AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 155:600–609 (2014)

Between Andes and Amazon: the Genetic Profile of the Arawak-Speaking Yanesha

Chiara Barbieri,[1] Paul Heggarty,[2] Daniele Yang Yao,[1] Gianmarco Ferri,[3] Sara De Fanti,[1] Stefania Sarno,[1] Graziella Ciani,[1] Alessio Boattini,[1] Donata Luiselli,[1]* and Davide Pettener[1]

[1]*Department of Biological, Geological and Environmental Sciences, University of Bologna, 40126 Bologna, Italy*
[2]*Department of Linguistics, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany*
[3]*Dipartimento di Medicina Diagnostica, Clinica e di Sanità Pubblica, Università degli Studi di Modena e Reggio Emilia, 41124 Modena, Italy*

KEY WORDS mtDNA; Y chromosome; STR; South America; language

ABSTRACT The Yanesha are a Peruvian population who inhabit an environment transitional between the Andes and Amazonia. They present cultural traits characteristic of both regions, including in the language they speak: Yanesha belongs to the Arawak language family (which very likely originated in the Amazon/Orinoco lowlands), but has been strongly influenced by Quechua, the most widespread language family of the Andes. Given their location and cultural make-up, the Yanesha make for an ideal case study for investigating language and population dynamics across the Andes-Amazonia divide. In this study, we analyze data from high and mid-altitude Yanesha villages, both Y chromosome (17 STRs and 16 SNPs diagnostic for assigning haplogroups) and mtDNA data (control region sequences and 3 SNPs and one INDEL diagnostic for assigning haplogroups). We uncover sex-biased genetic trends that probably arose in different stages: first, a male-biased gene flow from Andean regions, genetically consistent with highland Quechua-speakers and probably dating back to Inca expansion; and second, traces of European contact consistent with Y chromosome lineages from Italy and Tyrol, in line with historically documented migrations. Most research in the history, archaeology and linguistics of South America has long been characterized by perceptions of a sharp divide between the Andes and Amazonia; our results serve as a clear case-study confirming demographic flows across that 'divide'. Am J Phys Anthropol 155:600–609, 2014. © 2014 The Authors. American journal of physical Anthropology published by Wiley Periodocals, Inc.

INTRODUCTION

Particularly since the 1990s, population genetic studies have taken a keen interest in the indigenous populations of the Americas, given their unique demographic background (Schurr et al., 1990; Rothhammer and Silva, 1992; Torroni et al., 1992). The Americas were the last major continents to be settled by humans, probably by a small founding population around 15–17 kya (Tamm et al., 2007; Fagundes et al., 2008; Kitchen et al., 2008), yet they are home to surprisingly wide cultural and linguistic diversity (Nettle, 1999).

On a broad scale, the available genetic data clearly point towards a strong founder effect and a rapid spread through the continent (Wallace et al., 1985; Lewis et al., 2007), from the Bering Strait to Patagonia; this scenario is in agreement with archaeological indications of human presence in southern Chile from c. 14,500 ya (Dillehay et al., 2008), although no actual human remains of this antiquity have yet been found there. Uniparental genetic markers in native American populations are affected by this bottleneck (Torroni et al., 1993; Bortolini et al., 2003; Schurr and Sherry, 2004; Bisso-Machado et al., 2012; Battaglia et al., 2013): for mitochondrial DNA (mtDNA) only four macro haplogroups are found at appreciable frequencies (A, B, C, and D), while for the Y chromosome just a single lineage is predominant (macro-haplogroup Q). This limited genetic variability contrasts with high differentiation between-populations, probably arisen through drift (Yang et al., 2010; Bisso-Machado et al., 2012; Roewer et al., 2013).

On a local scale, meanwhile, the complex demographic histories of precolonial times are difficult to reconstruct, and different genetic data often hint at contradictory explanations (Wang et al., 2007; Bisso-Machado et al., 2012). Recent studies have focused on the dynamics behind the current distribution of language families, or on the role of the environment in shaping demographic

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: The European Research Council; Grant number: ERC-2011-AdG 295733 grant (Langelin); Grant sponsor: The Faculty of Mathematical, Physical and Natural Sciences of the University of Bologna.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

*Correspondence to: Donata Luiselli; Department of Biological, Geological and Environmental Sciences, Univeristy of Bologna, via Selmi 3, 40126 Bologna, Italy. E-mail: donata.luiselli@unibo.it

Received 3 July 2014; accepted 4 September 2014

DOI: 10.1002/ajpa.22616
Published online 17 September 2014 in Wiley Online Library (wileyonlinelibrary.com).

© 2014 THE AUTHORS. AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY PUBLISHED BY WILEY PERIODOCALS, INC.

trajectories (Rothhammer and Dillehay, 2009; Callegari-Jacques et al., 2011; Amorim et al., 2013; Ramallo et al., 2013; Roewer et al., 2013; Sandoval et al., 2013).

Both linguistic and environmental factors are unarguably important in understanding cultural diversity across South America. Much of the continent falls under one of two extreme environments: the Amazon rainforest; or the Andes, the world's second-highest mountain range running north-south through the tropics, parallel to the Pacific coast. The Central Andes saw the emergence of one of the world's few independent civilizations, as complex societies arose here up to five millennia ago. However, the main indigenous language family of the Andes, Quechua, expanded only in the last 1500 years or so. Its spread to the highlands of Ecuador and Bolivia dates to the Inca and early Spanish colonial periods, but Quechua's core range in the highlands of central and southern Peru dates back to earlier expansive processes, with an obvious candidate being the 'Middle Horizon' expansion, c. 550–1000 CE, originating from the site of Wari, near modern Ayacucho (Beresford-Jones and Heggarty, 2013). Both Wari and the Incas are widely presumed to have fostered significant population movements through the Central Andes highlands, leading to considerable genetic homogenization.

In the lowlands east of the Andes, the archaeological record is much less detailed, frustrated also by poorer visibility and preservation in the Amazonian rainforest. Certainly, the linguistic panorama is unusually fragmented and diverse. Four major language families are spoken—Arawak, Carib, Tupí, and Je— but in scattered distributions and interspersed with dozens of other minor families and isolate language lineages not demonstrably relatable to any other (Campbell, 1997). Arawak nonetheless ranks as the most widely dispersed family in the Americas, scattered from the Antilles to pockets even in northernmost Argentina. Various competing hypotheses disagree on where the family originated, when it began to expand, and how it did so. The traditional hypothesis places the homeland in the northern Amazon, between the Rio Negro and Orinoco, emerging from there in an agriculture-based demic expansion, primarily along major rivers, from c. 500 BCE (Aikhenvald, 1999). Walker and Ribeiro (2011), however, argue instead for Western Amazonia, based on a Bayesian phylogenetic analysis of basic lexical data from across the family, and propose manioc cultivation as a potential driver for population expansion. Others downplay the role of demographic growth and argue that Arawak dispersal was more through trade and cultural prestige (e.g., Hornborg et al., 2005).

