Origins and demographic dynamics of Tupí expansion: a genetic tale
Origens e dinâmica demográfica da expansão Tupi: uma história genética

Eduardo José Melo dos Santos[I], Andréa Luciana Soares da Silva[I]

Paloma Daguer Ewerton[I], Louise Yukari Takeshita[II], Maria Helena Thomaz Maia[I]

[I]Institutional affiliation: Human and Medical Genetics, Federal University of Pará, Belém, Brazil

[II]Institute of Integrative Biology, University of Liverpool, United Kingdom

Abstract: Tupí linguistic groups display a wide geographical dispersion in South America, probably originated, as pointed by linguistic, from Madeira-Guaporé Region (MGR) in Brazil. The present study reviewed genetic data on Tupians for autosomal and uniparental (Y-chromosome and mtDNA) markers, using it to evaluate Tupians geographic origin as well as the demographic dynamics of their dispersion from a genetic point of view. Comparison of genetic variability and mtDNA haplogroups D frequencies suggests a scenario where MGR is the Tupí homeland. The relationship between five estimators of genetic variability (Thetas-S, -Pi, -m2, -H and -k) shows that Tupí groups from MGR and non-MGR experienced different patterns of demographic dynamics, with an ancient Tupí expansion in MGR, followed by dispersion to other South America regions, probably associated to depopulation/founder effect events. Furthermore, other recent depopulation events could also be detected in both regions. Finally, the dispersion seems to be related to patrilocality, as suggested by comparison of uniparental markers genetic differentiation. This genetic model of dispersion dynamics may have an important impact in the interpretation of archeological and linguistic data, allowing to test if female associated technologies, like ceramic, are more extensively shared between dispersed populations than those which are not female-exclusive.

Keywords: Tupí homeland. Demographic. Tupí expansion. Patrilocality. Genetic markers.

Resumo: Grupos linguísticos Tupí exibem uma grande dispersão geográfica na América do Sul, provavelmente originadas, conforme apontada pela linguistica, na Região Madeira-Guaporé (MGR), Brasil. O presente estudo revisou dados geneticos de populacoes Tupí para marcadores autossomicos e uniparentais (Y-cromossomo and mtDNA), delineando abordagens para avaliar origem geográfica, bem como a dinâmica demográfica de sua dispersão, de um ponto de vista genético. Comparação da variabilidade genetica e das frequencias do haplogrupo D do mtDNA sugere um cenário onde MGR é o local de origem Tupí. A relação entre cinco estimadores de variabilidade genetica Thetas-S, -Pi, -m2, -H e -k) mostram que grupos Tupí da MGR não MGR experienciaram diferentes padrões de dinâmica demográfica, com uma antiga expansão Tupí em MGR, seguida de dispersão para outras regiões da América do Sul, provavelmente associada a eventos de depopulação/efeito fundador. Além disso, outros eventos recentes de depopulação também puderam ser detectados em ambas as regiões. Finalmente, a dispersão parece estar relacionada com práticas de patrilocalidade, como sugerido pela comparação de marcadores uniparentais. Este modelo genético de dinâmica de dispersão Tupí pode ter um impacto importante na interpretação dos dados arqueológicos e linguísticos, permitindo testar se tecnologias associadas ao sexo feminino, como a cerâmica, são mais amplamente compartilhadas entre populações dispersas do que tecnologias não exclusivas do sexo feminino.

Palavras Chave: Origem Tupí. Demografia. Expansão Tupí. Patrilocalidade. Marcadores Genéticos.

INTRODUCTION

Amerindian linguistic composition in Brazil encompasses between 154 and 170 languages clustered in 20 major groups (Moore, 2005), most of them located in Amazonia. This region shows a huge linguistic diversity, with at least 52 linguistic families (Epps, 2009). Considering the size and geographic dispersion, four main families can be highlighted: Tupí, Arawak, Carib e Macro-Je (Epps, 2009).

The Tupí linguistic family is widely dispersed in South America, branching in ten groups, most of them located in the State of Rondônia (Gabas, 2006), a region close to Madeira and Guaporé rivers (Madeira-Guaporé Region-MGR). Five branches (Arikém, Mondé, Puruborá, Ramaráma and Tuparí, besides some dialects of Kawahíb complex) are mainly located in MGR, while the dispersion of remaining branches reached regions outside Amazonia, noteworthy the branch Tupí-Guarani, which has occupied areas in the Brazilian South, Southeast and in the East coast (Gabas, 2006; Epps, 2009).

