	Caquetá
Sikuani	15	16	23,006	Guahiban	HGP	Guaviare
Guayabero	14	35	1,118	Guahiban	HGP	Guaviare
Saliba	23	16	1,929	Piaroa-Saliban	AG	Meta
Mur-Uitoto[e]		26	7,343	Huitotoan	AG	Putumayo
Murui	16	18		Huitotoan	AG	Putumayo
Uitoto	17	8		Huitotoan	AG	Putumayo
Puinave	20	19	6,604	Maku-Puinave	HGP	Inirida
Nukak	19	16	1,483	Maku-Puinave	HGP	Interfluvial
Pasto	21	14	69,789	Pasto	AG	Andean
Kamentsa	18	11	4,773	Kamentsa	AG	Andean
Inga	24	17	19,079	Quechuan	AG	Andean
Tikuna	25	18	7,102	Tikuna	AG	Amazonas
Cocama	26	17	792	Tupi	AG	Amazonas
Yagua	22	13	297	Peba-Yaguan	AG	Amazonas

(Continues)

TABLE 1 (Continued)

Population	Label in Figure 1	*n*	Census size[a]	Language family	Subsistence strategy[b]	River/place of residence
Guambiano		1	23,462	Barbacoan	AG	Andean
Nasa		1	138,501	Nasa	AG	Andean
Mestizo		9	NA	Mestizo[f]	NA	NA
Total		439				

[a]Adapted from Arango and Sánchez (2004).
[b]AG, agriculturalist; HGP, Hunter-gatherer populations, data from D-PLACE (Kirby et al., 2016) and HG (https://huntergatherer.la.utexas.edu/home, accessed on June 6, 2017).
[c]NA: data not available.
[d]Populations with label in italics were not considered in the population-based analyses.
[e]Census data reports the population size including groups that speak five dialectal varieties.
[f]Mestizo is an autonym used by people of mixed ancestry.

the authors (L.A.) in five departments (administrative divisions) of NWA, namely: Amazonas, Guainía, Guaviare, Meta and Putumayo (Table 1, Figure 1). The samples consisted of either saliva (n = 400), collected as 3 mL of saliva in 3 mL of lysis buffer (Quinque, Kittler, Kayser, Stoneking, & Nasidze, 2006), or blood samples ($n = 60$) stabilized with EDTA. Written informed consent was obtained from each participant, and from the community leader and/or local/regional indigenous organizations, after giving a full description of the aims of the study. Local translators and fieldwork assistants helped to explain and translate into the local languages when individuals or communities were not proficient in Spanish. Additionally, each participant answered a short questionnaire soliciting information regarding their birthplace, language, ethnic affiliation and that of their parents and grandparents. The study was approved by the ethics committee of the Universidad del Valle in Cali, Colombia and the Ethics Commission of the University of Leipzig Medical Faculty. All procedures were undertaken in accordance with the Declaration of Helsinki on ethical principles and an export permit was issued by the Colombian Ministry of Health and Social Protection.

2.2 | DNA sequencing and sequence processing

The DNA was extracted from blood samples with the "salting out" method (Miller, Dykes, & Polesky, 1988) and from the saliva samples with the QIAamp DNA Midi kit (Qiagen), starting from 2.0 mL of the saliva/buffer mixture. The concentration of DNA was quantified with a NanoDrop 8000 spectrophotometer (Thermo Scientific). We prepared genomic libraries with double indices and enriched for full mtDNA genomes using a hybridization-capture method described previously in Kircher, Sawyer, and Meyer (2012) and Maricic, Whitten, and Paabo (2010). From the enriched libraries, paired-end sequences of 100 bp length were generated on the Illumina Hiseq 2500 platform. Base-calling was performed using freeIbis (Renaud, Kircher, Stenzel, & Kelso, 2013), and Illumina adapters were trimmed and completely overlapping paired sequences were merged using leeHOM (Renaud, Stenzel, & Kelso, 2014a). The sequencing data were de-multiplexed using deML (Renaud, Stenzel, Maricic, Wiebe, & Kelso, 2014b) and the sequences aligned against the human reference genome 19 using BWA's *aln* algorithm (Li and Durbin, 2009). After duplicate removal using PicardTools v2.1.1 (https://github.com/broadinstitute/picard), we performed an iterative alignment for each library individually to obtain mtDNA consensus sequences. In the first step, we extracted all sequencing reads of a library that aligned either to the mitochondrial genome or to a list of nuclear copies of mtDNA (Li, Schroeder, Ko, & Stoneking, 2012). We subsequently aligned these reads to the revised Cambridge Reference Sequence (rCRS; Andrews et al. 1999) using BowTie2's *very-sensitive* algorithm (Langmead and Salzberg, 2012) and called a consensus sequence. In the second step, the reads were re-aligned to the library's respective consensus sequence generated in the first step, using the same BowTie2 settings. After the second alignment step, we called a final consensus sequence that was used throughout the rest of the analysis. Final sequences in fasta format were aligned to the rCRS (Andrews et al., 1999) with the multiple sequence alignment software Mafft (Katoh and Standley, 2013), and manually inspected for alignment errors with Bioedit ver. 7.2.5 (Hall, 1999). The two poly-C regions (np 303–315 and 16,183–16,194) were excluded from the subsequent analyses. Although one position (16,189) diagnostic for haplogroup B2 is therefore not considered in the haplogroup calling analysis, the additional substitutions defining this haplogroup that occur elsewhere in the mitochondrial genome enable unambiguous assignment of sequences to this lineage.

2.3 | Population samples

We considered populations with a sample size of 10 individuals or more, and merged populations with sample sizes smaller than 10 based on linguistic criteria when our initial analyses did not show significant genetic differences, as follows (Table 1). The Arawakan groups Achagua ($n = 6$) and Piapoco ($n = 18$) were merged into a single population, since their indigenous reservations are adjacent and individuals often intermarry (data available on request); one Bare ($n = 1$) individual was added to the Curripaco ($n = 16$) sample among whom he was living when sampled on the Atabapo River; the Yucuna ($n = 31$) and Matapi ($n = 8$) were merged into a single population, since they both speak Yucuna, live along the same river, and intermarry (data available on request); and the Murui ($n = 18$) and Uitoto ($n = 8$) were merged, as these two groups belong to the same language family, which is composed of several dialects that are mutually intelligible (http://glottolog.org/resource/languoid/id/huit1251, accessed on May 31, 2017). Finally,

following the latest classification of the Tukanoan family (Chacon, 2014), the Eastern Tukanoan groups Piratapuyo ($n = 8$) and Wanano ($n = 5$) were merged as Pira-Wanano; Tukano ($n = 8$) and Tatuyo ($n = 2$) were merged as Tuka-Tatuyo; and Tuyuca ($n = 7$), Yuruti ($n = 1$), Pisamira ($n = 1$), and Karapana ($n = 1$) were merged as Other-ET. The only group with a sample size smaller than 10 that we retained as a separate group in the analyses were the Carijona ($n = 8$), since this is the only Carib-speaking group living in NWA. Moreover, they are at risk of disappearing both physically and culturally, with <30 active speakers of Carijona scattered in two communities, and they occupy an important place in the ethno-history of the region (Franco, 2002). We excluded Barasano ($n = 4$), Kubeo ($n = 5$), Cofan ($n = 6$), Cabiyari ($n = 1$), Guambiano ($n = 1$), and Nasa ($n = 1$) individuals from all the analyses except the haplotype networks, since this analysis represents the evolutionary relationships among individual sequences. We furthermore excluded nine individuals with maternal ancestry tracing outside of NWA (labeled "Mestizo" in Table 1) from all analyses. After merging and filtering as described above, 412 sequences from 24 groups were kept in the population-based analyses.

2.4 | Data analysis

Based on information from D-PLACE (Kirby et al. 2016) and HG database (https://huntergatherer.la.utexas.edu/home, accessed on June 6, 2017), we divided the populations into agriculturalists (AG) and HGP. In the latter category, we placed the Nukak, who currently still practice a foraging way of life, as well as the Puinave, Sikuani, and Guayabero, who have all adopted agriculture only very recently (Kondo, 2002; Uribe Tobón and Instituto Colombiano de Cultura, 1992).

The haplogroup affiliation of the individual sequences was determined with Haplogrep (Kloss-Brandstatter et al., 2011), based on Phylotree build 16 (van Oven and Kayser, 2009). Haplogroup frequencies by population were estimated by simple counting, and a correspondence analysis (CA) based on the frequency of sub-haplogroups (e.g., A2a) was performed and visualized with the R-packages FactoMineR (Le, Josse, & Husson, 2008) and factoextra (Kassambara and Mundt, 2016), respectively.

Population-based statistical analyses were performed with Arlequin v3.5 (Excoffier and Lischer, 2010). These include the analysis of molecular variance (AMOVA), estimation of molecular diversity indices, the estimation of pairwise genetic distances based on Φ_{ST}, and Tajima's D test of selective neutrality. A multidimensional scaling analysis (MDS) was performed on the matrix of pairwise Φ_{ST} values to visualize the distances between populations. Additionally, we performed a Mantel test to evaluate the correlations between genetic distances and geographic distances. The matrix of geographic distances was built using the geographic coordinates of the location where the majority of samples for each ethnolinguistic group were collected and then calculating the great circle distances between locations via the R packages ade4 and geosphere (Dray and Dufour, 2007; Hijmans, 2016). Furthermore, a multiple regression analysis on distance matrices (MRM) (Goslee and Urban, 2007) with the form: MRM(as.dist(gen.dist) ~ as.dist(geo.dist) + as.dist(rivers.dist)) was performed. This analysis takes into consideration a matrix of geographic distances and a matrix of proximity along rivers as predictor variables of the genetic distances (pairwise Φ_{ST} values) between populations (Pugach et al., 2016; Yunusbayev et al., 2012). For the matrix of river distances, a value of zero was given to populations living along the same river or on rivers that are closely connected, and a value of one was given to populations living on different rivers.