This study focuses on the Yanesha (or 'Amuesha') of Central Peru, a population living in the Selva Central region and who speak a language of the southern branch of the Arawak family. The Yanesha are dispersed across a range of ecological settings at different altitudes, a transitional environment between the Andes and Amazon that may have favored contact flows both in the gene pool and in cultural characteristics (see references in Daigneault, 2010). Although the Yanesha language is fundamentally of Arawak (and thus Amazonian) origin, some aspects of its sound system, and a few characteristics even of its grammar, betray contact influences from several other languages. Above all, part of Yanesha's native Arawak lexicon has been augmented or replaced by a significant influx of loanwords. These came principally from the nearby Yaru dialect of Central Quechua, although a few were borrowed from other forms of Quechua, from nearby Amazonian languages, and from other unidentified source languages (Adelaar, 2006). Quechua influence seems to have largely ceased a few centuries ago, but instead Spanish has brought another layer of loanwords, since the Selva Central was first colonized by Spanish Franciscan friars from 1635. Nonetheless, the main sources of European migrations, starting in the 1850s, were Prussia and Italy (Contreras, 2007).

The latest census (INEI) puts the overall Yanesha population at 9213, in villages scattered across three main altitude levels: High Selva (HS) up to 1,800 m, Intermediate Selva (IS) up to 800 m and Low Selva (LS) up to 100 m. As a rule, the higher the altitude the greater the degree of urbanization and westernization, the lower communities being the most isolated. What historical sources fail to reveal is the extent to which the HS and IS settlements reflect recent incoming Quechua (and even European) populations from the Andes, who switched language to Arawak here, or original settlements by Arawak-speaking populations arriving from Amazonia.

This study provides novel genetic data for the Yanesha people, and discusses their genetic variability against the South American background. To this end, we analyzed uniparental markers in 214 HS and IS samples, providing Y chromosome data (17 STRs and 16 SNPs diagnostic for assigning haplogroups) and mtDNA data (control region sequences and 3 SNPs and one INDEL diagnostic for assigning haplogroups). The aims of our research are to: (1) characterize the genetic diversity of the Yanesha population, and uncover potential differences between the HS and IS communities; (2) estimate the degree of any sex-biased patterns of contact; (3) investigate the origin of the Yanesha population, by distinguishing Amazonian, Andean, and European gene flows; and (4) explore a putative genetic signature for Quechua and Arawak speakers in South America.

MATERIAL AND METHODS

Data collection and DNA extraction

A fieldwork expedition was carried out in the Reserva Comunal Yanesha (Peru, Pasco, and Junín departments). Buccal swabs were collected from unrelated individuals from 11 villages: 6 from the High Selva (HS) and 5 from the Intermediate Selva (IS) (Table 1, Supporting Information Fig. S1), covering two of the three altitude levels where Yanesha communities are found. This study and the collection of samples were approved by the representative of the Yanesha political association, the FECONAYA (Federación de Comunidades Nativas Yanesha), and a formal agreement was signed with representatives of each Yanesha village. All participants were volunteers, and signed an informed consent after being notified about the purpose of the study. The study was conducted in accordance with the Declaration of Helsinki. DNA was extracted by a Salting Out protocol modified from Miller et al. (1988).

Y chromosome genotyping

The Y chromosome was genotyped for all males in our sample (163 individuals). Sixteen SNPs were tested with a SNaPshot™ Multiplex kit (Applied Biosystems) as described in Brion et al. (2005). The first multiplex identified major European clades which were subsequently confirmed through their Y-STR profiles (Bayesian inference by

TABLE 1. Details on the Yanesha villages sampled

Village	Pop (census)	*N* Y ch	*n* mtDNA	Selva[a]	Altitude (m)	Admin. department (Peru)	District
Mayme	135	12	13	HS	1500	Pasco	Villarica
Milagros	111	9	16	HS	1500	Pasco	Villarica
Nagazu	276	21	23	HS	1500	Pasco	Villarica
Tsachopen	511	20	29	HS	1800	Pasco	Huancabamba
Union de la Selva Cacazu	173	11	17	HS	1200	Pasco	Villarica
Alto Yurinaqui	544	12	13	HS	1200	Junin	Perené
Alto Iscozacín	43	9	13	IS	300	Pasco	Palcazu
Loma Linda—Laguna	790	15	16	IS	300	Pasco	Palcazu
Nueva Esperanza	229	8	6	IS	300	Pasco	Palcazu
Siete de Junio[b]	1424	29	37	IS	300	Pasco	Palcazu
Shiringamazu	631	17	31	IS	300	Pasco	Palcazu
Total		163	214				

[a] HS: High Selva; IS: Intermediate Selva.
[b] Villa America; Centro Castilla; Pampa Cocha; Centro Esperanza; Centro Chispa; Palma; Comparachimas; Puerto Alegre; Conaz; Pampa Hermosa.

http://www.hprg.com/hapest5/). The individuals who belong to the native American Q haplogroup were further screened with a second multiplex to assign potential sublineages and to verify whether the common Q1a3a* was the only Amerindian haplogroup present (see Sevini et al., 2013 for details about the primers design and the protocol used). Haplogroups were named according to the current Y chromosome nomenclature given by the ISOGG consortium (http://www.isogg.org/tree/), but also according to the commonly used nomenclature from Karafet et al. (2008), for convenience. Y-STRs typing was performed using the AmpFlSTR® Yfiler™ (Applied Biosystems) following the manufacturer's instructions. Amplified products were separated and detected by an ABI PRISM® 3130 Genetic Analyzer (Applied Biosystems) and analysed using GeneMapper® ID v3.2 analysis software (Applied Biosystems). All 163 individuals were successfully genotyped for all 17 loci; their haplotypes are listed in Supporting Information Table S1.

mtDNA genotyping

All 214 samples were amplified and sequenced for the hyper variable segment (HVS)-I with the primers L15996 and H16401, using 3.1 BigDye Terminator protocol supplied by Applied Biosystems™ on an ABI PRISM 3730 genetic analyzer. Samples were sequenced in both directions to avoid sequencing errors. Sequences were checked with Chromas Lite 2.01 and aligned with DNA Alignment Software (http://www.fluxusengineering.com/align.htm). Polymorphic sites are referred to the RSRS (Behar et al., 2012). Sequences are available on GenBank under accession numbers KM090632-KM090845. All samples were additionally genotyped for four diagnostic SNPs for the main mtDNA Native Amerindian lineages A, B, C, and D, using two restriction enzymes (*Hae*III at position 663 for Haplogroup A and *Alu*I for both Haplogroups C and D at positions 5,176 and 13,262, respectively) and the INDEL 8281–8289d specific to Haplogroup B. D-loop sequence information was integrated to refine haplogroup assignment. Nomenclature was assigned according to the latest version of Phylotree Build 16 (http://www.phylotree.org, van Oven and Kayser, 2009).

Data analysis

Diversity values were calculated in R: for mtDNA, nucleotide diversity and variance were calculated with the *Pegas* package (Paradis, 2010); sequence diversity for mtDNA and the Y chromosome, as well as Y chromosome variance, were calculated using in-house scripts. Correspondence analysis (CA) was performed with the *ca* package (Nenadic and Greenacre, 2007). AMOVA and Φ_{ST} distance matrices were computed in Arlequin ver. 3.11. A neighbor-joining tree of the populations was generated from a Φ_{ST} distance matrix using the "nj" function within the *ape* package (Paradis et al., 2004). Nonmetric multidimensional scaling (nMDS) analyses were performed using the "isoMDS" function within the *MASS* package (Venables and Ripley, 2002). A Mantel test was performed between genetic (Φ_{ST}) and geographic distances using the *vegan* package (Oksanen et al., 2012).