Linguistic and genetic data support the origin of the Tupí family between 2800 and 3000 years ago (Urban, 1998; Amorim *et al.*, 2013). Its geographic origin is still controversial, although the Amazonia region has been considered the best candidate. Linguistic analyses point to the MGR as the putative origin, while some archeological data suggest the confluence of the Madeira and Amazon rivers as a better alternative (Rodrigues, 1964; Migliazza, 1982; Urban, 1996, 1998; Heckenberger *et al.*, 1998; Noelli, 1998). Morphological studies do not suggest any specific origin, but they also agree with an Amazonian origin (Neves *et al.*, 2011).

The linguistic-based proposition of MGR as geographic origin of Tupí family is well accepted, mainly due to higher Tupian linguistic variability of this region. Moreover, some archeological peculiarities of MGR also support this idea. This region has evidences of continuous human occupation for at least 9000 years, having the most ancient "dark soil" site (anthropogenic darkened soil), dating to 4700 years before the present (bp) (Zimpel-Neto, 2009). Old ceramic patterns found in such "dark soils" were indirectly associated to Tupian ceramic patterns (Zimpel-Neto, 2009). In this context, the origin of Tupí and their main dispersion events should have occurred along the last 3000 years (Marrero *et al.*, 2007), and some authors propose agriculture as the most important factor for their wide dispersion across South America (Epps, 2009).

Regarding the use of genetic data for infer the past of Amerindians, historically the earlier genetic datasets were restricted to blood groups. In the 1970s, with the development of electrophoresis and serological methods, the number of known genetic markers increased, including variants of many proteins from serum and blood cells. However, the genetic variability detected by those markers, denominated classical markers, is low (Rogers and Jorde, 1996).

In the 1990s, it was possible to obtain molecular data from direct DNA analysis, which revealed a substantial number of markers. This breakthrough allowed researchers to detect high levels of genetic variability on human populations, permitting more detailed inferences about their histories. More specifically, mtDNA and Y-chromosome polymorphisms, called uniparental markers, were also extensively explored, being used to reconstruct distinctly male and female biological histories (Mulligan *et al.*, 2004).

The main advantage of classical markers is that they were investigated for a large number of Amerindian tribes. However, they are mostly proteins and reflect only the variability of the coding regions of the genes. It is possible to suggest that some of them will be under natural selection pressure and therefore represents only a small part of human genome variability.

Despite modern molecular markers have been extensively studied in Amerindian populations, the number of populations investigated is still low when compared with the classical ones. Nonetheless, molecular markers still remain as a more indicated option to perform complex analysis, helping to infer more details about the demographic history of populations.

The main types of molecular markers are Single Nucleotides Polymorphisms (SNPs), Short Tandem Repeats

(STRs), *Alu* insertions and specific DNA sequences. These markers can be either found in autosomes (inherited equally from both parents), or in Y-chromosome or mtDNA, being inherited in a patrilineal and matrilineal way, respectively.

Briefly, SNPs are mutations involving one nucleotide in DNA sequences quite abundant in the human genome, mostly having two alternative forms (alleles) and low variability. *Alu* insertions are sequences of approximately 300 bases scattered in the genome, where their insertion is a polymorphism, i.e., an individual may have an *Alu* insertion in a given position, whereas others might not have any insertion in the same position. *Alu* insertions are restricted to primate genomes, are easy to genotype and have been used in many population studies (Novick *et al*., 1995). STRs are tandemly repeated sequence motifs from one to six bases. Their variation is related to the number of times the motif is repeated. These markers are also abundant in the genome, easy to genotype and have many alleles, being highly polymorphic (Tautz, 1993). Some DNA sequences are also highly polymorphic and commonly composed by few hundred of bases, where each base is a potential SNP. Patrilineal markers (*Alu* insertions, SNPs, and STRs found in the Y-chromosome) are used to reconstruct male demographic history, while matrilineal markers (mtDNA SNPs and mtDNA sequences) reconstruct female history (Mulligan *et al*., 2004).

All mitochondria have a genome compose by a single DNA strand. Some key positions, or SNPs, in the nucleotide sequence are likely used to define groups of sequences, denominated haplogroups. The combination of only four SNPs in mtDNA allows to define five haplogroups, exclusively observed in Amerindian populations, called haplogroups A, B, C, D and X. Additionally, the sequencing of a specific region of mtDNA with approximately 300 base pairs, called the hypervariable region (HVR I), provides a highly variable marker, very useful to infer the past demographic history of human female lineages (Mulligan *et al*., 2004).