The sharing of haplotypes between populations was estimated with in-house R scripts as the proportion of pairs of identical sequences shared between populations. Additionally, networks of haplotypes were constructed with the software Network ver. 4.6.1.3 and visualized with Network Publisher ver. 2.0.0.1 (http://www.fluxus-engineering.com). Finally, BSPs were constructed by population and by haplogroup (i.e., A2, B2, C1, and D1) with BEAST ver. 1.8.2 (Drummond, Suchard, Xie, & Rambaut, 2012). For this analysis, the best substitution model was estimated with jModeltest 2.1.7 (Darriba, Taboada, Doallo, & Posada, 2012), and BEAST was used to estimate whether a strict or a relaxed clock model best fits the data. This analysis was performed on both the complete sequences and the sequences partitioned

TABLE 2 Frequency of haplogroups for the 24 NWA ethnolinguistic groups included in the population analyses

Population	*n*	A2	B2	C1	D1
Yucu-Matapi	39	0.28	0.10	0.56	0.05
Curripaco	17	0.18	0.53	0.24	0.06
Ach-Piapoco	24	0.54	0.04	0.42	0.00
Carijona	8	0.13	0.25	0.63	0.00
Desano	17	0.29	0.12	0.41	0.18
Other-ET	10	0.50	0.00	0.30	0.20
Pira-Wanano	13	0.31	0.08	0.38	0.23
Siriano	10	0.40	0.10	0.40	0.10
Tanimuka	10	0.50	0.10	0.40	0.00
Tuka-Tatuyo	10	0.30	0.10	0.30	0.30
Siona	17	0.59	0.35	0.00	0.06
Coreguaje	19	0.11	0.16	0.74	0.00
Sikuani	16	0.25	0.00	0.75	0.00
Guayabero	35	0.43	0.23	0.34	0.00
Saliba	16	0.19	0.06	0.56	0.19
Mur-Uitoto	26	0.23	0.12	0.31	0.35
Puinave	19	0.11	0.42	0.47	0.00
Nukak	16	0.00	0.31	0.69	0.00
Pasto	14	0.36	0.14	0.21	0.29
Kamentsa	11	0.18	0.09	0.64	0.09
Inga	17	0.59	0.06	0.29	0.06
Tikuna	18	0.44	0.00	0.44	0.11
Cocama	17	0.29	0.35	0.29	0.06
Yagua	13	0.62	0.15	0.23	0.00

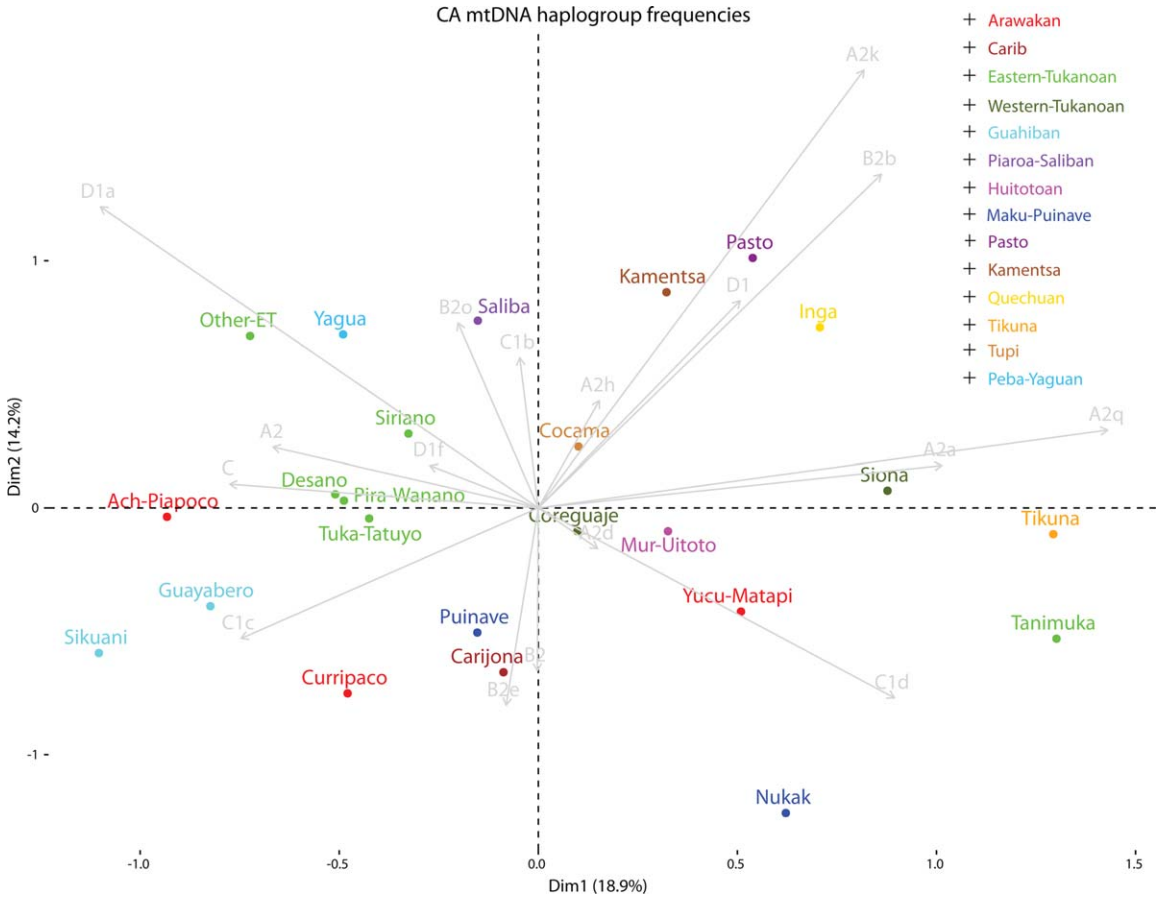

FIGURE 2 CA based on the sub-haplogroup frequencies by population. Populations are color-coded by linguistic affiliation

into coding (577–16023) and non-coding (16024-576) regions, applying the corresponding substitution rates reported previously in Soares et al. (2009).

3 | RESULTS

We generated 439 complete mitochondrial sequences to an average coverage per sample of 1,030×, which were deposited in GenBank with accession numbers: KY645515-KY645943 and MF152733-MF152739. All sequences belonged to one of the main Native American haplogroups, namely A2, B2, C1, and D1. Haplogroups A2 and C1 were the most frequent lineages in the NWA populations (excluding the so-called "Mestizos"), with more than half of all sequences belonging to A2 and C1 together. Table 2 provides a breakdown of the haplogroup frequencies for the ethnolinguistic groups included in the population analyses.

The CA (Figure 2) shows the clustering of populations based on the frequency of sub-haplogroups (Supporting Information Table S1). We observed differences among populations without a clear clustering by language family, with the exception of the Eastern Tukanoan groups. Most of these are clustered on the left side of the plot, although the Tanimuka do not cluster with the other Eastern Tukanoan groups. Additionally, Guayabero and Sikuani (who speak languages belonging to the Guahiban family) were located close to each other in the lower left pane of the plot. In addition to language affiliation, a few populations clustered due to geographic proximity, namely the Kamentsa, Pasto and Inga, who all live close to one another in the Andean foothills. In other cases, the relatively close proximity of populations in the plot could be attributed to their being settled along the same river or on rivers that are part of the same basin (Supporting Information Figure S1), such as the Curripaco and Puinave, who live on the Inírida and Atabapo Rivers.

3.1 | Molecular diversity indices

The genetic variation in these communities was assessed through different molecular diversity indices (Figure 3 and Supporting Information Table S2). On average, the gene diversity in these groups was high (0.9), but there were also differences amongst them. For example, Eastern Tukanoan groups showed consistently high values of gene diversity, with the exception of the Tanimuka, who had one of the lowest values (0.73). The Western Tukanoan groups Coreguaje (0.92) and Siona (0.82) showed lower values than Eastern Tukanoan groups. Among Arawakans, the Ach-Piapoco had the lowest value (0.77). The hunter-gatherer group Nukak showed the lowest gene diversity of all groups (0.64), with only four haplotypes observed among the 16 individuals analyzed. Additionally, agriculturalist groups tended to have higher gene diversities (average = 0.92) than hunter-gatherer groups (average = 0.80) (Mann-Whitney U test, p-value = .03).

