

Demographic Expansions in South America: Enlightening a Complex Scenario With Genetic and Linguistic Data

Virginia Ramallo,¹ Rafael Bisso-Machado,¹ Claudio Bravi,² Michael D. Coble,³ Francisco M. Salzano,¹ Tábita Hünemeier,¹ and Maria Cátira Bortolini^{1*}

¹*Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Caixa Postal 15053, 91501-970, Porto Alegre, RS, Brazil*

²*Laboratorio de Genética Molecular Poblacional, Instituto Multidisciplinario de Biología Celular (CCT La Plata CONICET-CICPBA), Argentina*

³*Armed Forces DNA Identification Laboratory, 1413 Research Blvd, Rockville, MD*

KEY WORDS Native Americans; genetic–linguistic comparisons; HVS I–II mtDNA sequences; fission–fusion process

ABSTRACT Native Americans are characterized by specific and unique patterns of genetic and cultural/linguistic diversities, and this information has been used to understand patterns of geographic dispersion, and the relationship between these peoples. Particularly interesting are the Tupi and Je speaker dispersions. At present, a large number of individuals speak languages of these two stocks; for instance, Tupi-Guarani is one of the official languages in Paraguay, Bolivia, and the Mercosul economic block. Although the Tupi expansion can be compared in importance to the Bantu migration in Africa, little is known about this event relative to others. Equal and even deeper gaps exist concerning the Je-speakers' expansion. This study aims to elucidate some aspects of these successful expansions. To meet this purpose, we analyzed Native American mtDNA complete control region from nine dif-

ferent populations and included HVS-I sequences available in the literature, resulting in a total of 1,176 samples investigated. Evolutionary relationships were explored through median-joining networks and genetic/geographic/linguistic correlations with Mantel tests and spatial autocorrelation analyses. Both Tupi and Je showed general traces of ancient or more recent fission–fusion processes, but a very different pattern of demographic expansion. Tupi populations displayed a classical isolation-by-distance pattern, while Je groups presented an intricate and nonlinear mode of dispersion. We suggest that the collective memory and other cultural processes could be important factors influencing the fission–fusion events, which likely contributed to the genetic structure, evolution, and dispersion of Native American populations. *Am J Phys Anthropol* 150:453–463, 2013. ©2013 Wiley Periodicals, Inc.

Native American populations have been widely studied for relationships between genetic, linguistic, and geographic data. Some of these studies showed a poor correlation between genetic and geographic distances, suggesting strong local differentiation (O'Rourke and Suarez, 1986; Santos et al., 2009). However, other investigators have noted specific genetic patterns when groups of different language stocks were compared (Salzano et al., 1988; Callegari-Jacques and Salzano, 1989; Kemp et al., 2010; Callegari-Jacques et al., 2011).

The Native Brazilian linguistic landscape includes between 154 and 170 languages divided in ~20 major stocks (Rodrigues, 2002; Moore, 2005). This enormous diversity has led to controversies regarding their relationship and hierarchical classification (Campbell, 1997; Bolnick et al., 2004), but two major linguistic stocks are normally recognized by all authors: Tupi and Je/Ge (Loukotka, 1968; Rodrigues, 1986; Greenberg, 1987; Campbell, 1997; Lewis, 2009).

The Macro-Tupi family (henceforth named Tupi) is characterized by a large number of speakers who inhabit a wide geographic area in South America with languages associated to its major branch, Tupi-Guarani (or just Guarani). The Guarani language currently presents a large geographical spread, with speakers in Brazil, Argentina, Paraguay, Bolivia, Peru, Venezuela, and French Guiana (Rodrigues, 2002). It is noteworthy that the Guarani speakers reached the southernmost limits of the

Tupi dispersion. The origin and expansion of the Tupi linguistic family are still controversial, but there is a consensus on two major points: (1) a common Amazonian

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

Grant sponsors: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (Programa de Apoio a Núcleos de Excelência).

Virginia Ramallo, Rafael Bisso-Machado, Tábita Hünemeier, and Maria Cátira Bortolini contributed equally to this work.

Present address: Michael D. Coble: National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899-8214

*Correspondence to: Maria Cátira Bortolini, Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Caixa Postal 15053, 91501-970 Porto Alegre, RS, Brazil. E-mail: maria.bortolini@ufrgs.br

Received 9 October 2012; accepted 5 December 2012

DOI 10.1002/ajpa.22219

Published online 22 January 2013 in Wiley Online Library (wileyonlinelibrary.com).

center of origin and (2) the Tupi differentiation occurred through distinct historic and cultural processes; however, they still retained several common features (Noelli, 1998, 2008; Rodrigues, 2000). Glottochronology information suggested that the Tupi origin occurred at the southern margin of the Amazon River mostly close to the Madeira-Guaporé basin, while archaeological data (based on site location, radiocarbon and thermoluminescence dating) point to a region farther north, at the confluence of the Madeira and Amazon rivers (Rodrigues, 1964; Migliazza, 1982; Urban, 1996, 1998; Heckenberger et al., 1998; Noelli, 2008). The morphological results also indicate an Amazonian origin for the Tupi group (Neves et al., 2011).

Regardless of the uncertainties, the expansion of the Tupi speakers, which occurred about 6,000–2,000 years before present (YBP), has been considered as one of the largest and most successful migration identified at the continental level (Rodrigues, 1964; Migliazza, 1982; Urban, 1996, 1998; Heckenberger et al., 1998; Noelli, 2008). These people were known for manufacturing ceramic utensils and have been described as skilled agriculturalists. Tupi speakers efficiently transformed the forest into productive areas, transporting their flora and also adopting new vegetables (Noelli, 2008). Territories under the domain of some Tupi peoples were conquered and managed for a long time. Due to these characteristics, the better term for these population shifts is expansion, meaning distention and spreading, conquering of new regions without abandoning the previous region (Noelli, 2008). Why the Tupi people left the border of the Amazon, however, may have been for multiple (and not mutually exclusive) reasons: demographic and/or social-cultural pressures, search for new cultivable lands, and/or prolonged droughts (Schmitz, 1997).

Linguistic and archaeological data have indicated that the Tupi dispersion was rapid and radial. This particular migration pattern would trigger an interesting phenomenon: languages spoken by geographically distant peoples could be closely related (Migliazza, 1982; Urban, 1996; Rodrigues, 2002). Figure 1 shows the putative routes of the Tupi expansion.

The postulated Je linguistic family origin, conversely, is believed to have occurred in an area between the São Francisco and Tocantins rivers and may have originated at about the same time as the Tupi. The relationship between these two major linguistic families, however, is subject to debate. Some researchers have suggested a minimum chronological depth of 7,000–5,000 YBP. The majority of the Je-speaking communities are located in the central and eastern regions of the Brazilian plateau, and an important movement to the south occurred approximately 3,000 YBP (Urban, 1998). However, the Je dispersion seems to be distinct from the pattern followed by the Tupi.

Only recently, genetic data have been used to delineate a more comprehensive view of the spread dynamics of the Tupi and Je people in South America. The conclusions of these studies are somewhat limited, given the number of populations investigated and the relatively few geographic areas covered (Marrero et al., 2007; Mazières et al., 2008).