Two comparative datasets of South American native populations were collected from available literature. For the Y chromosome, STR haplotypes for 17 loci were collected for 1,372 individuals from 60 populations covering various major language families (Supporting Information Table S2); only individuals belonging to native haplogroup Q were considered. For mtDNA, HVS-I sequences of 360 bp were collected for 2,695 individuals from 75 native populations (Supporting Information Table S3); only individuals belonging to native haplogroups A, B, C, and D were considered. Both datasets were screened to minimize the impact of missing data and sequencing errors, removing individuals with low quality results; only populations with 10 or more individuals were considered. Figure 1a locates the populations in the Y chromosome dataset and Figure 1b those in the mtDNA dataset; note that the populations in the two datasets do not overlap entirely. A further set of populations was selected as present in both datasets, giving an overlapping dataset of 39 populations (column K in Supporting Information Tables S2 and S3). The South American populations were divided into seven groups based on linguistic and geographic features relevant for our case study. These groups are: Andean, Andean Quechua, Andean Arawak, Amazonian, Amazonian Arawak, Northern South American, Northern South American Arawak. A further dataset was collected from 22 European populations, including 17 STR haplotypes for various non-Amerindian haplogroups (Supporting Information Table S4). nMDS plots and NJ trees for the comparative datasets were generated from Φ_{ST} (mtDNA sequences) and R_{ST} (Y chromosome STR) distance matrices as described above.

For each population in the dataset, the proportion of haplotypes identical or very similar to the HS Yanesha

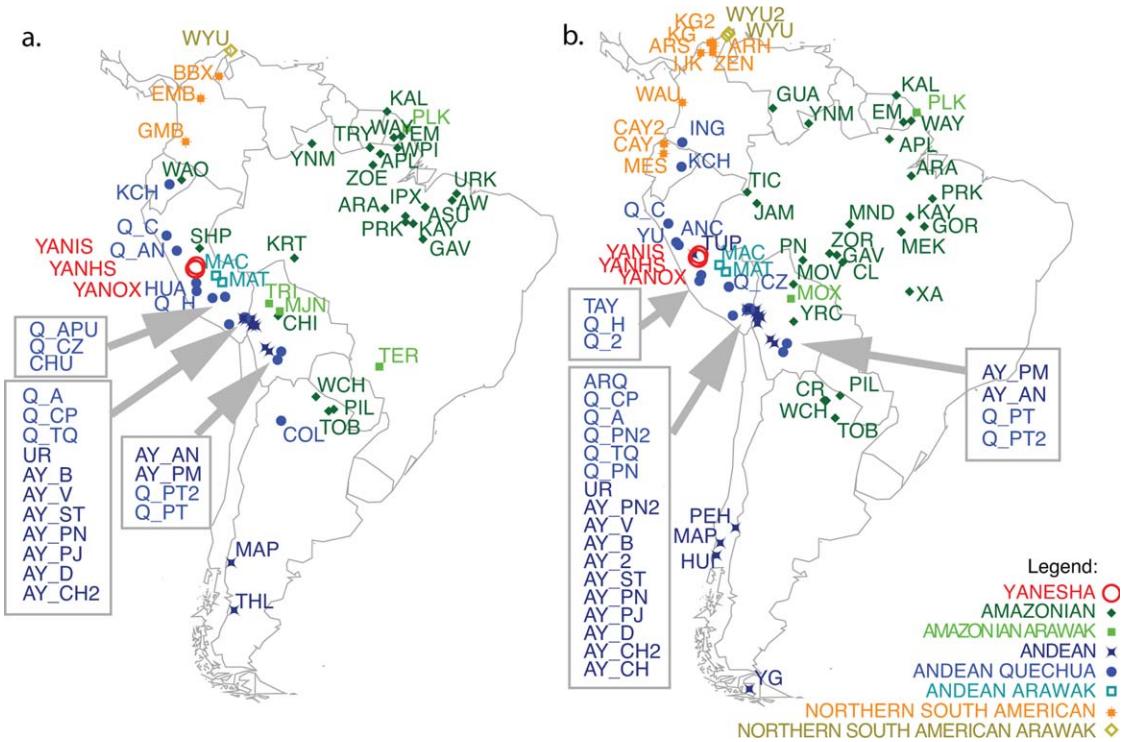

Fig. 1. Maps showing the location of the population samples included in the comparative datasets for (a) Y chromosome and (b) mtDNA. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.].

and IS Yanesha, respectively, was visualized on a map using the *maps* package (Becker et al., 2013). For the Y chromosome, a matrix of pairwise distances based on number of generations was calculated using an in-house R script (after excluding unstable loci DSY385a and b): each locus was divided by its mutation rate as reported in the Y-STR haplotype reference database (http://www.yhrd.org), and the TMRCA in the number of generations was averaged over the 15 loci. Three levels of threshold were considered: 0 generations difference (identical haplotypes); <25 generations difference (corresponding to a time frame of 750 years, with a generation time of 30 years); <50 generations difference (corresponding to 1,500 years), and <100 generations (corresponding to 3,000 years). For mtDNA, a matrix of pairwise sequence diversity was calculated with the dist.dna function within the *ape* package, and only the proportion of identical haplotypes was considered.

RESULTS

Y chromosome

The Yanesha populations show a predominance of native haplogroup Q, with different frequencies in the High Selva (65%) and Intermediate Selva (72%; Supporting Information Table S5). Interestingly, two samples in the HS Yanesha display the basal M346–Q1a3* mutation but not the downstream M3–Q1a3a1 marker. This pattern is very rare in South America (Bailliet et al., 2009). One individual from the IS Yanesha belongs to haplogroup M2-E1b1a* which is more common in sub-Saharan Africa (De Filippo et al., 2011). R1, the most common haplogroup in Europe, is at 20 and 10% in HS and IS, respectively, while E1b1b1, F, I, J2, K, L, and T occur more rarely, with a maximum frequency of 7% E1b1b1 in HS; all of these haplogroups are commonly found in Europe (Semino et al., 2004). Haplogroup proportions are also variable at the village level: the highest native ancestry is found in the most isolated IS villages (such as Siete de Junio, Shiringamazu and Nueva Esperanza, where it is above 75%). Nevertheless, the genetic variance between the 11 villages as measured via AMOVA (Table 2) is quite low (1.26% variance between villages, $P < 0.05$), but rises slightly when only native Q lineages are considered (6.46%, $P < 0.05$); in the latter case, analysis at the village level should be taken with caution because of the even more reduced sample size. Genetic variance is low and nonsignificant also between the two groups HS and IS, both for all haplotypes and for the Q haplotypes only.