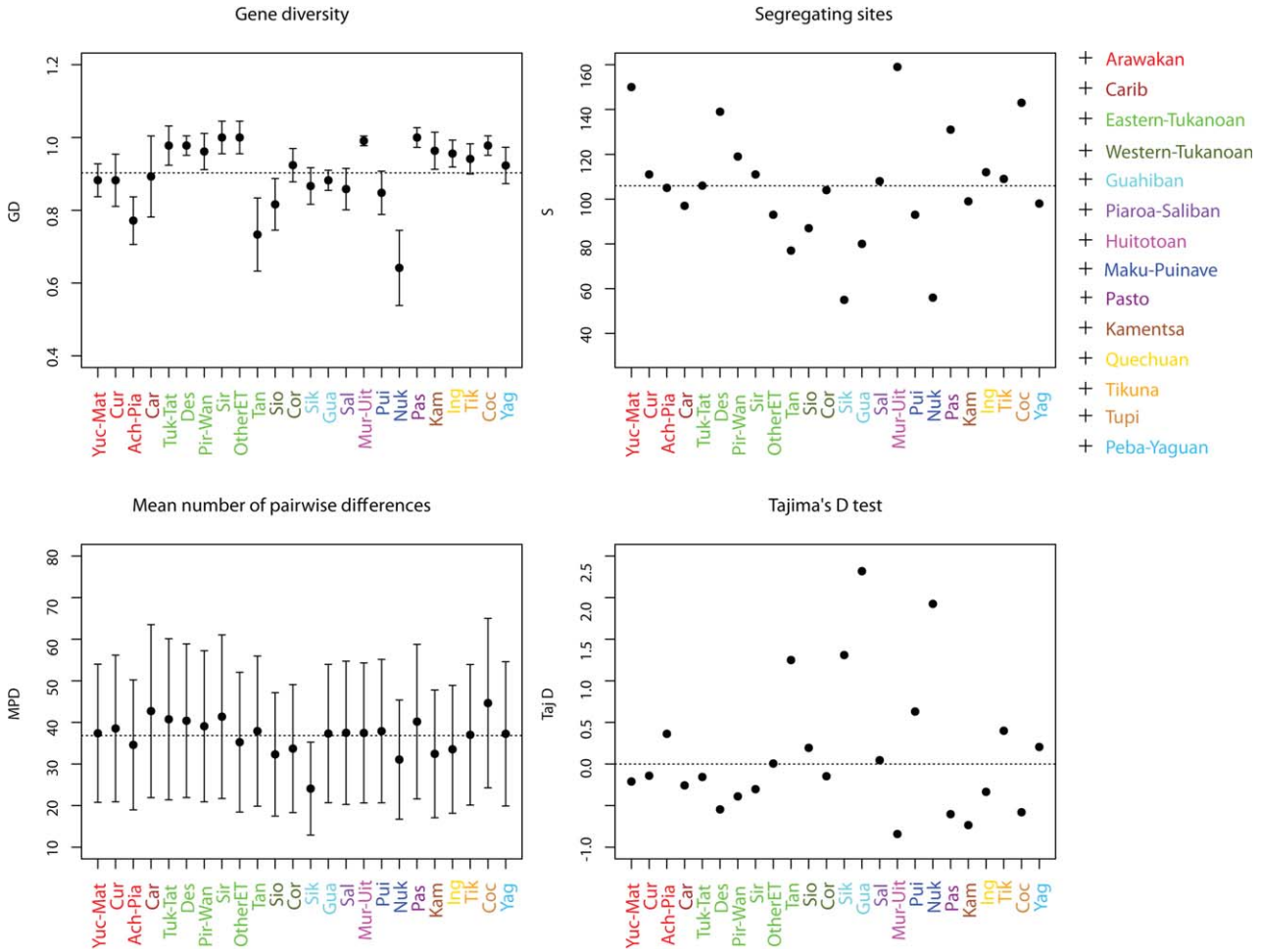

FIGURE 3 Molecular diversity indices by population. Dashed lines correspond to average values, except for Tajima's D test where it corresponds to zero. Populations are color-coded by linguistic affiliation as in Figure 2

The mean number of pairwise differences per population showed less variation, with an average of 41.07 ± 17.86 differences. The smallest values were found in Sikuani (24.08 ± 11.18) and Nukak (31.07 ± 14.33), and the largest values were observed in Cocama (44.64 ± 20.37), Carijona (42.71 ± 20.8), and Siriano (41.38 ± 19.66) The D values of Tajima's test of neutrality (Tajima, 1989) ranged from -0.735 to 2.318. Under neutrality, Tajima's D is expected to be equal to zero, and significant departures are interpreted as a result of selection or changes in population size. Although none of the D values were significant (all p-values > 0.2; Supporting Information Table S2), positive D values >1.2 were obtained for Guayabero, Nukak, Sikuani, and Tanimuka, which may reflect recent reductions in the size of these populations. This hypothesis was supported both by the distribution of pairwise differences by population (Supporting Information Figure S2), which showed increased frequencies for the category of small differences (0 and 1 differences) and for the category of large differences (50 or more), as well as by the Bayesian reconstruction of population size changes through time (BSP plots, Supporting Information Figure S3). Furthermore, the Tanimuka and Nukak had the lowest gene diversity values of any population analyzed.

3.2 | Shared haplotypes

A total of 216 different haplotypes were observed among the 412 sequences included in this analysis. Of these, 146 were unique haplotypes (i.e., found in single individuals) and 70 haplotypes were shared: 39 exclusively within populations, 18 exclusively between populations, and 13 both within and between populations. The shared haplotypes accounted for 64.6% of all the sequences analyzed. This amount of haplotype sharing between populations is quite high when compared with other population-based studies of complete mitochondrial genomes (Table 3). In other studies, the majority of shared haplotypes were generally observed within populations, with the exception of two African datasets from Burkina Faso and Zambia (Barbieri, Butthof, Bostoen, & Pakendorf, 2013; Barbieri et al. 2012), which showed low levels of sharing both within and between populations. The highest level of sharing between populations was observed for Siberian populations spread over a large geographic area (Duggan et al., 2013); the NWA populations analyzed in this study showed the second highest value of sharing between populations.

Figure 4 shows the proportion of pairs of sequences shared between and within NWA populations. Siriano, Other-ET, and Pasto

TABLE 3 Shared haplotypes in a worldwide sample of complete mitochondrial sequences sampled at the population level

Geographic region	No. sequences	No. haplotypes	%Unique haplotypes	Shared within population	Shared between populations	Sources
NW Amazonia	412	216	0.676	0.241	0.144	Present study
Burkina Faso	335	332	0.991	0.006	0.003	Barbieri et al. (2012)
SW Zambia	169	146	0.897	0.048	0.055	Barbieri et al. (2013)
Botswana/Namibia	218	128	0.75	0.188	0.133	Barbieri et al. (2014)
Philippines	365	233	0.734	0.227	0.077	Delfin et al. (2014)
Sumatra	72	48	0.771	0.229	0.021	Gunnarsdottir et al. (2011)
Taiwan	549	299	0.669	0.308	0.084	Ko et al. (2014)
Oceania	1,331	650	0.689	0.277	0.106	Duggan et al. (2014)
Siberia	525	244	0.574	0.336	0.217	Duggan et al. (2013)
Mexico[a]	113	90	0.867	0.133	0	Mizuno et al. (2014)

Note. The proportions do not sum up to 1 since some haplotypes are shared both within and between populations.
[a]The individuals from Mexico are all Native Americans from the Mazahua and Zapotec ethnic groups.

were the only groups without shared haplotypes within the populations, although Siriano and Other-ET did share with other populations. The majority of between-group haplotype sharing involved Arawakan and Eastern Tukanoan groups. The Arawakan groups shared mostly with groups living in close proximity (Supporting Information Figure S4), for example, Yucu-Matapi with Tanimuka; Curripaco with Puinave and Nukak; and Ach-Piapoco with Saliba and with the Guahiban groups Sikuani and Guayabero. Most Eastern Tukanoan groups, who practice linguistic exogamy, shared haplotypes among each other (except for Tanimuka, who shared only with Yucu-Matapi). In contrast, the

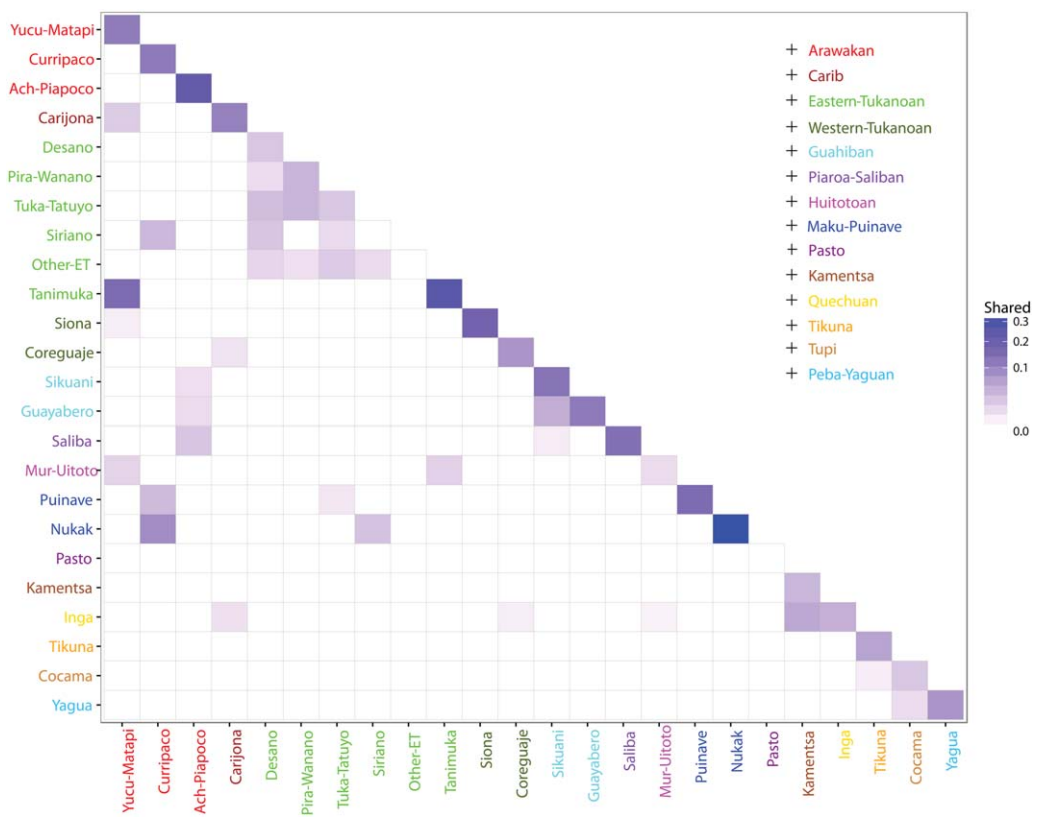

FIGURE 4 Matrix of shared haplotypes between populations. The color scale indicates the proportion of the total haplotypes that are shared within (on the diagonal) or between (below the diagonal) populations

TABLE 4 Analysis of molecular variance

	No. groups	Among groups	Within groups	Within populations	Global FST
One group	1		11.12**	88.88	0.1112
Language[a]	14	−1.33	12.37**	88.96**	0.1104
Geography[b]	6	1.04	10.24**	88.72**	0.1128
Rivers[c]	11	5.42**	5.99**	88.59**	0.1141

[a]1. Arawak: Yucu-Matapi, Curripaco, Ach-Piapoco; 2. Carib: Carijona; 3. Eastern-Tukanoan: Desano, Pira-Wanano, Siriano, Tuka-Tatuyo, Other-ET, Tanimuka; 4. Western-Tukanoan: Coreguaje, Siona; 5. Guahiban: Sikuani, Guayabero; 6.Huitoto: Mur-Uitoto; 7. Maku-Puinave: Puinave, Nukak; 8. Kamentsa; 9. Pasto; 10. Piaroa-Saliba: Saliba; 11. Peba-Yaguan: Yagua; 12. Quechua: Inga; 13.Tikuna; 14. Tupi: Cocama. [b]1. Saliba, Ach-Piapoco; 2. Sikuani, Guayabero, Nukak, Desano, Pira-Wanano, Siriano, Tuka-Tatuyo, Other-ET, Carijona; 3. Coreguaje, Siona, Mur-Uitoto, Inga, Kamentsa, Pasto; 4. Curripaco, Puinave; 5. Yucu-Matapi, Tanimuka; 6. Cocama, Tikuna, Yagua. [c]1. Meta: Saliba, Ach-Piapoco; 2. Vaupes: Desano, Pira-Wanano, Siriano, Tuka-Tatuyo, Other-ET, Carijona; 3. Guaviare: Guayabero, Sikuani; 4. Interfluve: Nukak; 5. Atabapo-Inirida: Curripaco, Puinave; 6. High-Putumayo: Inga, Kamentsa, Pasto; 7. Middle-Putumayo: Siona; 8. Lower-Putumayo: Mur-Uitoto; 9. Middle-Caqueta: Coreguaje; 10. Miritíi-Parana: Yucu-Matapi, Tanimuka; 11. Amazon: Cocama, Tikuna, Yagua.