The present investigation was performed to test if the expansion model proposed by linguists, archeologists, and morphologists agrees with the variation found in the complete control region of the mitochondrial DNA found in Tupi and Je speakers from nine different populations

located in their probable route of dispersion within the current Brazilian territory. Some of these populations have never been studied with respect to mitochondrial DNA (mtDNA) variation. These new mtDNA sequences were analyzed together with those from eight previously investigated populations affiliated with Tupi, Je, and other major South American linguistic families. For a more comprehensive understanding of this complex process, our work also takes into account specific aspects of the dynamics of each society. For example, collective memory understood as the common experience of sharing particular memories and registers of certain past events (Halbwachs, 1992), can be a determining factor in the behavior of members of a community. The collective memory creates a particular background of social identity. As a result, every group has its own collective memory and differs from the collective memory of other groups (Russell, 2006). Certain social practices (marriages) or demographic events (e.g., population fusion and fissions) can be strongly influenced by these different constructions of a group's past.

Our results give indications that the Tupi and Je speakers underwent radically different patterns of dispersal, therefore, unraveling part of the complex history of the South American prehistoric colonization.

MATERIALS AND METHODS

DNA samples

Ethical approval for this study was provided by the Brazilian National Ethics Commission (CONEP Resolution no. 123/98). Individual and tribal informed oral consent was obtained from all participants because in many cases these populations have little to no tradition of writing. Samples were obtained according to the Helsinki Declaration. The ethics committee approved the oral consent procedures, as well as the use of these samples in population and evolutionary studies.

DNA from 130 individuals (men and women) from nine Native American groups (Tupi: Cinta-Larga, Mundurukú, Parakanã; Je: Gorotire, Kuben-Kran-Kegn, Mekranoti, Xikrin [subgroups of the Kayapó complex; see item below "Evolutionary relationships"]; Carib: Arara; Isolated: Jamamadi) were amplified for the complete control region of mtDNA (rCRS positions 16,024–16,569 plus 001–576) according to the methods presented by Irwin et al. (2007). In addition, a portion of the mtDNA sequence located immediately at 5' of the control region (rCRS positions 15,878–16,023), which covers the position 15,930, was also sequenced to differentiate C1 Native American/Asian subhaplogroups (C1b, C1c, and C1d). Amplification products were purified and sequenced using the Applied Biosystems 3130 or 3730 Genetic Analyzer sequencers. Accession numbers for data presented here and reported to GenBank, <http://www.ncbi.nlm.nih.gov/Genbank/index.html>, are JQ995985–JQ996143.

Sequences were aligned and their quality, as well as accuracy, was evaluated using the Phred, Phrap, Coned (<http://www.phrap.org/phredphrapconsed.html>) software, and the Clustal W algorithm included in the BioEdit 7.0.9 software (Hall, 1999; <http://www.mbio.ncsu.edu/bioedit/bioedit.html>). Haplogroup frequencies were obtained through direct counting. Figure 2 gives the geographical location of the tested populations and of those studied by other authors and included for comparison.

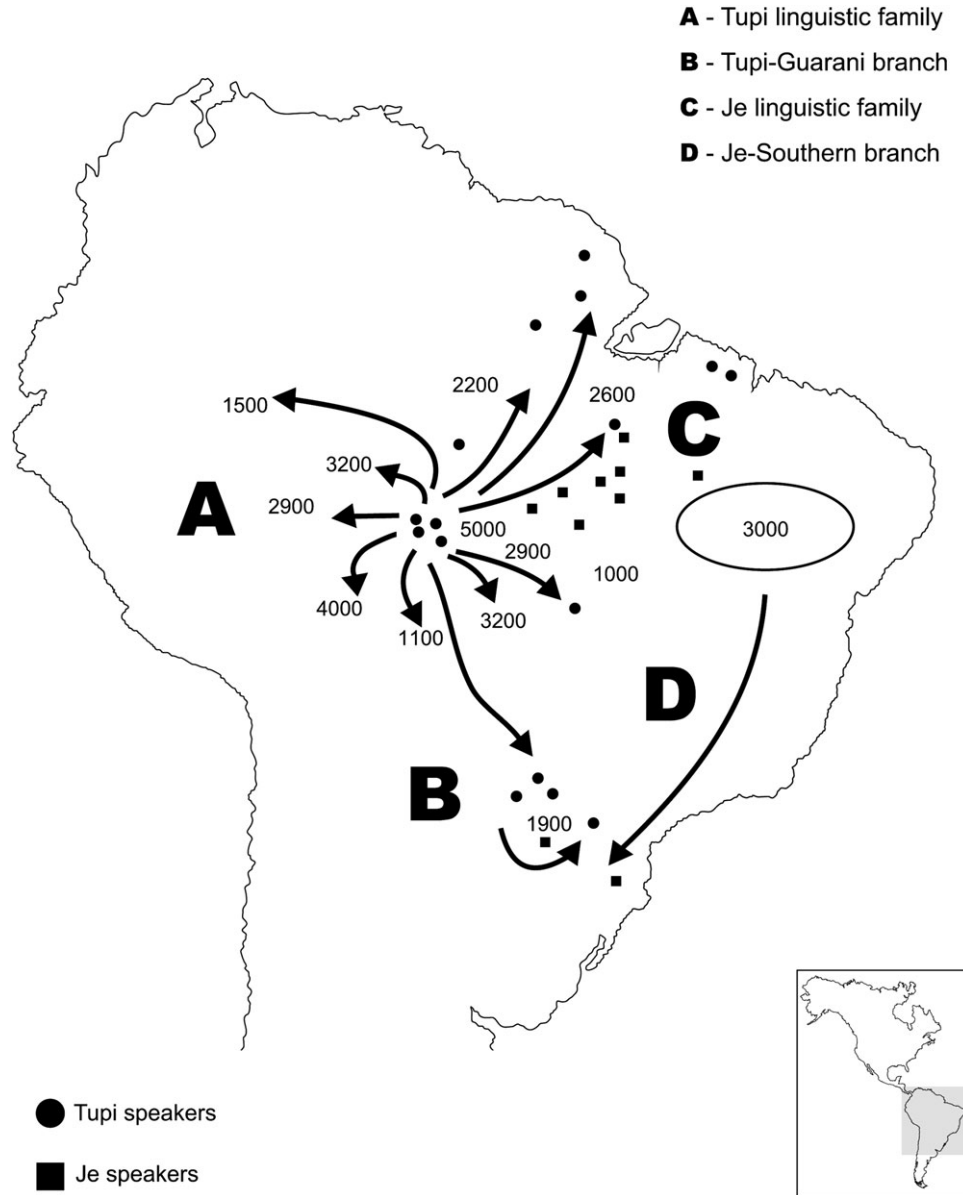


Fig. 1. Probable routes of dispersion of the Tupi and Je speakers (adapted from Schmitz et al., 1997) examined in this study. Arrows show possible routes together with estimated dates when migrations may have occurred in years before present.

Data analysis

Analyses were performed using the new mtDNA sequences obtained in the present study, as well as all those available in the literature. A total of 1,177 Native Americans (Tupi, Je, Carib, Arawakan, Yanoman, Arauan, and Ticuna speakers, see Fig. 2) from 37 different populations were considered. For this global analysis, only HVS-I sequences (rCRS positions 16,024–16,569) were used because of the lack of HVS-II (16,024–16,569) data in several of the published Amerindian populations. It should be noted that some of the analyses considered only the Tupi and Je linguistic stocks, because these were the main focus of the present investigation.