The 52 non-native haplotypes found in both HS and IS Yanesha were then combined and compared to a set of 17 loci STR from 22 European populations (Supporting Information Table S4). The European component in the Yanesha has its least R_{ST} distance (nonsignificant) with a set of haplotypes from Central Italy, followed by Northeast Italy and Sicily; they are also close to southern Italy and Austria. An nMDS plot visualizes the relationships between the Yanesha and European populations (Supporting Information Fig. S2); the Yanesha appear quite isolated, surrounded by the closest pairwise distance couples as well as by populations from the Iberian Peninsula and from Switzerland. Yanesha Eurasian haplotypes are shared directly only with Central, North and Northeastern Italy, Austrian Tyrol, South Germany, Northeast Spain and Central Spain.

The native Q haplotypes in HS and IS Yanesha were next compared to other Q haplotypes in a dataset of 62 South American populations. The nMDS plot based on R_{ST} distances (Fig. 2a) separates two clusters on the x

TABLE 2. AMOVA analysis of Y chromosome STR data and mtDNA data

	n pops	*n* groups	Between groups	Between pops	Between individuals
STR data					
All populations	62	1		46.67*	53.33
All Arawak	10	1		41.18*	58.82
All Quechua	15	1		32.79*	67.21
All Andes	28	1		28.83*	71.17
All Andes plus Yanesha	31	1		26.54*	73.46
All Amazon	27	1		47.21*	52.79
Yanesha and Europe (non-Q haplotypes)	14	1		9.53*	90.47
Yanesha HS vs. IS	11	2	0.74	0.85	98.42
Yanesha HS vs. IS (only Q haplotypes)	11	2	−0.82	6.95*	96.77
All Yanesha villages	11	1		1.26**	98.74
All Yanesha villages (only Q haplotypes)	11	1		6.46**	93.54
mtDNA data					
All populations	77	1		19.18*	80.82
All Arawak	9	1		7.4*	92.6
All Quechua	16	1		9.45*	90.55
All Andes	37	1		13.53*	86.47
All Andes plus Yanesha	40	1		11.89*	88.11
All Amazon	26	1		17.42*	82.58
Yanesha HS vs. IS	11	2	0.83	2.05	97.12
All Yanesha villages	11	1		2.52**	97.48

*$P < 0.001$.
**$P < 0.005$.

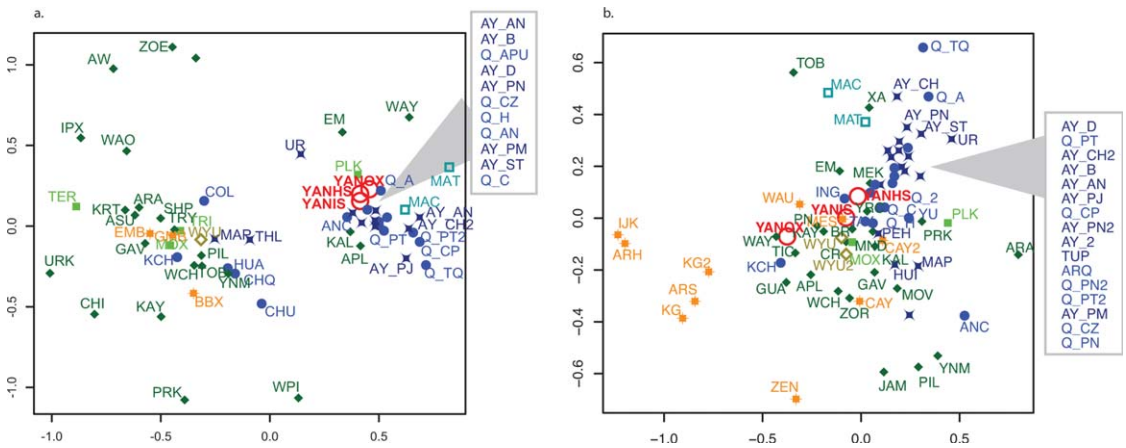

Fig. 2. MDS plots based on (a) Y chromosome R_{ST} distances (stress value: 16.3), and (b) mtDNA ΦST distances (stress value: 17.5). For the legend of major geographic/linguistic groups refer to Figure 1. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.].

axis, with a majority of central Andean populations on the right, and a majority of Amazonian and Northern populations on the left. The Yanesha populations fall towards the Andean cluster, together with five other populations from northeastern South America, between French Guyana and Brazil. The shortest R_{ST} distances (<0.2) to the two Yanesha populations are with a diverse set of Quechua and Aymara speakers, the Andean Arawak Machiguenga, the Amazonian Arawak Palikur, and the Amazonian Kalina and Apalai (Supporting Information Table S6a). The NJ tree based on R_{ST} distances (Supporting Information Fig. S3a) highlights the position of genetic isolates at the tip of the longest branches: these are more common among Amazonian populations.

The proportion of paternal genetic variance in our dataset is shown by the AMOVA results in Table 2. When only haplogroup Q is included, variance among populations across South America is very high (47%), especially when compared with the variance among the 23 European populations (10%), which include haplotypes from different haplogroups. Among Amazonian populations, variance is similarly high (47%), while variance among Andean populations is lower (29%), and even more so if the Yanesha are included in the Andean pool (27%). The 10 Arawak populations in the dataset have a variance of 41%, showing no homogeneity. All these values are highly significant ($P = 0$).

The haplotypes of the HS and IS Yanesha were compared against each population in the dataset to detect the pairwise proportions of identical haplotypes, and the proportion of haplotypes which diverged less than 25, 50, or 100 generations ago; maps in Figure 3a and Supporting Information Figure S4 help to visualize the geographic distribution of these relationships. HS Yanesha share identical haplotypes with Quechua and Aymara speakers from the Potosí, La Paz, Lake Titicaca, Cuzco, and Huancavelica regions, as well as with the Andean Arawak Machiguenga, while the IS Yanesha share

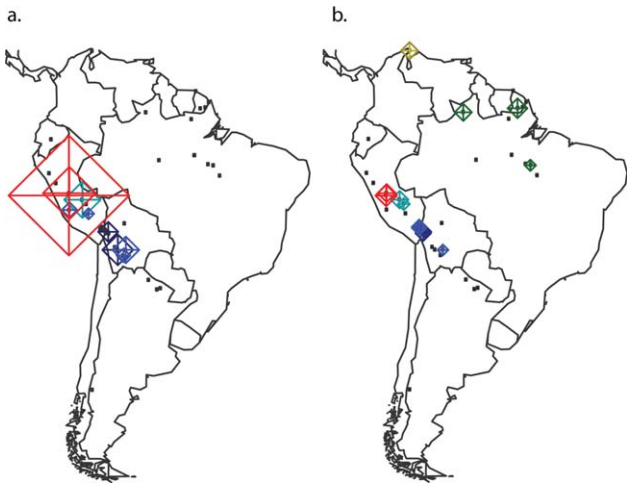

Fig. 3. Sharing of identical haplotypes between HS Yanesha and the populations from the overlapping dataset. (a) Populations which share identical haplotypes are marked and (b) populations which share >70% of their mtDNA HVSI sequences are marked. Square size is proportional to the amount of sharing.

identical haplotypes only with other Yanesha populations and with Quechua speakers from Kaquiabamba/Apurimac, Machiguenga, and Aymara from Santa Ana/La Paz (Supporting Information Fig. S4a). Haplotypes which diverged <25 generations ago for both HS and IS Yanesha are found in the Machiguenga and in several Aymara and Quechua populations from points lying between the Yanesha sampling area and as far as La Paz and Potosí in Bolivia (Supporting Information Fig. S4b). Haplotypes which diverged <50 generations ago are found also in the Kalina and Toba, in minimal percentage and only for HS Yanesha, while haplotypes which diverged <100 generations ago are found in almost all populations in the dataset, including Northern and Amazonian populations (Supporting Information Fig. S4c and S4d).