Genetic variability is the result of both selective pressures and stochastic events. Usually, the level of intrapopulation genetic variability can be measured by at least five different statistical measures or estimators. The first one and most simple is the number of alleles (k) that can be estimated for all types of genetic markers (SNPs, STRs, *Alu* insertions and DNA sequences). Heterozigosity (H) is a very robust statistics that evaluates the genetic variability for all kind of markers, considering the number of alleles and their frequencies. There is an extra statistical measure exclusive to STR markers called variance of the number of repeats (m2), which is based on the number of alleles, their frequencies and the degree of difference between alleles. Moreover, DNA sequence markers have two additional statistical measures, the nucleotide diversity (Pi) and number of polymorphic sites (S) (Chakraborty *et al*., 1988; Slatkin, 1995, Shriver *et al*., 1997, Tajima, 1989).

All five statistical measures can be independently used to estimate a summary statistic called Theta. Theta is an estimate that could be calculated from all genetic variability statistics aforementioned and is equal to 4Nu (for diploid loci), where N is the effective size of the population and u is the mutation rate. It is noteworthy that effective size is not the same as the census population size, but is related to the number of individuals reproductively active. However, this relation holds only if the population is under mutation-genetic drift equilibrium. Mutation is an evolutionary factor that introduces variability in the population while genetic drift is the random variation of allele frequencies that leads to a loss of variability. Thus, in absence of selective pressure, a given population retains its effective size for many generations and no expressive migrations occurs the population reach this equilibrium state and theta estimates obtained from different variability estimators should be very similar (Chakraborty *et al*., 1988; Slatkin, 1995, Shriver *et al*., 1997, Tajima, 1989).

Populations are not always in mutation-genetic drift equilibrium, sometimes they are in a called transient state, caused by fluctuations in population size over the time, as bottlenecks and rapid expansions. These demographic events affect the variability estimators in different ways. Thus, in transient state, theta estimates obtained from

distinct variability estimators will not be similar. Moreover, each kind of demographic event, expansion or bottleneck, causes different signatures in terms of the relation between thetas from different estimators. This feature allows evaluate these differences, inferring about past demographic history of the populations (Slatkin, 1994, 1995).

The dynamics of Tupí expansion was explored only recently by Ramallo et al., (2013), suggesting a radial expansion into new territories, starting putatively from MGR. However, no genetic studies have tested for the hypothesis of Tupí origin in MGR, except for one previous study from our group (Santos et al., 2013).

Genetic studies have accumulated data from Tupí populations for many decades. Here, the review the available Tupí genetic data allows inferring both its geographic origins and the demographic dynamics of their dispersion issues. The main results of present study found evidence for MGR as the Tupí homeland, and that this dispersion was accompanied by depopulation and female-biased migration.

MATERIAL AND METHODS

Genetic data review: Genetic data on several types of loci, like classic protein markers, STR, Alu insertions, Y-STRs and mtDNA were extensively gathered from literature. The sources were given accordingly while the results are presented in the next section.

Study design: In order to evaluate the geographic origin of Tupians two groups were defined, the MGR and non-MGR Tupians. Two approaches were used; the first compares the heterozygosity in populations from MGR and outside MGR for a number of genetic markers groups; the second compares the frequencies of mtDNA haplogroups from both regions. All comparisons were performed by Wilcoxon paired test and Mann-Whitney test.

The demographic dynamics inferences are based on Theta estimates obtained from five different measures of genetic variability, H, k, m2, Pi and S, as described above. The first approach evaluates the relation between Thetas k, H and m2 from STR data, as described in the results and

discussion section and the comparison was performed by paired Wilcoxon test. The second approach investigates the relationship between mtDNA sequence's Theta S and Pi, since they provide clues about population expansion or depopulation (Tajima, 1989).

The evaluation of migration processes during expansion of Tupians was made comparing uniparental markers, mtDNA and Y-chromosome markers, testing the trend across a number of studies by Wilcoxon paired test.

RESULTS AND DISCUSSION

GEOGRAPHIC ORIGIN OF TUPI GROUPS

As previously mentioned, the most accepted hypothesis about the demographic dynamics of Tupí supports MGR as homeland followed by a wide and rapid dispersion of Tupí-Guarani languages across the South America. Such kind of expansions is frequently associated with important technological or cultural innovation and agriculture is putatively responsible for this expansion (Urban, 1998; Gabas, 2006; Epps, 2009, Walker et al., 2012). The idea that after dispersion the original populations would display more variability than the dispersed ones proved to be true not only in linguistic, but also in genetic studies.

Indeed, at global scale, the sub-Saharan origin of modern humans is well established by paleontological and genetic data (Nei and Livshits, 1989). The inspection of genetic variability clearly demonstrates that Africans have higher heterozygosity than non-Africans (Rogers and Jorde 1996; Jorde et al., 2000). Y-chromosome data showed that the variability decreases with the distance from east Africa, compatible with sequential founder effect that is characteristic of population dispersion (Shi et al., 2010). The same effect was observed in Amerindian populations. Using 678 STRs in 29 populations Wang et al., (2007) demonstrated that the decline of genetic variability correlated with the distance from the Bering Strait region.