Genetic diversity

Sequences were aligned and processed using the MEGA 5 (Tamura et al., 2011) and DnaSP 4.50.1 (Librado and

Rozas, 2009) software. Native American mitochondrial haplogroups were defined considering the HVS-I polymorphic sites (plus HVS-II and position 15,930 when available) according to Achilli et al. (2008). The summary statistics (number of sequences, haplotype number, and haplotype diversity; Nei, 1987) were calculated using DnaSP 4.50.1 (Librado and Rozas, 2009). Population structure was estimated by an analysis of molecular variance (AMOVA; Weir and Cockerham, 1984; Excoffier et al., 1992; Weir, 1996) using Arlequin 3.5.1.2 (Excoffier and Lischer, 2010).

Evolutionary relationships

Median-joining haplotype networks (Bandelt et al., 1999) were constructed using the Network software (version 4.6.1.0; <http://www.fluxus-engineering.com>) separately for haplogroups A2, B2, C1, and D1 for the Tupi

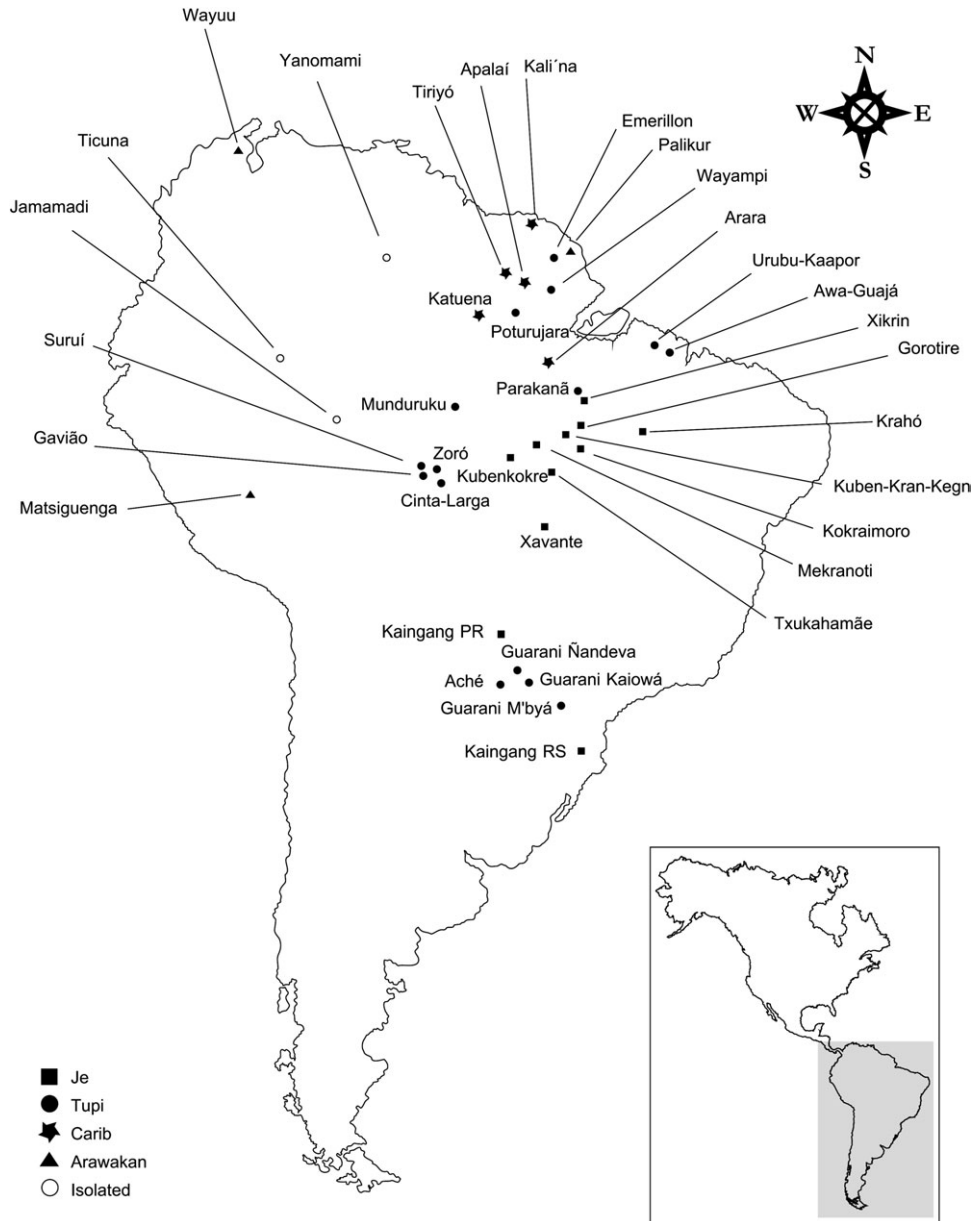


Fig. 2. Map of South America, showing the populations included in this study.

and Je stocks. Possible reticulations were resolved by MP calculation (Polzin and Daneschmand, 2003). To facilitate the graphs' interpretation, Tupi and Je populations were organized according to their relative geographic location: Northern Tupi (Mundurukú, Poturujara, Waiampi, Emerillon, Parakanã), Northeastern Tupi (Awa-Guajá, Urubu Kaapor), Central Tupi (Zoró, Cinta Larga, Surui, Gavião), and Southern Tupi (Ache, Guarani M'bya, Guarani Kaiowá, Guarani Nandeva); as well as Northern Je (Krahó and seven subgroups of the Kayapó complex, see below), Central Je (Xavante), and Southern Je (Kaingang).

To examine the evolutionary and demographic histories of the Kayapó complex (Northern Je branch), additional networks were constructed using control region sequences, extended to include the 15,930 sites. Data from seven subgroups (Gorotire, Kokraimoro, Kubenkokre, Kuben-Kran-Kegn, Mekranoti, Txukahamãe, and Xikrin) were

considered in the analyses. The Kayapó split from other members of the Je family probably occurred in the past 500 years (Urban, 1998). Successive events led to the formation of the aforementioned subgroups. This specific and relatively well-documented trajectory indicates that the Kayapó could be regarded as ideal for testing recent fission–fusion events, as well as complex nonrandom population processes highly dependent on cultural factors, which likely contributed to the genetic structure, evolution, and dispersion of Native American populations and of other prestate human groups (Neel and Salzano, 1967; Hunley et al., 2007; and references therein).

Genetic versus geographic/ historical-linguistic data

For both Tupi and Je samples, Mantel tests (Mantel, 1967) were performed using Arlequin 3.5.1.2 (Excoffier