Diversity values for both the HS and IS Yanesha and the reference dataset are reported in Supporting Information Table S2: the Yanesha populations show fairly high variance and haplotype diversity when compared with the other Amerindian populations.

mtDNA

All 214 samples show maternal Amerindian ancestry and neither European nor African haplogroups are present. Haplogroup B is the most widely represented in both populations, with 55% in the HS and 39% in the IS; haplogroup D is the least widely represented in HS (5%) but is present at 20% in IS (Supporting Information Table S5).

The haplogroup distribution is significantly different between the HS and IS villages ($\chi^2 = 14.91$, 3 *df*, $P < 0.01$), with haplogroups B and D, in particular, present in different proportions. The haplogroup distribution of a database of South American populations (mostly taken from the comprehensive review of Bisso-Machado et al., 2012) is listed in Supporting Information Table S7 and visualized using a CA plot in Supporting Information Figure S5. The CA plot summarizes known trends in haplogroup distributions across South America, with B characteristic of the Andes, A of the Northern regions of the continent, and C and D more common in Amazonia (with many exceptions probably due to strong drift effects in more isolated populations). The Yanesha populations, in the center of the plot, have a balanced haplogroup composition, with only the HS looking more similar to other Andean populations because of the higher presence of B.

Variability at the sequence level was first tested for the 11 Yanesha villages: their maternal genetic pool appears more homogeneous than does their Y chromosome pool (for the Q lineages only), with a low but significant between-populations variance of 2.52% ($P < 0.05$) (Table 2). The two groups of HS and IS villages cannot be distinguished (between group variance 0.83%, not significant).

The HS and IS Yanesha populations were then included in a dataset of 77 Amerindian populations for a comparison at a continental level (Supporting Information Table S3, Fig. 1b). Pairwise genetic distances (Φ_{ST}) were used to build a nMDS plot and an NJ tree. The Yanesha populations appear in the middle of the nMDS plot (Fig. 2b), together with the majority of Amazonian and Andean populations. Towards the top of the plot, a few Andean populations (Aymara, Quechua, and Arawak Machiguenga) form a separate cluster together with the Amazonian Xavante and Toba. To the left of the plot, five Northern populations are distinguished. The NJ tree (Supporting Information Fig. S3b) does not reveal more insights into maternal variation: it highlights the divergence of the five Northern populations, which are located at the end of long branches, and confirms that the HS and IS Yanesha are not particularly differentiated from the rest of the continent: they again appear at the center of the tree, on very short branches. The shortest Φ_{ST} distances are to the Brazilian Yuracare for both populations; other short distances (<0.03) are to populations that are different for the two Yanesha groups (Supporting Information Table S6b). HS Yanesha are similar to Quechua and Aymara populations from Cuzco, Puno, Tayacaja, and Potosí, as well as the Brazilian Mekranati and Emerillon from French Guyana. The IS Yanesha, on the other hand, show proximity to the neighboring Quechua speakers from Tayacaja and to Quechua from the Titicaca area, and to other Amazonian populations too: the Bolivian Arawak Moxo, the Brazilian Kayapo, Munduruku and Pacaas Novos, the Criollos of Gran Chaco, French Guyana Emerillon and finally a Northern population, the Arawak Wayuu from Colombia.

AMOVA analysis (Table 2) reveals a high and significant degree of variance between the populations in the dataset (19%). Contrary to what emerged from the Y chromosome analysis, the nine Arawak populations in our dataset are less heterogeneous than the 16 Quechua (7 vs. 9%). The Amazonian populations are more heterogeneous than the Andean ones (17 vs. 14%).

The haplotypes of the HS and IS Yanesha were compared to every other population in the dataset to detect the proportion of identical haplotypes; maps in Figure 3b and Supporting Information Figure S6 help to visualize the geographic distribution of these relationships.

HVSI sequences identical to sequences in both the HS and IS Yanesha are found in almost all the populations in the dataset (Supporting Information Fig. S6a). To identify the main contributions from recent contact, we selected only those populations that share more haplotypes than the average for the whole dataset

(Supporting Information Fig. S6b). The sharing pattern is again similar for HS and IS, and involves Andean populations from southern Peru, the Titicaca area, and parts of Bolivia and Chile; some populations scattered across the Amazon basin, and several populations from northern Colombia. To narrow this list down to only the very strongest shared haplotype signal, we retained only those populations with a far higher percentage of shared haplotypes than the average (for HS, we kept only populations that share >70% of their haplotypes, and for IS only populations that share >40%) (Fig. 3b, Supporting Information Fig. S6c): these populations include the Andean Arawak Machiguenga, Quechua, and Aymara from the Lake Titicaca area, Emerillon from French Guyana, four Northern populations including the Arawak Wayuu, and a few other scattered Amazonian populations. On average, HS shows higher percentages of haplotypes shared with other populations than does IS.

As with the Y chromosome, in their mtDNA diversity (nucleotide diversity and haplotype diversity), both HS and IS populations rank among the most diverse in the entire South American dataset (Supporting Information Table S3).

DISCUSSION

The genetic make-up of the Yanesha villages

The Yanesha population of central Peru makes for a valuable case study because of their distinctive patterns of pre- and postcolonial contact, in unique cultural and environmental contexts. Before the waves of European migration (from different sources and at different times) the Yanesha were subjected to the influences of both the Andean and Amazonian ecologies, effectively forming them into a bridge between those two environmental extremes.

The genetic data (17 Y chromosome STRs and HVSI mtDNA sequences) collected from 11 Yanesha villages were first analyzed to assess the degree of homogeneity of the whole population. The Yanesha villages appear genetically homogeneous for both markers, with negligible AMOVA variance between villages (Table 2). The villages can be ascribed to two different environments by altitude: we therefore divide them into High Selva (HS Yanesha, 1,800–1,200 m) and Intermediate Selva (IS Yanesha, 300 m)—see Table 1 and Supporting Information Figure S1. The two groups have a different haplogroup composition, but are not genetically distinguishable in their STR haplotypes or mtDNA pool (Table 2). Nevertheless, we retained this subdivision by altitude for the rest of the analysis in order to highlight possible different demographic behaviors. The two Yanesha populations in fact exhibit different patterns of contact with other populations in the South American dataset, with some reduced variability in the more isolated IS villages. In general, villages of the IS are less easily accessible; it is thus no surprise that for both Y chromosome and mtDNA, the IS Yanesha share fewer identical haplotypes with other indigenous populations. In mtDNA, haplogroup composition differs between HS and IS (CA plot Supporting Information Fig. S5): the HS villages have a higher incidence of haplogroup B, characteristic of Andean populations, while IS villages show higher incidence of haplogroup D, typical of Amazonian populations (Bisso-Machado et al., 2012).