Western Tukanoan groups Siona and Coreguaje shared primarily within their populations and did not share haplotypes with the Eastern Tukanoan groups.

The groups from the Andean foothills–Inga, Kamentsa, and Pasto–showed different patterns of shared haplotypes, despite the fact that they live in close geographic proximity. The Pasto, a group that has lost its native language and is largely incorporated into the admixed local population, shared no haplotypes with any population. The Kamentsa shared haplotypes only with the Inga, while the Inga also shared haplotypes with three other groups located further inside the Amazonian area—Carijona, Coreguaje, and Mur-Uitoto. Finally, of the three groups living on the banks of the Amazon River close to the town of Leticia, the Cocama shared with both the Yagua and Tikuna, whereas the latter two groups did not share with one another.

3.3 | Haplotype networks

The networks of haplotypes (Supporting Information Figure S5A–D) complement the patterns of sequence sharing, but in addition allow us to discern clusters of related (not just identical) haplotypes. We observed that some of these clusters included sequences from different language families while others were restricted to specific language families or to groups living in close geographic proximity (Supporting Information Table S3). For instance, Arawakan and Eastern Tukanoan groups exhibited several haplotypes within haplogroups A2, B2, and C1 (Supporting Information Figures S5A–C and Supporting Information Table S3) that were either shared or separated by only a few mutational steps. Notably, several of these clusters also included individuals speaking Maku-Puinave languages. Clusters of haplotypes restricted to specific groups are represented by clusters I and II of haplogroup D, which are exclusive to Eastern Tukanoan and Huitotoan populations, respectively. Furthermore, the haplotypes of the Quechuan populations and the Kamentsa, who live in close proximity in the Andean foothills, were either shared between them or closely related. Finally, the haplotypes of the Guayabero and Sikuani (Guahiban) were mostly differentiated from those of other populations and generally shared by several individuals within the family (Clusters II and III in Supporting Information Figure S5A; Cluster I in Supporting Information Figure S5C). The sequences belonging to cluster I in haplogroup C lack the diagnostic mutation A13263G for haplogroup C, but contain other diagnostic mutations that allow unambiguous assignment to haplogroup C. MtDNAs with this variant were previously identified in eastern Colombia by RFLP typing (Torres et al., 2006), where they occurred at high frequency in Guahibo, Piapoco, and Saliba groups. Given their high frequencies in the Guahiban groups, these haplotypes appear to belong to an autochthonous lineage that has diffused into other groups living in the Orinoco basin.

3.4 | Genetic structure and genetic distances

The AMOVA (Table 4) allows us to test different hypotheses about how genetic variation is structured in NWA. We defined groups *a priori* based on their language affiliation, geographic proximity, and distribution along major rivers or their tributaries to evaluate how much of the observed variation was explained by each grouping strategy. We observed that, of the three strategies, grouping populations by their distribution along rivers resulted in the largest among-group component of the genetic variance. In contrast, both language and geography were a poor predictor of the genetic structure, showing negative and non-significant values for the component of variance due to differences among groups. Although grouping by rivers performed better than grouping by geography or language, it still did not provide a very good description of the genetic structure, since the percentage of variance due to differences among populations within groups was still higher than the among groups component, suggesting the existence of other influences on substructure within populations.

The matrix of genetic distances between populations based on pairwise Φ_{ST} values (Supporting Information Figure S6) was used to construct an MDS plot (Figure 5). The populations did not form any clear clustering: the majority of populations were grouped together in the center of the plot (indicated by the inner circle in Figure 5) with an average pairwise $\Phi_{ST} = 0.03$, while around the main cluster a second group of populations showed higher differentiation (external circle, average $\Phi_{ST} = 0.07$). Sikuani, Siona, and the hunter-gatherer Nukak

appeared as outliers with high genetic differentiation (average $\Phi_{ST} = 0.22$). This picture did not change after adding an additional dimension to the MDS plot (Supporting Information Figure S7). Particularly striking were the small genetic distances between the Eastern Tukanoan groups, who clustered together in the center of the MDS plot. Although the Tanimuka appeared more distant from the main cluster of Eastern Tukanoan groups, their pairwise Φ_{ST} values were not significantly different (Supporting Information Figure S6) and the average Φ_{ST} (0.02) indicated low genetic differentiation among all Eastern Tukanoan groups. In contrast, the Coreguaje and the Siona, who speak languages of the Western Tukanoan branch, showed larger genetic distances, both with the Eastern Tukanoan groups and with each other. Populations from each of the other language families did not form clusters with their linguistic relatives. For example, Arawakan groups occupied different positions in the plot and their Φ_{ST} values were significantly different.

The results of the Mantel test showed a lack of significant correlation between geographic distances, estimated as great-circle distances, and the matrix of pairwise Φ_{ST} values ($r = 0.07$, p-value $= 0.28$). However, since rivers emerged as an important factor explaining the structure of genetic variation in the AMOVA results (Table 4), we also performed a MRM, where we added rivers as an additional predictor variable. Adding rivers to the regression model resulted in an increase in the amount of variation explained by the model (Table 5), with rivers being a significant predictor (p-value $= 0.01$). We then jack-knifed over populations (Pugach et al. 2016; Ramachandran et al., 2005) and identified three populations as outliers—Sikuani, Siona, and Nukak—groups that appeared as outliers in the MDS plot as well (Figure 5). We performed the multiple regression analysis excluding the outliers, and this resulted in an increase of 3.4% in the R square value, a better correlation between genetic and geographic distances, and geography becoming a significant predictor factor (p-value < 0.05) (Table 5 and Supporting Information Figure S8), although rivers were no longer a significant predictor of genetic subdivision.

3.5 | Bayesian demographic reconstruction

BSPs were generated for each haplogroup (A2, B2, C1, and D1) and population. All four haplogroups showed an increase in effective population size between 17,500 and 25,000 years before present. This signal was more evident for haplogroups A2 and C1, which had the highest number of sequences (Supporting Information Figure S9). In contrast, the BSP plots by population showed different outcomes. We observed four main trajectories (Supporting Information Figure S3): (a) a signal of population size increase shown by Yucu-Matapi, Curripaco, Desano, Siriano, Inga, Pasto, Mur-Uitoto, Tikuna, and Cocama (exemplified by Yucu-Matapi in Supporting Information Figure S3A); (b) population stability through time shown by Ach-Piapoco, Tanimuka, Coreguaje, Siona, Kamentsa, Puinave, and Yagua (exemplified by Coreguaje in Supporting Information Figure S3B); and (c) population contraction shown by Sikuani, Guayabero, and Nukak (exemplified by Nukak in Supporting Information Figure S3C), which is particularly

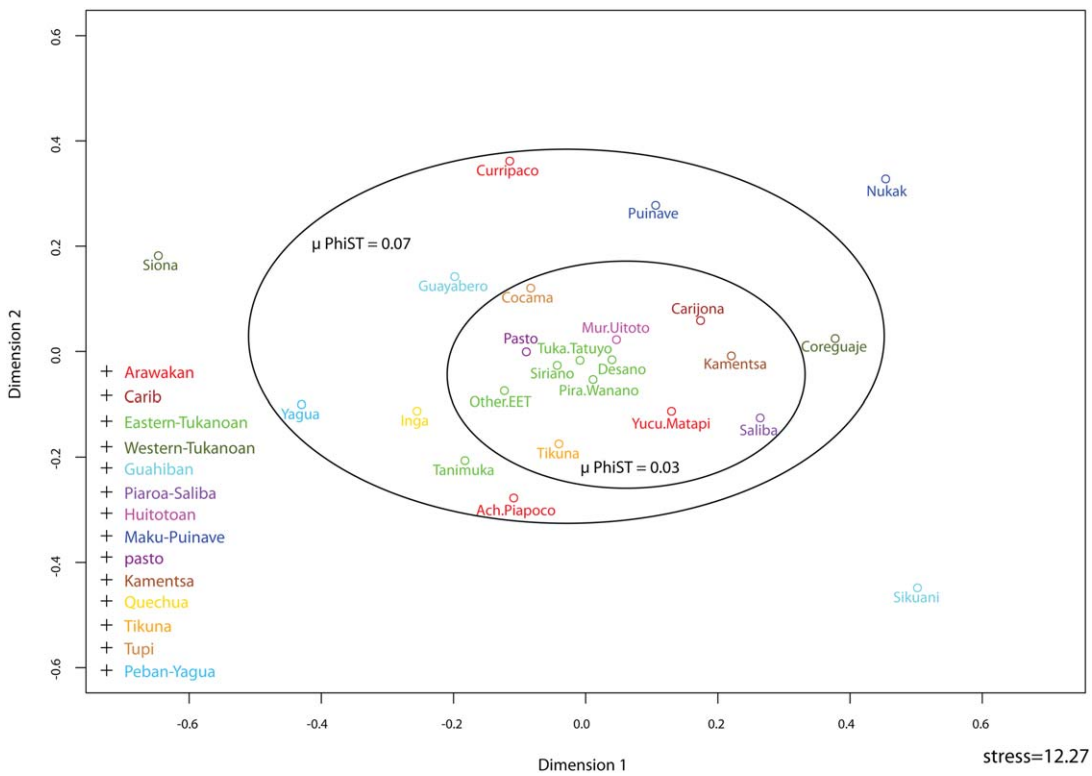

FIGURE 5 MDS plot based on Φ_{ST} genetic distances. Stress value is given in percentage. The inner circle indicates populations with low genetic differentiation and the outer circle indicates populations with moderate differentiation. μ PhiST is the average pairwise Φ_{ST} value within each circle