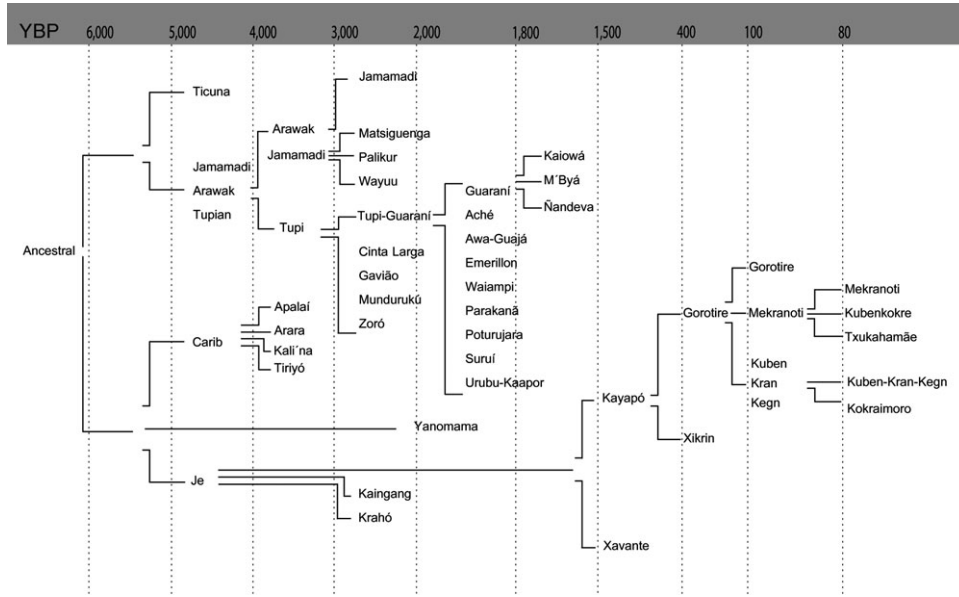


Fig. 3. Expansion and subdivision of the different groups based on time estimates of language or village splits. Based on Ethnologue (www.ethnologue.com), Carneiro-da-Cunha (1998 and references therein), Nichols (2008) and Salzano (1971) (YBP: years before present).

and Lischer, 2010) with 100,000 permutations to test for correlations between genetic, geographic, and historical distances at the 0.05 level of probability. Genetic distances were estimated as Kimura 2P F_{ST} values (Kimura, 1980). Geographic distances were calculated using latitude and longitude coordinates (<http://www.movable-type.co.uk/scripts/latlong.html>). Historical distances were estimated based on time estimates of language or village splits taken from Ethnologue (Lewis, 2009), Carneiro da Cunha (1998, and references therein), Nichols (2008), and Salzano (1971) (see Fig. 3). Note, however, that these dates are only estimates, so they should be considered with caution.

Spatial autocorrelation

Spatial autocorrelation analyses provide a measure of the genetic similarity of individuals as a function of the geographic distance separating them. Generally, patterns revealed by spatial autocorrelation are also reflected in the Mantel statistics. In fact, Mantel tests can be used to test spatial correlations in distance classes (Epperson, 2003). To quantitatively examine this geographic differentiation, we used the Genalex 6.4 software (Peakall and Smouse, 2006). The r statistic (which measures the genetic similarity between pairs of individuals within a given distance) was calculated for 10 classes and the results summarized in a correlogram. For large samples, r values range from -1 (negative autocorrelation, indicating genetic dissimilarity in a distance class) to $+1$ (positive autocorrelation or genetic similarity). Under the null hypothesis, the mtDNA haplotypes are randomly distributed in space. The distance class size at which the statistic r is no longer significant provides an approximation of the extent of detectable spatial genetic structure (Peakall et al., 2003). Statistical significance was based on 95% confidence intervals. P -values were obtained via randomization methods. For all analyses, we used 10,000 permutations.

RESULTS

Genetic diversity

A total of 177 different haplotypes were identified among the 1,177 HVS-I mtDNA sequences analyzed, which were approximately equally distributed among the four Native American haplogroups A2, B2, C1, and D1 (Table 1 and Supporting Information Table 1). Approximately 41% of the haplotypes are unique to individuals, while 40% were shared among individuals within a population and 7% were found between populations of the same linguistic group, while 12% were observed among linguistic stocks.

Overall, 81% of the haplotypes were population-specific and 19% were shared between populations (Supporting Information Table 1). Jamamadi, Ticuna, and Yanomami only shared the nodal haplotype for B2, C1, and D1 haplogroups with other populations and were absent for haplogroup A2.

Molecular variance (AMOVA) analysis revealed that the percentage of variation among all samples was $\sim 35\%$ ($F_{ST} = 0.354, P = 0.000$). F_{ST} values for each language group were also calculated (Table 2), Tupi sequences showed the highest values: $\sim 39\%$ of the observed variation was due to differences between populations of the same linguistic stock. Je and Carib sequences showed lower values (22% and 19%, respectively).

The total number of HVS-I mtDNA sequences for each language group (N), the number of haplotypes (H), and the haplotype diversity (Hd) are presented in Table 2. Je sequences showed 37 different haplotypes and the lower Hd value (0.892).

Evolutionary relationships

The Tupi A2, B2, and C1 networks show a star-like form (Supporting Information Fig. 1) with the exception of haplogroup D1, which has a major division caused by a change at position 16,179. Interestingly, this mutation is present

TABLE 1. Haplogroup population frequencies and references from which the primary data were obtained

Population	State/Country	N	A2	B2	C1	D1	Partial samples sizes and references
Tupi language family							
Ache	Paraguay	74	6	67	1	–	63 (Schmitt et al., 2004) + 11 (Yang et al., 2010)
Awa Guajá	Maranhão, Tocantins, Pará, BR	2	–	2	–	–	2 (Santos et al. 1996)*
Cinta-Larga	Mato Grosso, BR	11	9	–	1	1	8 (present study) + 3 (Dornelles et al., 2005)
Emerillon	French Guiana	30	9	21	–	–	29 (Mazières et al., 2008)
Gavião	Rondônia, BR	27	4	4	–	19	27 (Ward et al., 1996)
Guarani Kaiowá	Mato Grosso do Sul, BR	120	110	–	10	–	120 (Marrero et al., 2007)
Guarani M'byá	Paraná, BR	24	12	–	–	12	24 (Marrero et al., 2007)
Guarani Nandeva	Mato Grosso do Sul, BR	56	46	–	10	–	56 (Marrero et al., 2007)
Munduruku	Pará, Amazonas, BR	16	2	4	2	8	14 (present study) + 2 (Lobato-da-Silva et al., 2001)
Parakanã	Pará, BR	22	1	9	8	4	19 (present study) + 3 (Dornelles et al., 2005)
Poturuajara	Pará, BR	13	6	–	4	3	4 (Lobato-da-Silva et al., 2001) + 9 (Santos et al., 1996)*
Suruí	Rondônia, BR	24	4	–	–	20	24 (Bonatto and Salzano, 1997)
Zoró	Rondônia, BR	30	6	2	4	18	30 (Ward et al., 1996)
Urubu Kaapor	Maranhão, BR	3	1	–	–	2	3 (Dornelles et al., 2005)
Wayampi	Amapá, Pará, BR	71	41	10	–	20	3 (Lobato-da-Silva et al., 2001) + 21 (Santos et al., 1996)* + 47 (Mazières et al., 2008)
Je language family							
Gorotire	Mato Grosso, Pará, BR	11	3	2	2	4	11 (present study)
Kaingang PR	Paraná, BR	21	13	–	8	–	21 (Marrero et al., 2007)
Kaingang RS	Rio Grande do Sul, BR	53	22	3	28	–	53 (Marrero et al., 2007)
Kokraimoro	Mato Grosso, Pará, BR	2	1	1	–	–	2 (Lobato-da-Silva et al., 2001)
Krahó	Maranhão, Pará, Tocantins, BR	9	4	4	1	–	9 (Lobato-da-Silva et al., 2001)
Kubekokre	Mato Grosso, Pará, BR	4	–	4	–	–	4 (Lobato-da-Silva et al., 2001)
Kuben-Kran-Kegn	Mato Grosso, Pará, BR	19	11	5	1	2	19 (present study)
Mekranoti	Mato Grosso, Pará, BR	25	9	14	2	–	24 (present study) + 1 (Lobato-da-Silva et al., 2001)
Txukahamãe	Mato Grosso, Pará, BR	2	2	–	–	–	2 (Dornelles et al., 2005)
Xavante	Mato Grosso, BR	25	4	21	–	–	25 (Ward et al., 1996)
Xikrin	Mato Grosso, Pará, BR	4	1	1	1	1	3 (present study) + 1 (Dornelles et al., 2005)
Carib language family							
Apalaí	Pará, BR	102	40	1	30	31	102 (Mazières et al., 2008)
Arara	Pará, BR	29	1	14	14	–	20 (present study) + 9 (Santos et al., 1996)*
Kali'na	Venezuela, French Guiana, Guyana, Suriname, BR	27	2	12	11	2	27 (Mazières et al., 2008)
Tyrió	Pará, BR	3	–	–	1	2	3 (Santos et al., 1996)*
Katuenta	Amazonas, Pará, Roraima, BR; Guyana	9	3	1	3	2	9 (Santos et al., 1996)*
Arawakan language family							
Matsinguenga	Peru	44	2	41	–	1	37 (Mazières et al., 2008)
Palikur	Amapá, Guiana Francesa, BR	46	–	25	–	21	46 (Mazières et al., 2008)
Wayuu	Colombia	18	4	4	10	–	18 (Yang et al., 2010)
Arauan language family							
Jamamadi	Amazonas, BR	13	–	–	9	4	13 (present study)
Yanomam language family							
Yanomami	Roraima, Amazonas, BR	175	2	11	97	65	47 (Santos et al., 1996)* + 128 (Merriwether et al., 2000)
Ticuna language family							
Ticuna	Amazonas, Colombia, Peru, BR	12	5	1	–	6	12 (Yang et al., 2010)