Finally, in both populations, diversity values are high by comparison to the other populations in the database (Supporting Information Tables S2, S3): this effect is not driven only by the higher sample size, since other populations from the dataset with less than half individuals have similarly high values of diversity (for example Kichwa and Wayampi). We assume a negligible effect of isolation (since the Yanesha did indeed experience gene flow from other populations), and/or a higher effective population size.

We conclude that the different environments inhabited by the Yanesha do not represent a barrier to internal or external exchange. The contact experienced by the Yanesha was probably for commercial reasons: in fact the Yanesha controlled the salt trade, owning the *Cerro de la Sal* (Salt Mountain) on the Paucartambo River between San Luis de Shuaro and Villa Rica. The villages at higher altitude acted as a gateway to the extensive Andean exchange network already in existence long before the Incas (Lumbreras, 1974). While the genetic pool of the Yanesha villages is homogeneous overall, the IS villages are more isolated and tend to exchange less, and a minor difference between the HS and IS villages can be discerned in their mtDNA haplogroup composition.

Recent contact with European colonists

Native and non-native components in Amerindian populations can be readily distinguished by their haplogroup composition. While all Yanesha individuals belong to one of the four native mtDNA haplogroups A, B, C, or D, only 70% of the Y chromosome haplogroups in the Yanesha belong to the native Q lineage. The remaining 30% is made up of a patchwork of typically Eurasian or even African haplogroups, likely introduced by European colonists in recent centuries (Supporting Information Table S5).

The pattern of contact in this situation is typically sex-biased, with immigrant men marrying local women. From the pattern of shared haplotypes and from the nMDS plot of pairwise STR similarity, the closest European populations in our database are those from the Italian peninsula (Central, Sicily, and northeast in particular): this would reflect a recent flow from Italian immigrants into the area following the War of the Pacific in 1884. Some similarities are also shared with Austrian Tyrolese, due to the establishment of German-speaking colonies in the Pozuzo valley in the nineteenth century. This was a consequence of the politics of the Peruvian government, anxious to colonize its Amazonian territories, and specifically the Selva Central, in order to counter Brazilian expansion from the west. A contingent from Tyrol and Prussia emigrated to Peru in 1857 (Contreras, 2007), and to this day the architecture and customs of the Pozuzo valley still bear traces of German tradition.

Ancient contacts between Andean and Amazonian cultures

Looking at the native component in both mtDNA and the Y chromosome, we attempted to disentangle the genetic contribution from the Amazonian and the Andean sources in the Yanesha. The Yanesha language well reflects the past stratification of those inputs: it belongs to theArawak language family, but certain characteristics in its sound and grammatical structures, and pervasive lexical borrowing, reflect very intense contact with Quechua-speaking inhabitants of the Andes. Indeed

so far-reaching were these effects that Yanesha was initially considered of uncertain classification, or even a language isolate (Adelaar, 2006).

These different contributions can be visualized in the shared haplotypes map (Fig. 3, Supporting Information Fig. S4). The mtDNA HVSI is less informative for limited recent contact effects, since it retains a level of shared haplotypes compatible with first settlement of South America. The Yanesha share haplotypes with most populations across the dataset, up to averages of 40–50% of the total number of sequences per population. Further resolution in this maternal marker, particularly full mtDNA sequence data, should in principle help distinguish where such sharing of partial haplotypes is due to chance (or drift), or to actual population contacts.

More fine-grained insights are to be had from the 17 STRs that define the Y chromosome haplotypes. The sharing pattern for our Yanesha samples clearly points towards Andean populations, speakers of either Quechua or Aymara (as described in the original source reference for those samples), distributed through the Andean highlands from the Yanesha area southwards as far as the Lake Titicaca region and central-western Bolivia. Most language contact impacts in Yanesha point to a specific source within the Quechua family: the now moribund Yaru dialect, spoken in highland regions geographically contiguous to the Yanesha. There is also a small set of loaned Quechua vocabulary reminiscent specifically of the political and cultural dominance of the Incas. The stronger linguistic signal, however, is not this late interaction with the Incas, but an intense and prolonged interaction with the local Yaru. Our database may simply lack genetic samples sufficiently representative of the Yaru population specifically. The wide area of the south-central Andes across which we found haplotypes most shared with the Yanesha did fall under the Incas' brief control (c. 1472–1532), but also that of the Wari Middle Horizon (c. 550–1000), hypothesized as having driven the first main phase of Quechua dispersal, and which fits also with the time scale of up to 50 generations for the haplotypes concerned. And whether local Yaru or from further afield, a plausible scenario of predominantly trader and/or military immigrants to the Yanesha communities would explain the sex-bias in the genetic contributions. Finally, over the past few decades, the construction of new roads to facilitate extraction of natural resources has favored further colonization, bringing new populations from the highlands into the Yanesha regions. The trend towards mating outsider males (exogamy) is increasing, enhancing the admixture effects.

Genetic structure within the South American continent

Within South America, genetic structure broadly follows the two contrasts between the environments of the Andes and Amazonia, and their respective cultural trajectories over many millennia (Tarazona-Santos et al., 2001). In particular, within the Y chromosome native haplogroup Q, the STR haplotypes show very high variance between populations within Amazonia, but greater homogeneity in the Andes, confirming a tendency previously reported (Luiselli et al., 2000; Tarazona-Santos et al., 2001; Fuselli et al., 2003; Barbieri et al., 2011; Bisso-Machado et al., 2012—see Table 2). The NJ tree for the Y chromosome R_{ST} distances (Supporting Information Fig. S3a) appears in line with this trend: to the left, most Andean populations appear on short, close branches, while to the right most Amazonian populations are on long, separate branches). In the paternal line, linguistic affiliation to either the Quechua or the Arawak family does not make for a strong criterion by which to group populations (Table 2). In fact, all highland populations are more homogeneous than the subset of Quechua populations, probably due to some Quechua speakers who do not originate in the core Andean region, such as the Ecuadorian lowland Kichwa or Argentinean Colla (as reported in Roewer et al., 2013), and whose paternal lines are more similar to Amazonian populations. Among Arawak-speaking populations, meanwhile, variance is significantly higher. More problematic is the proximity to some populations of the northeast of the continent, such as the Arawak Palikur and the Kalina, Apalai and Emerillon from the dataset in Mazieres et al. (2008). These are the only Amazonian populations to show proximity to the Andean cluster on the paternal side (Fig. 2a).

The maternal line shows higher homogeneity at sequence level, while the haplogroup distribution displays regional gradients (Supporting Information Fig. S5, see also Bisso-Machado et al., 2012). The main structure on the continent-wide level is the clustering of some Andean populations visible on the nMDS plot, and the separation of northern populations, which appear in a drifted position on the NJ tree. In our dataset the Quechua speakers show higher homogeneity than do Andean populations in general (Table 2). Unlike the Y chromosome pattern, the nine Arawak-speaking populations appear more homogeneous than the Quechua-speaking ones, despite the greater geographic distances; this might be an artifact of the general homogeneous signal from Amazonian populations, while Andean populations are distributed more widely.