TABLE 5 Multiple regression analysis on distance matrices

		gen.dist all populations			
		Reg.coefficient	*p*-value	R^2	*p*-value
Simple regression	geo.dist	3.28×10^{-8}	0.444	0.009	0.444
Multiple regression	geo.dist	9.64×10^{-10}	0.983	0.036	0.126
	rivers.dist	5.71×10^{-2}	0.011		
		gen.dist without outliers			
Simple regression	geo.dist.no.outliers	6.70×10^{-8}	0.026	0.067	0.026
Multiple regression	geo.dist.no.outliers	5.76×10^{-8}	0.112	0.070	0.040
	rivers.dist.no.outliers	1.38×10^{-2}	0.458		

striking for Sikuani (Supporting Information Figure S3D). These differences in the effective population size through time suggest that these populations have independent demographic histories.

4 | DISCUSSION

We have investigated the genetic diversity of ethnolinguistic groups from NWA at the level of complete mitochondrial genomes. This area is underrepresented in previous studies, and our data help to fill a gap in our knowledge about the genetic diversity of modern human populations. We have found that NWA harbors a considerable amount of genetic diversity, with evidence for contact among different ethnolinguistic groups, contrary to the common picture of Amazonian populations as being small and isolated with low genetic diversity (Fuselli et al., 2003; Wang et al., 2007). NWA populations show values of nucleotide diversity as high as or higher than those observed in most other non-African populations (Supporting Information Figure S10), and they display the second-highest amount of sequence sharing in a world-wide comparison (Table 3). The complete mitochondrial genome is the maximum level of resolution one can achieve to differentiate individuals and populations at the maternal level, so the presence of identical sequences among populations living in distant geographic areas likely indicates their common ancestry and/or recent contact.

4.1 | Lack of genetic structure along linguistic lines

Although our dataset includes populations speaking languages belonging to different language families, we found that linguistic affiliation is a poor predictor of genetic structure, as shown by the AMOVA (Table 4) and the CA plot based on sub-haplogroup frequencies (Figure 2 and Supporting Information Table S1). These results indicate that language does not constitute a barrier to gene flow, and that groups have been interacting with neighboring groups for some time, especially along rivers, which in our analyses performed better in explaining the patterns of genetic diversity. Archeological and linguistic evidence demonstrates that NWA has been an area of intense contact and movement of peoples of different cultural traditions, as evidenced by the diffusion of ceramic styles (Heckenberger, 2002; Lathrap, 1970; Zucchi, 2002) and shared subsistence strategies, by the existence of language areas and contact-induced linguistic change (Aikhenvald, 1999), and the generalized multilingualism among groups (Sorensen, 1967; Stenzel, 2005). Likewise, ethnographic studies provide additional evidence of contact among groups. For example, both Arawakans and Eastern Tukanoans share a ceremonial complex for male initiation known as Yurupari, in which sacred flutes and trumpets are only played by males, as well as sharing myths concerning the hero Kúwai (Hugh-Jones, 1979; Jackson, 1983; Vidal, 2002). In addition, the Eastern Tukanoan groups from the Pira-Parana and Apaporis Rivers (Barasano, Makuna, and Tanimuka) reveal Arawakan influence, since they also practice dances with masks during the season of high abundance of the palm tree fruit pupunha (*Bactris gasipaes*) (Hugh-Jones, 1979).

The genetic distances among populations provide additional evidence in this regard. Although the global Φ_{ST} value of 0.11 indicates moderate differentiation (Hartl and Clark, 2007), this value is driven by three populations, namely the Siona, Sikuani, and Nukak. These are highly differentiated from the other populations, most likely reflecting the effects of genetic drift due to bottlenecks, as indicated by the positive Tajima's D values (Figure 3) and the distribution of pairwise differences (Supporting Information Figure S2). When we exclude these populations, we observe an average pairwise Φ_{ST} of 0.07, and populations appear close together in the MDS plot (Figure 5), indicating low genetic differentiation among NWA populations.

In this general picture, the Eastern Tukanoan groups stand apart, since they cluster together in the CA and MDS plots (Figures 2 and 5), and their pairwise genetic distances are small and non-significant (Supporting Information Figure S6). Linguists have proposed a time depth for the Tukanoan family of 2000–2500 years, based on a comparison of the diversity in Tukanoan languages with the diversity in Romance and Germanic languages (Chacon, 2014). The time depth of the Eastern Tukanoan branch (and thus the time to the most recent common ancestor of the Eastern Tukanoan languages) would be even more recent, and as such might indicate that the peoples speaking these languages share recent common genetic ancestry as well (at least on the maternal side). However, the Eastern Tukanoan groups practice linguistic exogamy, and the close genetic relationships among these populations might be the result of this marital system, in which women move among different ethnolinguistic groups. The consequences of the linguistic exogamy are also evident in the gene diversity values and the

patterns of shared haplotypes. Eastern Tukanoans show the highest gene diversity values and share more haplotypes among themselves than with other non-Eastern Tukanoan groups. In addition, their haplotypes tend to be closely related, as seen in the phylogenetic networks (Supporting Information Figure S5). Analyses of the Y-chromosome as well as nuclear markers will help to disentangle the effects of linguistic exogamy versus recent common ancestry on the patterns of genetic variation among Eastern Tukanoan groups.

The Tanimuka stand apart from the other Eastern Tukanoan groups in the analyses, and this may reflect their settlement further south, along the Apaporis and Mirití-Paraná Rivers. Moreover, they do not participate in the linguistic exogamic system with other Eastern Tukanoan groups, but interact mainly with the Arawakan groups Yucuna and Matapi. These interactions are reflected in the patterns of haplotype sharing (Figure 4) as well as in their language, which shows evidence of Arawakan influence (Barnes, 1999; Chacon, 2014).

4.2 | The role of rivers in structuring genetic variation

Besides language, geography is another important factor in structuring the patterns of genetic variation in human populations (Ramachandran et al., 2005; Schönberg, Theunert, Li, Stoneking, & Nasidze, 2011; Wang et al., 2007). One of the most salient characteristics of the physical landscape of NWA is the high density of rivers that drain the area, and their importance for human populations was earlier recognized by explorers and ethnographers that traveled through the region (Koch-Grünberg, 1995; Wallace, 1853). We found that the distribution of populations along rivers is an additional important factor influencing their genetic structure. AMOVA (Table 4) shows that clustering populations according to the rivers where they are distributed explains more of the genetic variation that is due to differences among groups than does grouping them by linguistic affiliation, that is, populations living in the same river basin or along closely connected rivers are genetically more similar than those living on different rivers. This pattern is also observed in the distribution of sub-haplogroups among populations, which drive their location in the CA plot (Supporting Information Figure S1). For example, the Curripaco and Puinave, who live on the Inírida and Atabapo Rivers, are located close together in the plot. The presence of Coreguaje, Yucu-Matapi, and Mur-Uitoto in the center of the plot could reflect their presence in a region where the Putumayo and Caquetá Rivers are separated by their shortest distance, therefore facilitating contact among people inhabiting the basins and tributaries of these two rivers. Indeed, one Murui individual was sampled in a Coreguaje community, and two Uitoto individuals were sampled in the Mirití-Parana region, thus providing evidence for the movement of people among these groups. The results of the MRM analysis provide additional evidence in this regard: even though no correlation between genetic distances and geographic distances was observed via the Mantel test, adding river distances as an additional predictor variable resulted in an increase of around 3% of the R-square value (Table 5), indicating that rivers contribute to explaining a slightly higher percentage of the variation observed in the genetic distances.

Rivers in Amazonia serve a double function in providing a means of communication as well as subsistence, and the wide distribution of certain cultural traits (e.g., the production of Saladoid-Barrancoid ceramics and circular plaza village settlement patterns) has been associated with the expansion of Arawakan-speaking populations along the extensive system of NWA waterways (Heckenberger, 2002; Hornborg, 2005; Lathrap, 1970; Lowie, 1948). They also mark a distinction in subsistence strategies between the more numerous "river people" who build canoes, settle along rivers, and rely on horticulture and fishing, and the "forest people" who inhabit the interfluvial areas, settle away from the major rivers, and base their subsistence on foraging (Epps and Stenzel, 2013). Additionally, the rivers have profound meanings and are embedded in the cosmogonies of several NWA indigenous groups. The Eastern Tukanoan creation myths describe the journeys that the ancestors of the people made to settle this world on board an anaconda canoe that traveled up the Vaupes River; from the anaconda's body all the Tukanoan siblings emerged (Chernela, 2010; Jackson, 1983). Arawakan groups also describe a series of ever returning voyages from the sacred center of the world and the place of emergence of the first ancestors at the rapids of Hípana on the Aiary River, covering the major arteries of the Rio Negro, Orinoco, and Amazon Rivers (Wright, 2002; Zucchi, 2002). Therefore, our findings about the role of rivers in structuring the genetic variation are in keeping with the ethnographically demonstrated role that rivers play for NWA populations.