BR: Brazil

* The information on the population corresponding for each sequence was obtained directly from the author.

only in north Tupi populations. In general, all haplotypes were present in the four regions, 32% being unique to individuals, 48% shared within populations, and 20% shared among populations, with 80% of population-specific haplotypes. The unique transversion observed in Tupi individu-

als (16,239A), associated to the 16,266T transition, defines a derived A2aa lineage (van Oven and Kayser, 2009). Three different A2aa haplotypes are represented by six individuals belonging to Poturuajara (North Tupi) and fourteen Guarani-Nandeva (South Tupi) individuals (Marrero

TABLE 2. Intrapopulation diversity parameters in the language groups analyzed

	<i>N</i>	<i>H</i>	<i>H_d</i>	<i>F_{ST}</i>
Arawakan	108	28	0.947	0.306
Carib	170	36	0.931	0.189
Isolated	200	46	0.935	0.248
Je	175	37	0.892	0.236
Tupi	524	60	0.932	0.389

N, sample size ; *H*, different haplotypes; *H_d*, haplotype diversity; *F_{ST}* values

et al., 2007). Lineage A2ab (defined by transition 16,291T; Behar et al., 2012) is also found in Tupi individuals (Central: Cinta-Larga, Gavião, and Zoró; South: Guarani-Ñandeva, Guarani-Kaiowá, and Guarani-M'byá). Among the 40 individuals used to define the C1 haplogroup network, none coincided with the nodal haplotype.

Overall, Je populations presented 42% unique haplotypes, 30% shared within populations, and 28% shared among populations, indicating 72% of population-specific haplotypes. Network haplogroups/haplotypes (Supporting Information Fig. 2) presented a main division in haplogroup C1 generated by a mutation at position 16,051, which is shared by 24 individuals of the Kaingang population and two individuals from the Mekranoti, a subgroup of the Kayapó complex. Haplogroup D1 is the least represented, with only three haplotypes present in northern Je populations.

Fifty-seven complete control region sequences were investigated in four Kayapó subgroups: Gorotire (*n* = 11), Kuben-Kran-Kegn (*n* = 18), Mekranoti (*n* = 25), and Xikrin (*n* = 3; Supporting Information Table 2). Additionally, HVS-I sequences obtained from the literature were used including other three Kayapó subgroups: Kokraimoro (*n* = 2), Kubenkokre (*n* = 4), and Txukahamãe (*n* = 2). Initially, networks were constructed using only the mtDNA complete control region sequences. Supporting Information Figure 3 shows few haplotypes shared between derived tribes and ancestral groups. This pattern was also observed when the HVS-I sequences of the literature were included in the analyses (Supporting Information Fig. 4). Interestingly, ~60% of the haplotypes were tribe-specific (three haplotypes in A2; five in B2; four in C1; and one in D1 haplogroups), which is lower than the value observed among the Tupi (80%), suggesting that the fission model of village propagation is prominent and that the proportion of exclusive haplotypes increases with the depth of the branches. However, the high proportion of exclusive haplotypes in the Kayapó villages is rather unexpected when we consider that the fission between some of these villages occurred just 100 years ago.

The overall frequency of C1 is low among all populations, while D1 is common among the Tupi but rare among the Je speakers. Geographically, the major A2 distributions in both linguistic stocks were concentrated in only a few populations, while B2 was highly represented in the Je, but not in the Tupi samples, with an exception: the Ache (two distinct haplotypes present in 90% of the individuals). Interestingly, the origin of these hunter-gatherer Tupi speakers of Paraguay is controversial, but our results support the idea that they could be a Tupi-Guarani group who took refuge in the Paraguay forest and lost their agricultural lifestyle. Additionally, during their expansion from Amazonia or after their establishment in the region, the Ache may have made

TABLE 3. Mantel test considering the linguistic and geographic distributions

		C.C.	<i>P</i> -value	Partial C.C.	<i>P</i> -value
Tupi groups	mtDNA/Language	-0.184	0.82	-0.130	0.72
	mtDNA/Geography	0.286	0.02	0.256	0.03*
Je groups	mtDNA/Language	-0.321	0.92	-0.298	0.88
	mtDNA/Geography	0.126	0.27	-0.008	0.44

C.C., correlation coefficient; mtDNA, mitochondrial DNA.

* Significant *P*-value > 0.05.

use of a typical Tupi warfare cultural custom: capturing, not killing their non-Tupi enemies, and incorporating women and children (Callegari-Jacques et al., 2008). Our results suggest that captured females may have had a leading role in this gene background introgression. Regardless of the source of origin for the two Ache B2 haplotypes, their high frequencies stress the important role of drift in the evolutionary trajectory of this group, a process already observed with Y-chromosome markers (Bortolini et al., 2003).

Genetic versus geographic and historical-linguistic data

The linguistic variable used in this study was years of separation. Pairwise genetic difference matrices (Kimura 2*P* distance) for Tupi and Je populations and respective *P*-values are summarized in Supporting Information Table 3. Mantel tests were conducted separately for each language stock. The results for the Tupi speakers showed that the mtDNA distances are positively and significantly correlated at the 0.05 level of probability with geographic but not with historical/linguistic, distances. Je groups did not show any significant association between genetics and these variables (Table 3), even when only the seven Kayapó subgroups were considered. These results are in agreement with those observed by Hunley et al. (2007) who investigated mtDNA HVS-I sequences in 1,381 individuals from 17 Native American populations. These researchers concluded that the general pattern of relationship among languages is not the same as that among gene pools, although some concordance between these variables was observed in shallow branches.

To explore in more detail the correlation (or absence of it) between genetic, geographic, and historical/linguistic diversities, a spatial autocorrelation analysis was performed.