The only other Arawak speakers from the Andean area present in our dataset are two Machiguenga populations (Mazières et al., 2008; Sandoval et al., 2013). On this limited basis only, we sought to explore a hypothetical common history for the dispersal of the Arawak family into high altitude areas—although Yanesha and Machiguenga belong to quite different branches of that family. A putative genetic proximity is detectable only in the paternal line: Yanesha and Machiguenga share a high number of haplotypes and show close R_{ST} distances, but this signal might be the result of intense contact with immigrants from the highlands, perhaps particularly in the Inca period. Other factors indicate a different demographic history for the Machiguenga: the low levels of diversity suggest a marked degree of isolation, while the maternal line is closer to that of other Andean populations, in particular from the Titicaca area.

Overall, our results raise questions for both main approaches to accounting for the dispersal of the Arawak language family, whether by demic or cultural spread. Most plausibly, both would have had a role, as this Yanesha case study well illustrates. Linguistic data reveal Yanesha as a member of the Arawak family, but radically transformed by several powerful contact impacts, not least from the contiguous highland Quechua. Our Y chromosome data duly detect the robust input of a component of Andean origin—but into a geographically widespread substrate, going back to a greater time depth. This older signal points towards populations genetically close (within the available published data) to

others far down the Amazon and into the Guianas, matching the second of Walker and Ribeiro's "bi-modal" estimates for the Arawak homeland (2011). The Yanesha case only confirms the need to test the rival theories of Arawak origins on fuller data-sets that target coverage of the hypothesized homelands currently in northern (Aikhenvald, 1999) or western (Walker and Ribeiro, 2011) Amazonia.

ACKNOWLEDGMENTS

This work was supported by the European Research Council ERC-2011-AdG 295733 grant (Langelin) to DP and by a grant from the Faculty of Mathematical, Physical and Natural Sciences of the University of Bologna to CB and DYY. We are grateful to all the Yanesha people who participated in this project, and in particular to Marcial Ortiz Espiritu for fieldwork assistance. We thank Giorgio Piracci for fieldwork planning and Margherita Banella for help in collecting samples. We thank Giorgia Mattei for assistance in the lab analysis. We also thank Prof. Willem Adelaar for helpful insights on Yanesha language and culture.

LITERATURE CITED

Adelaar W. 2006. The Quechua impact in Amuesha. In: Aikhenvald AY, Dixon RMW, editors. Grammars in contact: a cross-linguistic typology. Latrobe, Australia: Research Centre for Linguistic Typology. p 290–311.

Aikhenvald AY. 1999. The Arawak language family. In: Dixon RMW, Aikhenvald AY, editors. The Amazonian languages. Cambridge: Cambridge Language Surveys. p 65–106.

Amorim CEG, Bisso-Machado R, Ramallo V, Bortolini MC, Bonatto SL, Salzano FM, Hünemeier T. 2013. A Bayesian approach to genome/linguistic relationships in native south americans. PLoS One 8:e64099.

Bailliet G, Ramallo V, Muzzio M, García A, Santos MR, Alfaro EL, Dipierri JE, Salceda S, Carnese FR, Bravi CM, Bianchi NO, Demarchi DA. 2009. Brief communication: restricted geographic distribution for Y-Q* paragroup in South America. Am J Phys Anthropol 140:578–582.

Barbieri C, Heggarty P, Castrì L, Luiselli D, Pettener D. 2011. Mitochondrial DNA variability in the Titicaca basin: matches and mismatches with linguistics and ethnohistory. Am J Hum Biol 23:89–99.

Battaglia V, Grugni V, Perego UA, Angerhofer N, Gomez-Palmieri JE, Woodward SR, Achilli A, Myres N, Torroni A, Semino O. 2013. The first peopling of South America: new evidence from Y-chromosome haplogroup Q. PLoS One 8: e71390.

Becker RA, Wilks AR, Brownrigg R, Minka TP. 2013. maps: Draw geographical maps, 2013. R package version, 2–3.

Behar DM, Van Oven M, Rosset S, Metspalu M, Loogväli EL, Silva NM, Kivisild T, Torroni A, Villems R. 2012. A "copernican" reassessment of the human mitochondrial DNA tree from its root. Am J Hum Genet 90:675–684.

Beresford-Jones D, Heggarty P. 2013. Andes: archaeology. In: Ness I, Bellwood P, editors. The encyclopedia of global human migration. Oxford: Blackwell Publishing. p 410–416.

Bisso-Machado R, Bortolini MC, Salzano FM. 2012. Uniparental genetic markers in South Amerindians. Genet Mol Biol 35: 365–387.

Bortolini M-C, Salzano FM, Thomas MG, Stuart S, Nasanen SPK, Bau CHD, Hutz MH, Layrisse Z, Petzl-Erler ML, Tsuneto LT, Hill K, Hurtado AM, Castro-de-Guerra D, Torres MM, Groot H, Michalski R, Nymadawa P, Bedoya G, BradmanN, Labuda D, Ruiz-Linares A. 2003. Y-chromosome evidence for differing ancient demographic histories in the Americas. Am J Hum Genet 73:524–539.

Brion M, Sobrino B, Blanco-Verea A, Lareu MV, Carracedo A. 2005. Hierarchical analysis of 30 Y-chromosome SNPs in European populations. Int J Legal Med 119:10–15.

Callegari-Jacques SM, Tarazona-Santos EM, Gilman RH, Herrera P, Cabrera L, dos Santos SEB, Morés L, Hutz MH, Salzano FM. 2011. Autosome STRs in native South America—testing models of association with geography and language. Am J Phys Anthropol 145:371–381.

Campbell L. 1997. American indian languages. Oxford: Oxford University Press.

Contreras W. 2007. Afiche conmemorativo del Sesquicentenario de la llegada de los Colonos Autro-Alemanes a Pozuzo-Perú (1857–1859). Asociación de Historia y Cultura de Pozuzo, Club Cultural de Prusia y Museo Schafferer. ESERGRAF Ediciones y servicios graficos, RUC.

Daigneault AL. 2010. An ethnolinguistic study of the Yanesha' (Amuesha) language and speech community in Peru's Andean Amazon, and the traditional role of Ponapnora, a female rite of passage. PhD dissertation, Department of Anthropology, Université de Montréal.

Dillehay TD, Ramírez C, Pino M, Collins MB, Rossen J, Pino-Navarro JD. 2008. Monte Verde: Seaweed, Food, Medicine, and the Peopling of South America. Science 320:784–786.

Fagundes NJR, Kanitz R, Bonatto SL. 2008. A reevaluation of the Native American mtDNA genome diversity and its bearing on the models of early colonization of Beringia. PLoS One 3:e3157.

De Filippo C, Barbieri C, Whitten M, Mpoloka SW, Gunnarsdóttir ED, Bostoen K, Nyambe T, Beyer K, Schreiber H, de Knijff P, Luiselli D, Stoneking M, Pakendorf B. 2011. Y-chromosomal variation in sub-Saharan Africa: insights into the history of Niger-Congo groups. Mol Biol Evol 28:1255–1269.

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. Mol Biol Evol 20:1682–1691.

Hornborg A, Gassn R, Heckenberger M, Hill J, Neves E, SantosGranero F. 2005. Ethnogenesis, regional integration, and ecology in prehistoric Amazonia: toward a system perspective 1. Curr Anthropol 46:589–620.

INEI. Censos Nacionales 2007: XI de Población y VI de Vivienda. Crecimiento y Distribución de la población. Lima, Peru.

Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF. 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. Genome Res 18:830–838.

Kitchen A, Miyamoto MM, Mulligan CJ. 2008. A three-stage colonization model for the peopling of the Americas. PLoS One 3:e1596.

Lewis CM, Lizárraga B, Tito RY, López PW, Iannacone GC, Medina A, Martínez R, Polo SI, De la Cruz AF, Cáceres AM, Stone AM. 2007. Mitochondrial DNA and the peopling of South America. Hum Biol 79:159–178.

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

Lumbreras LG. 1974. The peoples and cultures of Ancient Peru. Washington, D.C.: Smithsonian Institution Press.

Mazières S, Guitard E, Crubézy E, Dugoujon JM, Bortolini MC, Bonatto SL, Hutz MH, Bois E, Tiouka F, Larrouy G, Salzano FM. 2008. Uniparental (mtDNA, Y-chromosome) polymorphisms in French Guiana and two related populations—implications for the region's colonization. Ann Hum Genet 72:145–156.

Miller SA, Dykes DD, Polesky HF. 1988. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 16:1215.

Nenadic O, Greenacre M. 2007. Correspondence analysis in R, with two-and three-dimensional graphics: the ca package. J Stat Softw 20:1–13.

Nettle D. 1999. Linguistic diversity of the Americas can be reconciled with a recent colonization. Proc Natl Acad Sci 96: 3325–3329.

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MRH, Wagner H. 2012. vegan: Community Ecology Package. R package version 2.0–5. http://www.cran.r-project.org/web/packages/vegan/index.html.

Paradis E. 2010. Pegas: an R package for population genetics with an integrated–modular approach. Bioinformatics 26:419.

Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20:289–290.

Ramallo V, Bisso-Machado R, Bravi C, Coble MD, Salzano FM, Hünemeier T, Bortolini MC. 2013. Demographic expansions in South America: enlightening a complex scenario with genetic and linguistic data. Am J Phys Anthropol 150:453–463.

Roewer L, Nothnagel M, Gusmão L, Gomes V, González M, Corach D, Sala A, Alechine E, Palha T, Santos N, Geppert M, Willuweit S, Nagy M, Zweynert S, Baeta M, Núñez C, Martínez-Jarreta B, González-Andrade F, Fagundes de Carvalho E, da Silva DA, Builes JJ, Turbón D, Lopez Parra AM, Arroyo-Pardo E, Toscanini U, Borjas L, Barletta C, Ewart E, Santos S, Krawczak M. 2013. Continent-wide decoupling of Y-chromosomal genetic variation from language and geography in native South Americans. PLoS Genet 9: e1003460.

Rothhammer F, Dillehay TD. 2009. The late Pleistocene colonization of South America: an interdisciplinary perspective. Ann Hum Genet 73:540–549.

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

Sandoval JR, Lacerda DR, Jota MSA, Salazar-Granara A, Vieira PPR, Acosta O, Cuellar C, Revollo S, Fujita R, Santos FR. The Genographic Project Consortium. 2013. The genetic history of indigenous populations of the Peruvian and Bolivian Altiplano: the legacy of the Uros. PLoS One 8:e73006.

Schurr TG, Ballinger SW, Gan YY, Hodge JA, Merriwether DA, Lawrence DN, Knowler WC, Weiss KM, Wallace DC. 1990. Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. Am J Hum Genet 46:613–623.

Schurr TG, Sherry ST. 2004. Mitochondrial DNA and Y chromosome diversity and the peopling of the Americas: evolutionary and demographic evidence. Am J Hum Biol 16:420–439.

Semino O, Magri C, Benuzzi G, Lin AA, Al-Zahery N, Battaglia V, Maccioni L, Triantaphyllidis C, Shen P, Oefner PJ, Zhivotovsky LA, King R, Torroni A, Cavalli-Sforza LL, Underhill PA, Santachiara-Benerecetti AS. 2004. Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. Am J Hum Genet 74:1023–1034.

Sevini F, Yao DY, Lomartire L, Barbieri A, Vianello D, Ferri G, Moretti E, Dasso MC, Garagnani P, Pettener D, Franceschi C, Luiselli D, Franceschi ZA. 2013. Analysis of population substructure in two sympatric populations of Gran Chaco, Argentina. PLoS One 8:e64054.

Tamm E, Kivisild T, Reidla M, Metspalu M, Smith DG, Mulligan CJ, Bravi CM, Rickards O, Martinez-Labarga C, Khusnutdinova EK, Fedorova SA, Golubenko MV, Stepanov VA, Gubina MA, Zhadanov SI, Ossipova LP, Damba L, Voevoda MI, Dipierri JE, Villems R, Malhi RS. 2007. Beringian standstill and spread of Native American founders. PLoS One 2:e829.

Tarazona-Santos E, Carvalho-Silva DR, Pettener D, Luiselli D, De Stefano GF, Labarga CM, Rickards O, Tyler-Smith C, Pena SDJ, Santos FR. 2001. Genetic differentiation in South Amerindians is related to environmental and cultural diversity: evidence from the Y chromosome. Am J Hum Genet 68: 1485–1496.

Torroni A, Schurr TG, Cabell MF, Brown MD, Neel J V, Larsen M, Smith DG, Vullo CM, Wallace DC. 1993. Asian affinities and continental radiation of the four founding Native American mtDNAs. Am J Hum Genet 53:563–590.

Torroni A, Schurr TG, Yang CC, Szathmary EJ, Williams RC, Schanfield MS, Troup GA, Knowler WC, Lawrence DN, Weiss KM. 1992. Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. Genetics 130:153–162.

van Oven M, Kayser M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mutat 30:386–394.

Venables WN, Ripley BD. 2002. MASS: modern applied statistics with S. New York: Springer.

Walker RS, Ribeiro LA. 2011. Bayesian phylogeography of the Arawak expansion in lowland South America. Proc R Soc B Biol Sci 278:2562–2567.

Wallace DC, Garrison K, Knowler WC. 1985. Dramatic founder effects in Amerindian mitochondrial DNAs. Am J Phys Anthropol 68:149–155.

Wang S, Lewis Jr. CM, Jakobsson M, Ramachandran S, Ray N, Bedoya G, Rojas W, Parra MV, Molina JA, Gallo C, Mazzotti G, Poletti G, Hill K, Hurtado AM, Labuda D, Klitz W, Barrantes R, Bortolini MC, Salzano FM, Petzl-Erler ML, Tsuneto LT, Llop E, Rothhammer F, Excoffier L, Feldman MW, Rosenberg NA, Ruiz-Linares A. 2007. Genetic variation and population structure in Native Americans. PLoS Genet 3: e185.

Yang NN, Mazières S, Bravi C, Ray N, Wang S, Burley M-W, Bedoya G, Rojas W, Parra MV, Molina JA, Gallo C, Poletti G, Hill K, Hurtado AM, Petzl-Erler, ML, Tsuneto LT, Klitz W, Barrantes R, Llop E, Rothhammer F, Labuda D, Salzano FM, Bortolini MC, Excoffier L, Dugoujon JM, Ruiz-Linares A. 2010. Contrasting patterns of nuclear and mtDNA diversity in Native American populations. Ann Hum Genet 74:525–538.