The lack of fit between genetic and simple geographic distances may be the result of relatively recent movements and the displacement of ethnolinguistic groups from their traditional territories. Population dynamics and population sizes were drastically altered during the last five centuries, starting with early colonial times (16th and 17th centuries), when many groups were decimated by newly introduced epidemics and moved away from the accessible margins of the major rivers to avoid the slave raids of the Spanish, Portuguese, and Dutch colonizers (Santos-Granero 2002). Similar perturbations happened during the time of the Christian missions in the 18th century, when many groups were forced to relocate to multiethnic mission settlements, and finally during the rubber boom between the 19th and beginning of the 20th centuries, when the groups who managed to escape the mercenaries exploiting the rubber fields resettled in remote areas in the headwaters of small rivers (Dixon and Aikhenvald, 1999; Hill and Santos-Granero, 2002; Stenzel, 2005). The inferred reduction in population size of the Tanimuka, Sikuani, Guayabero, and Nukak, as indicated by their low diversity values, the positive Tajima's D values (Figure 3), the distribution of pairwise differences (Supporting Information Figure S2), and the reconstruction of effective population sizes (Supporting Information Figure S3C,D), might be a result of these social upheavals.

4.3 | The impact of subsistence strategies on the genetic diversity

NWA contains groups with different subsistence strategies, with manioc (*Manihot esculenta*) the main staple among horticulturalist groups, who are best described as riverine horticultural societies, given their close association with rivers. The Nukak, in contrast, are traditionally foragers,

who still rely on hunting and gathering, and move throughout the extensive area between the Guaviare and Inírida Rivers. Furthermore, the Guayabero, Sikuani, and Puinave are traditional foragers who have only recently undergone the transition to agriculture, and are therefore considered as HGP together with Nukak in our analyses (Table 1). Our data show that agricultural societies (AG) have higher levels of diversity on average than forager groups (HGP) as indicated by the Mann-Whitney U test (p-value = .03), while the HGP groups have larger values of Tajima's D statistic (Figure 3) and do not show signals of population expansion (Supporting Information Figure S3). These findings agree with patterns reported for other HGPs around the world (Aime et al., 2013; Excoffier and Schneider, 1999; Oota et al., 2005) and contrast with the genetic signature of an agricultural way of life, namely higher effective population size (Patin et al., 2014), higher levels of diversity, and significantly negative values of Tajima's D test (Aime et al., 2013).

However, subsistence strategies are flexible and diverse among NWA populations. Horticulturalists complement their diet with occasional hunting and/or gathering of several kinds of palm fruit, and extensive exchanges between AG and HGP groups have been reported (Jackson, 1983; Milton, 1984). In this system, HGP populations usually provide meat and several products from the forest, such as the poison curare for the tips of darts and arrows, in exchange for different cultivated products, such as manioc and other trade goods (Epps and Stenzel, 2013; Jackson, 1983; Milton, 1984). Nonetheless, this exchange seems to be exclusively restricted to goods and labor, with little or no intermarriage documented between AG and HGP groups (Aikhenvald, 1996). In contrast, we observed shared haplotypes between AG and HGP groups, which likely reflects recent intermarriage or recent common ancestry. For example, the most frequent haplotype in the Arawakan AG group Curripaco (Haplotype H_84 in Supporting Information Figure S4) is observed at high frequency in the HGP Nukak (and in the Eastern Tukanoan AG group Siriano). Moreover, the HGP Puinave share several haplotypes with the AG group Curripaco (H_219, H_161, H_117 in Supporting Information Figure S4), a likely result of intermarriage between these groups, since there are communities on the Inírida River where one finds individuals from both groups. Similarly, the Guahiban HGP groups Sikuani and Guayabero exhibit a haplotype at high frequency (H_43 in Supporting Information Figure S4) that is shared with the AG Ach-Piapoco as well as further haplotypes related to haplotypes found in AG Arawakan groups (Clusters I and II in Supporting Information Figure S5B,C). This may reflect contact among them, since there are Piapoco communities on the lower Guaviare River as well as Sikuani communities on the Meta River, places where these groups overlap.

However, it is difficult to determine the direction of the gene flow or to distinguish between contact and common ancestry as explanations for shared mtDNA haplotypes. Nevertheless, it is plausible that, where haplotypes are shared, the source population is the one in which the haplotype is present at higher frequency. For instance, the shared haplotype between the HGP Puinave and the AG Curripaco (H_219 in Supporting Information Figure S4) has a likely origin in Puinave, because of its higher frequency and the presence of related haplotypes in Puinave (Cluster I Supporting Information Figure S5B). The source of the shared haplotype among the HGP Nukak and the AG Curripaco and Siriano (H_84 in Supporting Information Figure S4) is more difficult to infer, since its frequency is similar in the Nukak and in the Curripaco; furthermore, three other haplotypes present in the HGP Nukak and Guayabero are only one mutation apart from it (Cluster II in Supporting Information Figure S5B). Therefore, it is likely that this haplotype, too, moved from the HGP populations into the AG Curripaco. A similar explanation could be given for H_43 in Figure S4, which is part of the Cluster I in Supporting Information Figure S5C, moving from the HGP Guayabero and Sikuani into the AG Ach-Piapoco.

Thus, these observations seem to fit a scenario of asymmetric gene flow in which women move from HGP to AG, a pattern that has been reported for populations in Central and Southern Africa (Barbieri et al., 2014; Destro-Bisol et al., 2004; Verdu et al., 2013). However, this scenario will need to be further refined by analyses of Y-chromosome and genome-wide data, which will allow us to determine whether the gene flow among groups was sex-biased (i.e., involving the movement of only females or only males among groups) and to make inferences about the time and magnitude of these events.

In conclusion, this study provides new data from this remote and little-studied part of the world, which allow insights into the impact of cultural practices on the patterns of genetic variation and on the population dynamics of NWA groups. Although our current data do not allow us to distinguish whether the population movements took place prior to or as a consequence of European contact, analyses of Y-chromosome variation and genome-wide data will shed further light on the genetic history of NWA. Furthermore, historical genetic studies will benefit from more archeological work in NWA, since huge areas remain completely unexplored.

ACKNOWLEDGMENTS

We are grateful to all sample donors, communities, community leaders, and regional indigenous organizations. L.A. especially gives thanks to Consejo Regional Indígena del Guaviare, Asociación de Autoridades Tradicionales y Cabildos de los Pueblos Indígenas del Municipio de Leguízamo y Alto Resguardo Predio Putumayo, Asociación de Cabildos Indígenas del Trapecio Amazónico, Organización Zonal Indígena Del Putumayo, the staff of Parques Nacionales Naturales in Puerto Leguízamo, the office of Indigenous Affairs in Puerto Inírida, Rafael Rodriguez, William Yucuna, the late Gustavo Arias, and all people who helped during the fieldtrips for their valuable collaboration and warm welcome during our stay in their communities. We also acknowledge Roland Schroeder for laboratory technical assistance, and Enrico Macholdt, Alexander Huebner, Irina Pugach, and Michael Dannemann for advice with the data analyses. B.P. acknowledges the LABEX ASLAN (ANR-10-LABX-0081) of Université de Lyon for its financial support within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) of the French government operated by the National Research Agency (ANR). L.A. was supported by a graduate grant from COLCIENCIAS; research was supported by funds from the Max Planck Society.

ORCID

Leonardo Arias http://orcid.org/0000-0001-6460-7934

REFERENCES

Aceituno, F. J., Loaiza, N., Delgado-Burbano, M. E., & Barrientos, G. (2013). The initial human settlement of Northwest South America during the Pleistocene/Holocene transition: Synthesis and perspectives. *Quaternary International*, *301*, 23–33.

Achilli, A., Perego, U. A., Lancioni, H., Olivieri, A., Gandini, F., Hooshiar Kashani, B., … Torroni, A. (2013). Reconciling migration models to the Americas with the variation of North American native mitogenomes. *Proceedings of the National Academy of Science of the United States of America*, *110*, 14308–14313.

Aikhenvald, A. Y. (1996). Areal diffusion in Northwest Amazonia: The case of Tariana. *Anthropological Linguistics*, *38*, 73–116.

Aikhenvald, A. Y. (1999). The Arawak language family. In: Dixon RMW, and Aikhenvald AY, editors. *The amazonian languages* (pp. 65–102). Cambridge; New York: Cambridge University Press.

Aime, C., Laval, G., Patin, E., Verdu, P., Segurel, L., Chaix, R., … Austerlitz, F. (2013). Human genetic data reveal contrasting demographic patterns between sedentary and nomadic populations that predate the emergence of farming. *Molecular Biology & Evolution*, *30*, 2629–2644.

Andrews, R. M., Kubacka, I., Chinnery, P. F., Lightowlers, R. N., Turnbull, D. M., & Howell, N. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nature Genetics*, *23*, 147.

Arango, R., & Sánchez, E. (2004). *Los pueblos indígenas de Colombia en el umbral del nuevo milenio : población, cultura y territorio : bases para el fortalecimiento social y económico de los pueblos indígenas*. Bogotá: Departamento Nacional de Planeación. 526 p.

Bailliet, G., Rothhammer, F., Carnese, F. R., Bravi, C. M., & Bianchi, N. O. (1994). Founder mitochondrial haplotypes in Amerindian populations. *American Journal of Human Genetics*, *55*, 27–33.

Barbieri, C., Butthof, A., Bostoen, K., & Pakendorf, B. (2013). Genetic perspectives on the origin of clicks in Bantu languages from southwestern Zambia. *European Journal of Human Genetics*, *21*, 430–436.

Barbieri, C., Guldemann, T., Naumann, C., Gerlach, L., Berthold, F., Nakagawa, H., … Pakendorf, B. (2014). Unraveling the complex maternal history of Southern African Khoisan populations. *American Journal of Physical Anthropology*, *153*, 435–448.

Barbieri, C., Heggarty, P., Castri, L., Luiselli, D., & Pettener, D. (2011). Mitochondrial DNA variability in the Titicaca basin: Matches and mismatches with linguistics and ethnohistory. *American Journal of Human Biology*, *23*, 89–99.

Barbieri, C., Whitten, M., Beyer, K., Schreiber, H., Li, M., & Pakendorf, B. (2012). Contrasting maternal and paternal histories in the linguistic context of Burkina Faso. *Molecular Biology & Evolution*, *29*, 1213–1223.