Spatial autocorrelation

Tupi and Je groups showed different plots of *r* values (Fig. 4). Because their populations present distinct geographic locations, occupying areas of different surfaces, the average distance interval for the 10 classes in the Tupi populations was 318 km, and was 245 km for the Je. All correlation coefficients fluctuate in a range of 0.5/0.6 and -0.3/-0.4. The two correlograms showed non-random space haplotype distributions and a positive and significant autocorrelation in the first distance class, which is expected when sequences are sampled in the same population or between populations in close proximity, possibly due to high levels of the gene flow or a common origin. The overall significance of the two correlograms was tested following Smouse et al.'s (2008)

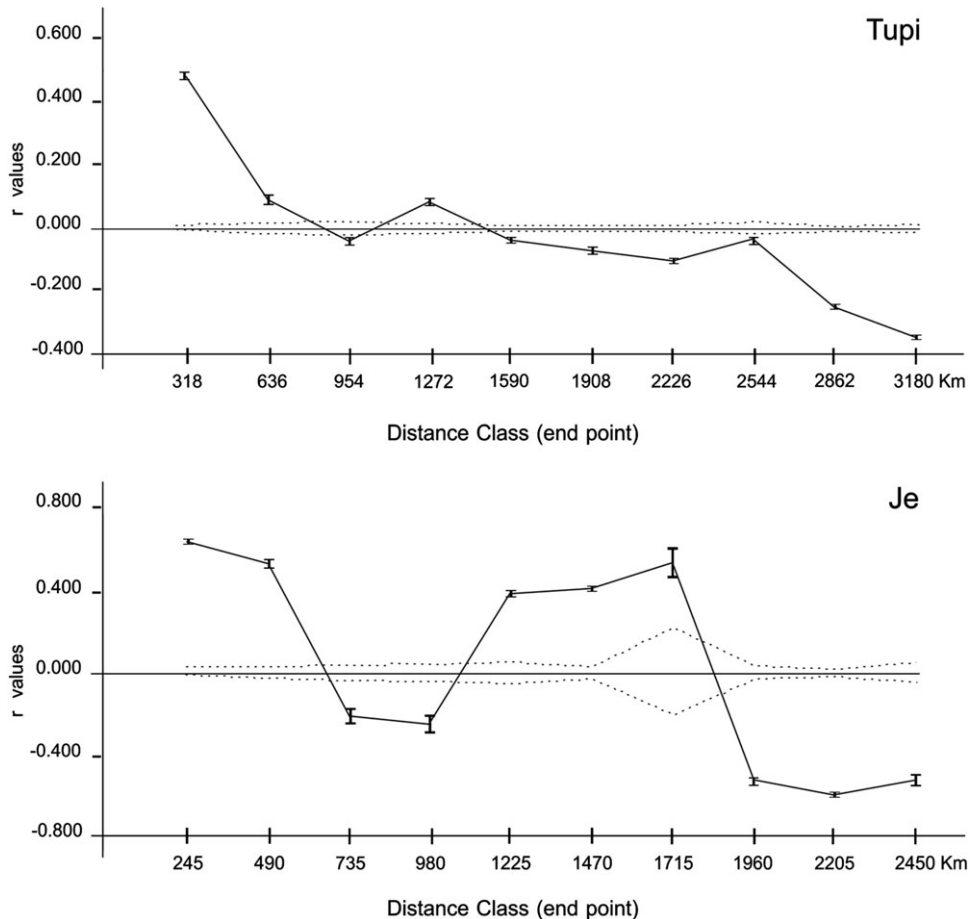


Fig. 4. Spatial structure analysis for the Tupi and Je populations. Upper and lower confidence limits (dotted lines) bound the 95% confidence interval on the null hypothesis of no spatial structure for the combined dataset as determined by 10,000 permutations.

procedures and showed a different spatial genetic structure for both Tupi and Je populations ($P = 0.010$). For the Tupi, the autocorrelations decreased more or less monotonically as the distances increased, indicating isolation by distance (Wright, 1943) and corroborating the Mantel test results (Table 3). For the Je, however, there is an increase in the autocorrelations from the distance classes 1,225 to 1,715 km, followed by a sharp decrease in the three most distant classes, indicating that some Je populations (Xavante and Kayapó groups) located far from each other have higher or similar levels of genetic identity than others located geographically closer. No correlation between geography and genetic distances was found in the Mantel test (Table 3) either, suggesting that other factors besides geography have influenced the mtDNA sequence distributions in Je populations.

DISCUSSION

Since his departure from Africa during an interglacial period that occurred at about 135,000–74,000 YBP (Armitage et al., 2011), anatomically modern *Homo sapiens* has experienced a most extraordinary migration, as compared to all the other primate species. Modern man has colonized all of the continents except Antarctica. Human arrival in the Americas occurred more recently, through one or more streams of migration from Asia,

between 18,000 and 2,000 YBP (González-José et al., 2008; Reich et al., 2012). Consequently, this migration brought a significant degree of genetic and cultural differentiation among groups. However, the transmission mode operating in these two systems is different: while genetic inheritance is vertical (from parents to offspring), cultural norms (language, for instance) are vertically and horizontally (between unrelated individuals) transmitted, and this may obscure the relationships between them (Hunley et al., 2007).

Here, we have conducted a genetic investigation involving the two main linguistic stocks found in Brazil. Taken together, our results reflect remarkable differences between the Tupi and Je expansions. Tupi speakers seem to have undertaken a radial expansion into new territories with an increase in genetic diversity as populations moved from their origin with subsequent isolation (as observed by F_{ST} values higher than that seen for isolated languages; Table 2). The shared presence of A2aa, as well as the A2ab lineage, found in Tupi individuals from populations located geographically distant from each other (A2aa: six samples from Poturuajara/North and fourteen from Guarani-Nandeva/South; A2ab: four Cinta-Larga, three Gavião, and six Zoró/Centra; seven Guarani-Nandeva, three Guarani-Kaiowá, and six Guarani-M'byá/South), can be considered a trace of their common origin. Subsequent processes of population

fissions could also lead to lineage differentiation: for example, the major division of the D1 network, which contained sequences only sampled among the Wayampi. The Mantel test and spatial autocorrelation analyses revealed a positive and significant correlation between their genetic and geographic distances, with a classical isolation-by-distance pattern.

We also observed that Tupi villages that fractionated long ago may have higher genetic similarity than other groups that separated recently. Note that these fissions generally occur along kinship lines and that they are generally associated with social tensions, such as warfare, beliefs in sorcery, revenge, disease, and other autochthonous conditions (Neel and Salzano, 1967; Salzano, 2009). This absence of significant correlation between genetic distances with historical/linguistic information may be due to several not mutually exclusive factors: (a) related females, carrying a particular mtDNA lineage that will be present in the new village but no longer in the original village; (b) gene flow, meaning that the villages could have fused again, through alliances or marriage exchanges, especially when the above-cited reasons for the fissions have been forgotten; (c) more recent Post-Columbian women movements; and (d) as mentioned in the “Materials and Methods” section, inaccuracies that may exist in the historical/linguistic dates of split. Additionally, it is possible to see that the Tupi expansion was also characterized by some level of absorption of non-Tupi women, including Je-speakers (e.g., the Ache case commented on in the Results section) who arrived in the South at least 1,000 years before the Tupi (Urban, 1998).