In this context, we estimated the variability of genetic markers previously described in the literature using

the statistical measure of Heterozigozity (H). The main observation was that H is higher in Tupian from MGR than in those outside MGR, for all kind of markers (Table 1), suggesting MGR as Tupí homeland.

An additional approach to track dispersion and migration paths is the identification of clines in allele frequencies. We analyzed the frequencies of the major mtDNA haplogroups (A, B, C and D) in Tupí populations (Figure 1). MGR populations show the highest frequencies of haplogroup D (all over 60%), while Tupians outside MGR displayed frequencies lower than 30%, except for Mundurukú (55%), which is geographically closer to MGR (Mann-Whitney test, Z(U) =3.12; p=0.0018). This scenario is compatible with a Tupí dispersion center in MGR and, during the dispersion, contact with non-Tupí populations, which had higher frequencies of non-D haplogroups, lowered haplogroup D frequencies in Tupí outside MGR.

DEMOGRAPHIC DYNAMICS OF TUPI DISPERSION

Estimates of genetic variability have a direct relationship with the effective size of population and the mutation rate of each marker. Hence, large populations should have high values of these estimators. Moreover, markers such as STRs may lead to higher values than SNPs or classical markers because their higher mutation rates. Given the importance of population size in the degree of genetic variability, it is understandable that sudden demographic variations have a significant impact on those estimates and can be traced back

Table 1. Heterozygozity estimate to Madeira-Guaporé Region (MGR) and Non-Madeira-Guaporé (No-MGR) region populations.

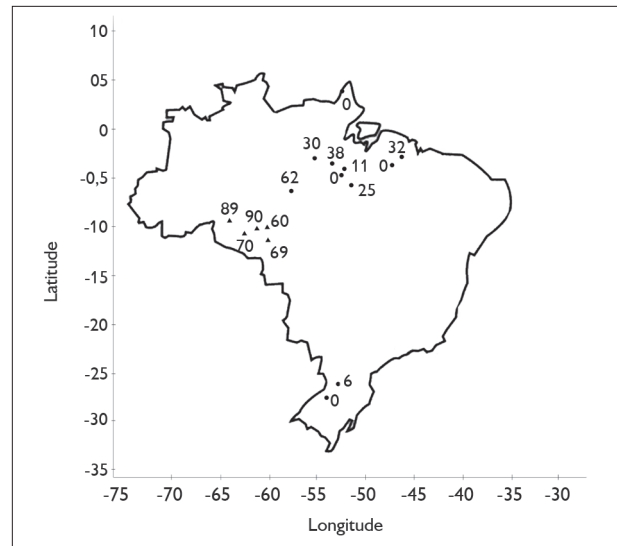| Markers	MGR	No-MGR
Classics[1]	0.193	0.172
STR[2]	0.683	0.632
Y-STR[3]	0.253	0.164
Alu insertions[4]	0.235	0.213
mtDNA sequence[5]	0.865	0.845

Figure 1. Frequencies (%) of mtDNA haplogroup D in MGR Tupi populations and in Tupi populations outside MGR (Source: Silva *et al*., 2001; Marrero *et al*., 2007). Triangles represent MGR populations (Zoró, Gavião, Suruí, Cinta Larga, Karitiana) and circles populations outside MGR (Mundurukú, Asuriní do Trocará, Asuriní do Coatinemo, Araweté, Urubú-Kaapór, Parakanã, Awá-Guajá, Wayampi, Zoé; Ache; Guarani). Comparison between MGR and outside MGR populations was statistically significant (Mann-Whitney test; Z (U) = 3.12; p=0.0018).