Barnes, J. (1999). Tucano. In: Dixon RMW, and Aikhenvald AY, editors. *The amazonian languages* (pp. 207–225). New York: Cambridge University Press.

Bodner, M., Perego, U. A., Huber, G., Fendt, L., Rock, A. W., Zimmermann, B., … Parson, W. (2012). Rapid coastal spread of First Americans: Novel insights from South America's Southern Cone mitochondrial genomes. *Genome Research*, *22*, 811–820.

Bolnick, D. A., & Smith, D. G. (2003). Unexpected patterns of mitochondrial DNA variation among Native Americans from the southeastern United States. *American Journal of Physical Anthropology*, *122*, 336–354.

Bonatto, S. L., & Salzano, F. M. (1997). Diversity and age of the four major mtDNA haplogroups, and their implications for the peopling of the New World. *American Journal of Human Genetics*, *61*, 1413–1423.

Campbell, L. (1997). *American Indian languages: The historical linguistics of native America* (pp. 528). New York: Oxford University Press.

Chacon, T. (2014). A revised proposal of proto-tukanoan consonants and Tukanoan family classification. *International Journal of American Linguistics*, *80*, 275–322.

Chacon, T., & Cayón, L. (2013). Considerações sobre a exogamia linguística no Noroeste Amazônico. *Revista de Letras da Universidade Católica de Brasília*, *6*, 6–20.

Chernela, J. M. (2010). *The Wanano Indians of the Brazilian Amazon: A Sense of Space*. Austin: University of Texas Press. 207 p.

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, *9*, 772.

de Saint Pierre, M., Gandini, F., Perego, U. A., Bodner, M., Gomez-Carballa, A., Corach, D., … Olivieri, A. (2012). Arrival of Paleo-Indians to the southern cone of South America: New clues from mitogenomes. *PLoS One*, *7*, e51311.

Dediu, D., & Levinson, S. C. (2012). Abstract Profiles of Structural Stability Point to Universal Tendencies, Family-Specific Factors, and Ancient Connections between Languages. *Plos One*, *7*(9):e45198.

Delfin, F., Min-Shan Ko, A., Li, M., Gunnarsdottir, E. D., Tabbada, K. A., Salvador, J. M., … Stoneking, M. (2014). Complete mtDNA genomes of Filipino ethnolinguistic groups: a melting pot of recent and ancient lineages in the Asia-Pacific region. *European Journal of Human Genetics*, *22*, 228–237.

Denevan, W. M. (1992). The Pristine Myth: The landscape of the Americas in 1492. *Annals of the Association of American Geographers*, *82*, 369–385.

Destro-Bisol, G., Donati, F., Coia, V., Boschi, I., Verginelli, F., Caglia, A., … Capelli, C. (2004). Variation of female and male lineages in sub-Saharan populations: the importance of sociocultural factors. *Molecular Biology & Evolution*, *21*, 1673–1682.

Dixon, R. M. W., & Aikhenvald, A. Y. (1999). *The Amazonian languages* (Vol. *xxviii*, pp. 446). Cambridge University Press.

Dray, S., & Dufour, A. B. (2007). The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software*, *22*, 1–20.

Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology & Evolution*, *29*, 1969–1973.

Duggan, A. T., Whitten, M., Wiebe, V., Crawford, M., Butthof, A., Spitsyn, V., … Pakendorf, B. (2013). Investigating the prehistory of Tungusic peoples of Siberia and the Amur-Ussuri region with complete mtDNA genome sequences and Y-chromosomal markers. *PLoS One*, *8*, e83570.

Epps, P., & Stenzel, K. (2013). *Upper Rio Negro: Cultural and linguistic interaction in northwestern Amazonia* (pp. 13–50). Rio de Janeiro: Museu do Índio – FUNAI, Museu Nacional.

Eriksen, L. (2011). Nature and culture in prehistoric Amazonia [PhD]. Lund: Lund University.

Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*, 564–567.

Excoffier, L., & Schneider, S. (1999). Why hunter-gatherer populations do not show signs of pleistocene demographic expansions. *Proceedings of the National Academy Sciences of the United States of America*, *96*, 10597–10602.

Fagundes, N. J., Kanitz, R., Eckert, R., Valls, A. C., Bogo, M. R., Salzano, F. M., … Bonatto, S. L. (2008). Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. *American Journal of Human Genetics*, *82*, 583–592.

Fehren-Schmitz, L., Llamas, B., Lindauer, S., Tomasto-Cagigao, E., Kuzminsky, S., Rohland, N., … Haak, W. (2015). A re-appraisal of the early Andean human remains from Lauricocha in Peru. *PLoS One*, *10*, e0127141.

Franco, R. (2002). *Los carijonas de Chiribiquete* (pp. 318). Bogotá: Fundación Puerto Rastrojo.

Fuselli, S., Tarazona-Santos, E., Dupanloup, I., Soto, A., Luiselli, D., & Pettener, D. (2003). Mitochondrial DNA diversity in South America and the genetic history of Andean highlanders. *Molecular Biology & Evolution*, *20*, 1682–1691.

Gaya-Vidal, M., Moral, P., Saenz-Ruales, N., Gerbault, P., Tonasso, L., Villena, M., … Dugoujon, J. M. (2011). mtDNA and Y-chromosome diversity in Aymaras and Quechuas from Bolivia: different stories and special genetic traits of the Andean Altiplano populations. *American Journal of Physical Anthropology*, *145*, 215–230.

Gnecco, C., & Mora, S. (1997). Late Pleistocene early Holocene tropical forest occupations at San Isidro and Pena Roja, Colombia. *Antiquity*, *71*, 683–690.

Goslee, S. C., & Urban, D. L. (2007). The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, *22*, 1–19.

Gunnarsdottir, E. D., Li, M., Bauchet, M., Finstermeier, K., & Stoneking, M. (2011). High-throughput sequencing of complete human mtDNA genomes from the Philippines. *Genome Research*, *21*, 1–11.

Hall, T. A. (1999). BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, *41*, 4.

Hartl, D. L., & Clark, A. G. (2007). *Principles of population genetics* (Vol. *xv*, pp. 652). Sunderland, MA: Sinauer Associates, Inc.

Heckenberger, M. J. (2002). Rethinking the Arawakan Diaspora: Hierarchy, Regionality, and the Amazonian Formative. In: Hill JD, and Santos-Granero F, editors. *Comparative Arawakan Histories: Rethinking Language Family and Culture Area in Amazonia*. Urbana: University of Illinois Press. p 99–122.

Heckenberger, M. J. (2008). Amazonian Mosaics: Identity, Interaction, and Integration in the Tropical Forest. In H. Silverman & W. H. Isbell (Eds.), *The handbook of South American archaeology* (pp. 941–961). New York, NY: Springer New York.

Hijmans, R. J. (2016). geosphere: Spherical Trigonometry. R package version 1.5-5. https://CRAN.R-project.org/package=geosphere.

Hill, J. D., & Santos-Granero, F. (2002). *Comparative Arawakan histories: Rethinking language family and culture area in Amazonia* (pp. 340). Urbana: University of Illinois Press.

Hock, H. H., & Joseph, B. D. (2009). *Language History, Language Change, and Language Relationship: An Introduction to Historical and Comparative Linguistics*. Berlin: Mouton De Gruyter. 588 p.

Hornborg, A. (2005). Ethnogenesis, regional integration, and ecology in prehistoric Amazonia - Toward a system perspective. *Current Anthropology*, *46*, 589–620.

Hugh-Jones, S. (1979). *The palm and the Pleiades: initiation and cosmology in northwest Amazonia* (Vol. *xvi*, pp. 332). Cambridge; New York: Cambridge University Press.

Jackson, J. E. (1983). *The fish people: Linguistic exogamy and Tukanoan identity in northwest Amazonia* (Vol. *xix*, pp. 287). Cambridge: Cambridge University Press.

Kassambara, A., & Mundt, F. (2016). factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package version 1.0.3 ed: https://CRAN.R-project.org/package=factoextra.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology & Evolution*, *30*, 772–780.

Keyeux, G., Rodas, C., Gelvez, N., & Carter, D. (2002). Possible migration routes into South America deduced from mitochondrial DNA studies in Colombian Amerindian populations. *Human Biology*, *74*, 211–233.

Kirby, K. R., Gray, R. D., Greenhill, S. J., Jordan, F. M., Gomes-Ng, S., Bibiko, H. J., Blasi, D. E., Botero, C. A., Bowern, C., Ember, C. R. et al. (2016). D-PLACE: A Global Database of Cultural, Linguistic and Environmental Diversity. *Plos One*, *11*(7):e0158391.

Kircher, M., Sawyer, S., & Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research*, *40*, e3.

Kivisild, T. (2015). Maternal ancestry and population history from whole mitochondrial genomes. *Investigating Genetics*, *6*, 3.

Kloss-Brandstatter, A., Pacher, D., Schonherr, S., Weissensteiner, H., Binna, R., Specht, G., & Kronenberg, F. (2011). HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Human Mutation*, *32*, 25–32.

Ko, A. M., Chen, C. Y., Fu, Q., Delfin, F., Li, M., Chiu, H. L., Stoneking, M., & Ko, Y. C. (2014). Early Austronesians: into and out of Taiwan. *Am J Hum Genet*, *94*(3):426–436.

Koch-Grünberg, T. (1995). *Dos años entre los indios: viajes por el noroeste brasileño*, 1903-1905. Bogotá: Editorial Universidad Nacional. 364 p.

Kondo, R. (2002). *En pos de los guahibos: Prehistóricos, históricos y actuales: Con pistas lingüísticas* (pp. 181). Bogotá, Colombia: Editorial Alberto Lleras Camargo.

Landaburu, J. (2000). Clasificación de las lenguas indígenas de Colombia. In: P. M. S. González, M. M. L. Rodríguez, & Instituto Caro y Cuervo (Eds.), *Lenguas indígenas de Colombia, una visión descriptive* (pp. 848). Santafé de Bogotá: Instituto Caro y Cuervo.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*, 357–359.