Network analysis considering the Je indicated several populations sharing lineages derived from a baseline haplotype. The major haplogroup C1 division includes samples from distant places, such as Kaingang and one of the Kayapó subgroups, Mekranoti. In the correlogram of the Je populations, the r values become positive at three distance classes and then dramatically decline to a much lower and negative value. The absence of significant correlation between genetic distances with geography and historical/linguistic data are an indication (at some level) of kinship influences, independent of the geographic distances and/or historical splits. Cases where isolation by distance patterns was absent or weak often indicate that a more complex suite of cultural, ecological, or physical factors may be responsible for structuring the populations.

For example, the combined Kayapó HVS-I plus HVS-II sequences showed lineages that may be present only in the derived group and not in the parental group anymore. The amount of exclusive lineages in the different Kayapó subgroups (60%) was somewhat unexpected, because the fragmentation of several Kayapó villages occurred just 100 years ago (Fig. 3). This may be due to simple stochastic processes, such as genetic drift. However, the ethnographic and historical information about the social dynamics of these populations should also be considered. The motives for the recent Kayapó fissions—warfare, beliefs in sorcery, revenge—seem to be still present in the collective memory of these groups (Neel and Salzano, 1967; Salzano, 2009), preventing subsequent fusions. This phenomenon appears to be, at least partly, responsible for other results, such as the proportion of shared lineages being higher among linguistic stocks (12%) than between populations of the same linguistic group (7%).

Finally, the Xavante population presents a particular set of social-cultural characteristics, like bilocal marital residence, community organization based on high endogamy, and a sororal polygyny marriage type. Hünemeier et al (2012) demonstrated that culture-specific traits are important to understand certain biological parameters, like fitness. The reproductive success of important men in the Xavante hierarchy is well documented in ethnographic studies (Coimbra et al., 2002). However, caution is needed regarding sociocultural practices because they are not stable over time. The traditional native culture could be very different from the present practices after European contact. Be that as it may, the perpetuation over a certain number of generations of a social practice like sororate unions could partly explain the genetic differences found in our results. If powerful men could marry more than one woman, and preferably sisters, all their children would share the same mitochondrial lineage, whose frequency could significantly increase. Additionally, cultural practices like those mentioned above have prevented fusion between the Xavante and the other Je populations located in the Brazilian Central Plateau, such as the Kayapó (Hünemeier et al., 2012).

On the basis of these results, we suggest that the intra- and inter-Tupi and Je group diversity could also have been mediated by females due to some specific sociocultural factors.

CONCLUSION

Our results suggest that the two main Brazilian-native linguistic stocks expanded in different ways. While the Tupi expanded through space following a well-defined radial pattern from a common origin, the Je presented an intricate and nonlinear mode of dispersion. Both the groups showed general traces of ancient fission–fusion processes mediated by women. Among the Je, these events did not produce congruent genetic, geographic, and historical scenarios, and this observation can be attributed, in part, to sociocultural factors. For instance, the reasons behind fissions that occurred only a few generations ago may continue to persist in the collective memory of the group, preventing fusions.

ACKNOWLEDGEMENTS

The authors are very grateful to the individuals who donated the samples analyzed here and to the Fundação Nacional do Índio (FUNAI) for logistic support in the sample collection. The authors also thank Jodi Irwin, Rebecca Just, Melissa Scheible, Jessica Saunier, Erin Gorden, and Rachel Rubenstein, from the Armed Forces DNA Identification Laboratory’s Research Section, for skilled technical assistance.

LITERATURE CITED

- Achilli A, Perego UA, Bravi CM, Coble MD, Kong QP, Woodward SR, Salas A, Torroni A, Bandelt HJ. 2008. The phylogeny of the four pan-American mtDNA haplogroups: implications for evolutionary and disease studies. *PLoS One* 3:e1764.
- Armitage SJ, Jasim SA, Marks AE, Parker AG, Usik VI, Uerpmann HP. 2011. The southern route “Out of Africa”: evidence for an early expansion of modern humans into Arabia. *Science* 331:453–456.
- Bandelt H-J, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16:37–48.

- Behar DM, van Oven M, Rosset S, Metspalu M, Loogväli E-L, 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.
- Bolnick DA, Shook BA, Campbell L, Goddard I. 2004. Problematic use of Greenberg's linguistic classification of the Americas in studies of Native American genetic variation. *Am J Hum Genet* 75:519–522.
- Bortolini MC, Salzano FM, Thomas MG, Stuart S, Nasanen SP, Bau CH, 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, Bradman N, Labuda D, Ruiz-Linares A. 2003. Y-chromosome evidence for differing ancient demographic histories in the Americas. *Am J Hum Genet* 73:524–539.
- Callegari-Jacques SM, Salzano FM. 1989. Genetic variation within two linguistic Amerindian groups: relationship to geography and population size. *Am J Phys Anthropol* 79:313–320.
- Callegari-Jacques SM, Hill K, Hurtado AM, Rodrigues LT, Bau CH, Salzano FM. 2008. Genetic clues about the origin of Aché hunter-gatherers of Paraguay. *Am J Hum Biol* 20:735–737.
- Callegari-Jacques SM, Tarazona-Santos EM, Gilman RH, Herrera P, Cabrera L, Santos SE, Morés L, Hutz MH, Salzano FM. 2011. Autosomal 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: the historical linguistics of native America*. New York: Oxford University Press.
- Carneiro da Cunha M. 1998. *História dos índios no Brasil*, 2nd ed. São Paulo: Companhia das Letras.
- Coimbra CEA, Flowers NM, Salzano FM, Santos RV. 2002. The Xavante in transition. Health, ecology, and bioanthropology in Central Brazil. Ann Arbor: The University of Michigan Press.
- Epperson BK. 2003. *Geographical genetics*. Princeton, NJ: Princeton University Press.
- Excoffier L, Lischer HEL. 2010. Arlequin suite ver. 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Res* 10:564–567.
- Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479–491.
- González-José R, Bortolini MC, Santos FR, Bonatto SL. 2008. The peopling of America: craniofacial shape variation on a continental scale and its interpretation from an interdisciplinary view. *Am J Phys Anthropol* 137:175–187.
- Greenberg JH. 1987. *Languages in the Americas*. Stanford, CA: Stanford University Press.
- Halbwachs M. 1992. *On collective memory*. Chicago: University of Chicago Press.
- Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* 41:95–98.
- Heckenberger MJ, Neves EG, Petersen JB. 1998. De onde surgem os modelos? As origens e expansões Tupi na Amazônia Central. *Rev Antropol* 41:69–96.
- Hünemeier T, Gómez-Valdés J, Ballesteros-Romero M, Soledad de Azevedo, Neus Martínez-Abadías, Mireia Esparza, Torstein Sjøvold, Sandro L. Bonatto, Francisco Mauro Salzano, Maria Cátira Bortolini, Rolando González-José. 2012. Cultural diversification promotes rapid phenotypic evolution in Xavante Indians. *Proc Natl Acad Sci USA* 109:73–77.
- Hunley KL, Cabana GS, Merriwether DA, Long JC. 2007. A formal test of linguistic and genetic coevolution in native Central and South America. *Am J Phys Anthropol* 132:622–631.
- Irwin JA, Saunier JL, Strouss KM, Sturk KA, Diegoli TM, Just RS, Coble MD, Parson W, Parsons TJ. 2007. Development and expansion of high-quality control region databases to improve forensic mtDNA evidence interpretation. *Forensic Sci Int Genet* 1:154–157.
- Kemp BM, González-Oliver A, Malhi RS, Monroe C, Britt Schroeder K, McDonough J, Rhatt G, Resendéz A, Peñalosa-Espinosa RE, Buentello-Malo L, Gorodesky C, Smith DG. 2010. Evaluating the farming/language dispersal hypothesis with genetic variation exhibited by populations in the Southwest and Mesoamerica. *Proc Natl Acad Sci USA* 107:6759–6764.
- Kimura M. 1980. A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. *J Mol Evol* 16:111–120.
- Lewis MP, editor. 2009. *Ethnologue: languages of the world*, 16th edition. Dallas: Summer Institute of Linguistics. Online version: <http://www.ethnologue.com/>.
- Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452.
- Lobato-da-Silva DF, Ribeiro-dos-Santos AKC, Santos SEB. 2001. Diversidade genética de populações humanas na Amazônia. In: Guimarães Vieira IC, Cardoso da Silva JM, Oren DC, D'Inea MA, editors. *Diversidade Humana e Cultural na Amazônia*. Belém: Museu Paraense Emílio Goeldi. p 167–193.
- Loukotka Č. 1968. *Classification of South American Indian languages*. Los Angeles: University of California.
- Mantel N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res* 27:209–220.
- Marrero AR, Silva-Junior WA, Bravi CM, Hutz MH, Petzl-Erler ML, Ruiz-Linares A, Salzano FM, Bortolini MC. 2007. Demographic and evolutionary trajectories of the Guarani and Kaingang natives of Brazil. *Am J Phys Anthropol* 132:301–310.
- Mazieres 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.
- Migliazza EC. 1982. Linguistic prehistory and the refuge model in Amazonia. In: Prance GT, editor. *Biological diversification in the tropics*. New York: Columbia University Press. p 497–519.
- Moore D. 2005. Brazil: language situation. In: Brown K, editor. *Encyclopedia of languages and linguistics*. Amsterdam: Elsevier. p 117–127.
- Neel JV, Salzano FM. 1967. Further studies on the Xavante Indians. X. Some hypotheses-generalizations resulting from these studies. *Am J Hum Genet* 19:554–574.
- Nei M. 1987. *Molecular evolutionary genetics*. New York: Columbia University Press.
- Neves WA, Bernardo DV, Okumura M, Almeida TF, Strauss AM. 2011. Origin and dispersion of the Tupiguarani: What does cranial morphology say? *Bol Mus Para Emílio Goeldi Cienc Hum* 6:95–122.
- Nichols J. 2008. Language spread rates as indicators of glacial-age peopling of the Americas. *Curr Anthropol* 49:1109–1117.
- Noelli FS. 1998. The Tupi: explaining origin and expansions in terms of archeology and of historical linguistics. *Antiquity* 72:648–663.
- Noelli FS. 2008. The Tupi expansion. In: Silverman H, Isbell W, editors. *Handbook of South American archaeology*. New York: Springer. p 659–670.
- O'Rourke DK, Suarez BK. 1986. Patterns and correlates of genetic variation in South Amerindians. *Ann Hum Biol* 13:13–31.
- Peakall R, Smouse PE. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* 6:288–295.
- Peakall R, Ruibal M, Lindenmayer DB. 2003. Spatial autocorrelation analysis offers new insights into gene flow in the Australian bush rat, *Rattus fuscipes*. *Evolution* 57:1182–1195.
- Polzin T, Daneschmand SV. 2003. On Steiner trees and minimum spanning trees in hypergraphs. *Operat Res Lett* 31:12–20.
- Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, Ray N, Parra MV, Rojas W, Duque C, Mesa N, Garcia LF, Triana O, Blair S, Maestre A, Dib JC, Bravi CM, Bailliet G, Corach D, Hünemeier T, Bortolini MC, Salzano FM, Petzl-Erler ML, Acuña-Alonso V, Aguilar-Salinas C, Canizales-Quinteros S, Tusié-Luna T, Riba L, Rodríguez-Cruz M, Lopez-Alarcón M, Coral-Vazquez R, Canto-Cetina T, Silva-Zolezzi I, Fernandez-Lopez JC, Contreras AV, Jimenez-Sanchez G, Gómez-Vázquez

- MJ, Molina J, Carracedo A, Salas A, Gallo C, Poletti G, Witonsky DB, Alkorta-Aranburu G, Sukernik RI, Osipova L, Fedorova SA, Vasquez R, Villena M, Moreau C, Barrantes R, Pauls D, Excoffier L, Bedoya G, Rothhammer F, Dugoujon JM, Larrouy G, Klitz W, Labuda D, Kidd J, Kidd K, Di Rienzo A, Freimer NB, Price AL, Ruiz-Linares A. 2012. Reconstructing Native American population history. *Nature* doi: 10.1038/nature11258.
- Rodrigues A. 1986. Línguas brasileiras: para o conhecimento das línguas indígenas. São Paulo: Edições Loyola.
- Rodrigues AD. 1964. Classificação do tronco linguístico Tupi. *Rev Antropol* 12:99–104.
- Rodrigues AD. 2000. Flexão relacional no tronco linguístico Macro-Jê. *Bol Assoc Bras Lingüíst* 25:219–231.
- Rodrigues AD. 2002. Línguas brasileiras: para o conhecimento das línguas indígenas, 4th ed. São Paulo: Edições Loyola.
- Russell N. 2006. Collective memory before and after Halbwachs. *French Rev* 79:792–804.
- Salzano FM. 1971. Demographic and genetic interrelationships among the Cayapo Indians of Brazil. *Soc Biol* 18:148–157.
- Salzano FM. 2009. The fission-fusion concept. *Curr Anthropol* 50:959–959.
- Salzano FM, Black FL, Callegari-Jacques SM, Santos SEB, Weimer TA, Mestriner MA, Pandey JP, Hutz MH, Rieger TT. 1988. Genetic variation within a linguistic group: Apalai-Wayana and other Carib tribes. *Am J Phys Anthropol* 75:347–356.
- Santos SEB, Ribeiro-Rodrigues EM, Ribeiro-Dos-Santos AK, Hutz MH, Tovo-Rodrigues L, Salzano FM, Callegari-Jacques SM. 2009. Autosomal STR analyses in native Amazonian tribes suggest a population structure driven by isolation by distance. *Hum Biol* 81:71–88.
- Schmitz PI. 1997. Migrantes da Amazônia: a tradição Tupi-Guarani. In: Kern AA (org) *Arqueologia pré-histórica do Rio Grande do Sul*, 2nd ed. Porto Alegre: Mercado Aberto. p 295–330.
- Smouse PE, Peakall R, Gonzales E. 2008. A heterogeneity test for fine-scale genetic structure. *Mol Ecol* 17:3389–3400.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28:2731–2739.
- Urban G. 1996. On the geographical origins and dispersion of Tupian languages. *Rev Antropol* 39:61–104.
- Urban G. 1998. A história da cultura brasileira segundo as línguas nativas. In: Carneiro da Cunha M, organizer. *História dos índios no Brasil*. São Paulo: Companhia das Letras. p 87–102.
- van Oven M, Kayser M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* 30:386–394.
- Ward RH, Salzano FM, Bonatto SL, Hutz MH, Coimbra CEA, Santos RV. 1996. Mitochondrial DNA polymorphism in three Brazilian Indian tribes. *Am J Hum Biol* 8:317–323.
- Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358–1370.
- Weir BS. 1996. The second National Research Council report on forensic DNA evidence. *Am J Hum Genet* 59:497–500.
- Wright S. 1943. Isolation by distance. *Genetics* 28:114–138.
- Yang NN, Mazières S, Bravi C, Ray N, Wang S, Burley MW, 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.