Lathrap, D. W. (1970). *The upper Amazon* (pp. 256). London: Thames & Hudson.

Le, S., Josse, J., & Husson, F. (2008). FactoMineR: An R package for multivariate analysis. *Journal of Statistical Software*, *25*, 1–18.

Lee, E. J., & Merriwether, D. A. (2015). Identification of whole mitochondrial genomes from Venezuela and implications on regional phylogenies in South America. *Human Biology*, *87*, 29–38.

Lewis, C. M., Jr., Lizarraga, B., Tito, R. Y., Lopez, P. W., Iannacone, G. C., Medina, A., … Stone, A. C. (2007). Mitochondrial DNA and the peopling of South America. *Human Biology*, *79*, 159–178.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*, 1754–1760.

Li, M., Schroeder, R., Ko, A., & Stoneking, M. (2012). Fidelity of capture-enrichment for mtDNA genome sequencing: influence of NUMTs. *Nucleic acids research*, *40*(18):e137.

Llamas, B., Fehren-Schmitz, L., Valverde, G., Soubrier, J., Mallick, S., Rohland, N., … Haak, W. (2016). Ancient mitochondrial DNA provides high-resolution time scale of the peopling of the Americas. *Science Advances*, *2*, e1501385.

Lowie, R. H. (1948). The tropical forests: An introduction. In: J. H. Steward (Ed.), *Handbook of South American Indians* (pp. 1–56). Washington: U.S. G.P.O.

Malhi, R. S., Schultz, B. A., & Smith, D. G. (2001). Distribution of mitochondrial DNA lineages among Native American tribes of Northeastern North America. *Human Biology*, *73*, 17–55.

Maricic, T., Whitten, M., & Paabo, S. (2010). Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One*, *5*, e14004.

Marrero, A. R., Silva-Junior, W. A., Bravi, C. M., Hutz, M. H., Petzl-Erler, M. L., Ruiz-Linares, A., … Bortolini, M. C. (2007). Demographic and evolutionary trajectories of the Guarani and Kaingang natives of Brazil. *American Journal of Physical Anthropology*, *132*, 301–310.

Meggers, B. J. (1954). Environmental limitation on the development of culture. *American Anthropologist*, *56*, 801–824.

Meggers, J. (1948). The archeology of the Amazon Basin. In J. H. Steward (Ed.), *Handbook of South American Indians* (pp. 149–166). Washington: U.S. G.P.O.

Melton, P. E., Briceno, I., Gomez, A., Devor, E. J., Bernal, J. E., & Crawford, M. H. (2007). Biological relationship between Central and South American Chibchan speaking populations: evidence from mtDNA. *American Journal of Physical Anthropology*, *133*, 753–770.

Miller, S. A., Dykes, D. D., & Polesky, H. F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research*, *16*, 1215.

Milton, K. (1984). Protein and Carbohydrate Resources of the Maku Indians of Northwestern Amazonia. *American Anthropologist*, *86*, 7–27.

Nettle, D. (1999). Linguistic diversity of the Americas can be reconciled with a recent colonization. *Proceedings of the National Academy of Science of the United States of America*, *96*, 3325–3329.

Nimuendajú, C. (1950). Reconhecimento dos rios Içána, Ayarí e Uaupés. *Journal de la Société des Américanistes*, *39*, 125–182.

Oota, H., Pakendorf, B., Weiss, G., von Haeseler, A., Pookajorn, S., Settheetham-Ishida, W., … Stoneking, M. (2005). Recent origin and cultural reversion of a hunter-gatherer group. *PLoS Biology*, *3*, e71.

Patin, E., Siddle, K. J., Laval, G., Quach, H., Harmant, C., Becker, N., … Quintana-Murci, L. (2014). The impact of agricultural emergence on the genetic history of African rainforest hunter-gatherers and agriculturalists. *Nature Communication*, *5*, 3163.

Perego, U. A., Achilli, A., Angerhofer, N., Accetturo, M., Pala, M., Olivieri, A., … Kong, Q. P. et al, (2009). Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. *Current Biology*, *19*, 1–8.

Perego, U. A., Angerhofer, N., Pala, M., Olivieri, A., Lancioni, H., Hooshiar Kashani, B., … Torroni, A. (2010). The initial peopling of the Americas: A growing number of founding mitochondrial genomes from Beringia. *Genome Research*, *20*, 1174–1179.

Pugach, I., Matveev, R., Spitsyn, V., Makarov, S., Novgorodov, I., Osakovsky, V., … Pakendorf, B. (2016). The complex admixture history and recent southern origins of Siberian populations. *Molecular Biology & Evolution*, *33*, 1777–1795.

Quinque, D., Kittler, R., Kayser, M., Stoneking, M., & Nasidze, I. (2006). Evaluation of saliva as a source of human DNA for population and association studies. *Analytical Biochemistry*, *353*, 272–277.

Ramachandran, S., Deshpande, O., Roseman, C. C., Rosenberg, N. A., Feldman, M. W., & Cavalli-Sforza, L. L. (2005). Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proceedings of the National Academy of Science of the United States of America*, *102*, 15942–15947.

Renaud, G., Kircher, M., Stenzel, U., & Kelso, J. (2013). freeIbis: an efficient basecaller with calibrated quality scores for Illumina sequencers. *Bioinformatics*, *29*, 1208–1209.

Renaud, G., Stenzel, U., & Kelso, J. (2014a). LeeHom: Adaptor trimming and merging for Illumina sequencing reads. *Nucleic Acids Research*, *42*, e141.

Renaud, G., Stenzel, U., Maricic, T., Wiebe, V., & Kelso, J. (2014b). deML: Robust demultiplexing of Illumina sequences using a likelihood-based approach. *Bioinformatics*, *31*(5):770–772.

Santös-Granero, F. (2002). The Arawakan Matrix: Ethos, Language, and History in Native South America. In: Hill JD, and Santos-Granero F, editors. *Comparative Arawakan histories: Rethinking language family and culture area in Amazonia* (pp. 25–50). Urbana: University of Illinois Press.

Schönberg, A., Theunert, C., Li, M., Stoneking, M., & Nasidze, I. (2011). High-throughput sequencing of complete human mtDNA genomes from the Caucasus and West Asia: high diversity and demographic inferences. *European Journal of Human Genetics*, *19*, 988–994.

Schurr, T. G. (2004). The peopling of the New World: Perspectives from molecular anthropology. *Annual Review of Anthropology*, *33*, 551–583.

Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Rohl, A., … Richards, M. B. (2009). Correcting for purifying selection: an improved human mitochondrial molecular clock. *American Journal of Human Genetics*, *84*, 740–759.

Sorensen, A. P. (1967). Multilingualism in the Northwest Amazon. *American Anthropologist*, *69*, 670–684.

Stenzel, K. (2005). *Multilingualism in the northwest Amazon, revisited. Memorias del Congreso de Idiomas Indígenas de Latinoamérica-II* (pp. 1–28). Austin: University of Texas.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*, 585–595.

Torres, M. M., Bravi, C. M., Bortolini, M. C., Duque, C., Callegari-Jacques, S., Ortiz, D., … Ruiz-Linares, A. (2006). A revertant of the major founder Native American haplogroup C common in populations from northern South America. *American Journal of Human Biology*, *18*, 59–65.

Torroni, A., Schurr, T. G., Cabell, M. F., Brown, M. D., Neel, J. V., Larsen, M., … Wallace, D. C. (1993). Asian affinities and continental radiation of the four founding Native American mtDNAs. *American Journal of Human Genetics*, *53*, 563–590.

Uribe Tobón, C. A., & Instituto Colombiano de Cultura, H. (1992). *Geografía humana de Colombia*. Santa Fe de Bogotá: Instituto Colombiano de Cultura Hispánica.

van Oven, M., & Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human Mutation*, *30*, E386–E394.

Verdu, P., Becker, N. S., Froment, A., Georges, M., Grugni, V., Quintana-Murci, L., … Austerlitz, F. (2013). Sociocultural behavior, sex-biased admixture, and effective population sizes in Central African Pygmies and non-Pygmies. *Molecular Biology & Evolution*, *30*, 918–937.

Vidal, S. M. (1997). Liderazgo y confederaciones multiétnicas amerindias en la amazonia luso-hispana del. *Siglo XVIII Revista Antropológica*, *87*, 19–46.

Vidal, S. M. (2002). *Secret religious cults and political leadership: Multiethnic confederacies from Northwestern Amazonia* (pp. 248–268). Comparative Arawakan Histories: University of Illinois Press.

Wallace, A. R. (1853). *A narrative of travels on the Amazon and Rio Negro: With an account of the native tribes, and observations on the climate, geology, and natural history of the Amazon valley* (Vol. *viii*, pp. 541). London: Reeve and Co.

Wang, S., Lewis, C. M., Jakobsson, M., Ramachandran, S., Ray, N., Bedoya, G., ... Gallo, C. et al, (2007). Genetic variation and population structure in native Americans. *PLoS Genetics*, *3*, e185.

Wright, R. M. (2002). *Prophetic traditions among the Baniwa and other Arawakan peoples of the northwest Amazon* (pp. 269–294). Comparative Arawakan Histories: University of Illinois Press.

Yunusbayev, B., Metspalu, M., Jarve, M., Kutuev, I., Rootsi, S., Metspalu, E., ... Villems, R. (2012). The Caucasus as an asymmetric semipermeable barrier to ancient human migrations. *Molecular Biology & Evolution*, *29*, 359–365.

Zucchi, A. (2002). A new model of the Northern Arawakan expansion. In J. D. Hill & F. Santos-Granero (Eds.), *Comparative Arawakan histories: Rethinking language family and culture area in Amazonia* (pp. 199–222). Urbana: University of Illinois Press.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Arias L, Barbieri C, Barreto G, Stoneking M, Pakendorf B. High-resolution mitochondrial DNA analysis sheds light on human diversity, cultural interactions, and population mobility in Northwestern Amazonia. *Am J Phys Anthropol*. 2018;165:238–255. https://doi.org/10.1002/ajpa.23345