---

[1] Calculated from 24 classical markers (Guerreiro *et al.*, 1994). Heterozigosity (H) estimated from three populations classified as MGR: Suruí, Karitiana and Cinta Larga; and six Non-MGR populations: Wayampi, Urubú-Kaapór, Asurini do Trocará, Asurini do Coatinemo, Parakanã and Araweté.
[2] Calculatad from 11 STR (Callegari Jacques *et al.*, 2011). H estimated from three MGR populations: Gavião, Zoró and Suruí) and eight No-MGR populations: Wayampi, Emerillon, Zoé, Urubú-Kaapór, Awá-Guajá, Parakanã, Guaraní and Aché).
[3] Data from four Y-chromosome STR (Tarazona-Santos *et al.*, 2001; Palha *et al.*, 2010). H estimated from four MGR populations: Karitiana, Zoró, Suruí and Gavião; and five Non-MGR populations: Wayampi, Zoé, Urubú-Kaapór, Awá-Guajá and Parakanã.
[4] Data from 12 Alu insertions (Battilana *et al.*, 2006). H estimated from four MGR populations: Cinta Larga, Gavião, Suruí and Zoró; and two Non-MGR populations: Guaraní and Aché.
[5] H estimative from hyper variable region I of mtDNA (source: Ewerton *et al.*, 2011). H from three MGR populations: Gavião, Suruí and Zoró; and six Non-MGR populations: Wayampi, Zoé, Urubú-Kaapór, Awá-Guajá, Mundurukúand Guaraní. Comparisons of H estimative between both MGR and Non-MGR populations were statistically significant (Wilcoxon paired test; *p-value*=0.04).

by analysis of degree of genetic variability in each marker. An important feature in this analysis is that each estimator responds differently to demographic events, like population bottlenecks (intense depopulation) and expansions.

Aiming to answer some questions about the demographic dynamics of the Tupí, we reviewed autosomal STR data available for Tupí populations living in MGR and outside MGR. For this kind of markers it is possible to estimate three measures of genetic variation: k, H and m2. In a theoretical model of a stable population, it is expected that Theta k, Theta H and Theta m2 (Thetas estimated from k, H and m2, respectively) values are similar to each other. Nonetheless, given that actual populations are not stationary and some dramatic alterations in the population size may occur, theta values tend to be conflicting. For example, if a population experienced a severe bottleneck, the most expected scenario regarding Theta values is: Theta k < Theta H < Theta m2 (King et al., 2000; Cournet et al., 1997). This expectation fits well to present findings using autosomal STR, as presented in Figure 2. Ewerton (2011) showed that Theta k < Theta H using paired Wilcoxon test. This finding held not only for Tupians, but also for non-Tupians tribes, suggesting a general depopulation process. We performed also paired Wilcoxon test and showed that H < Theta m2 as well (p<0.01 for all comparisons).

Once this expectation is able to detect recent depopulation events (Garrigan; Hedrick, 2003), we suggest that the observed pattern reflects a recent depopulation experienced by both MGR and non-MGR Tupí populations (and probably for non-Tupian populations). Some recent events such as European colonization caused intense depopulation in Amerindians, and are likely to be the reason for this recent bottleneck scenario. Of course European colonization is not the unique explanation for the depopulation signature. During the expansion process, the non-MGR Tupians probably experienced depopulation events. Otherwise, MGR tribes have experienced strong reduction of their population sizes in the last many decades due to the exploitation of natural resources of the region
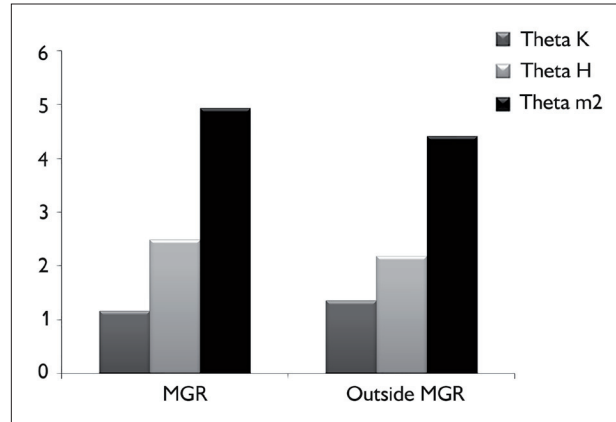

Figure 2. Theta estimators obtained from number of alleles (Theta k), heterozygosity (Theta H) and variance in number of repetitions (Theta m2). All statistics were estimated for 11 STRs in Tupí populations from MGR (Zoró, Gavião, Suruí) and outside MGR (Ache, Urubú-Kaapór, Awá-Guajá, Wayampi, Zoé, Guarani, Parakanã). Source: Ewerton 2011.

and intense contact with diseases. Hence, the signature of depopulation is likely to be a cumulative effect of many independent recent events, putatively not restricted to Tupians. Additionally, is important to highlight that this signature reflects recent events. H, K and m2 take more than N (effective size) generations to come back to equilibrium (Garrigan; Hedrick, 2003). Thus, even events occurred few generations ago are able to disrupt the mutation-genetic drift equilibrium.

Additionally, in agreement with the previous results and conclusion, the genetic variability showed also higher values in MGR populations than in non-MGR ones. Restricting the comparison to Theta H the difference was statistically significant (Wilcoxon paired test, p=0.02).

As previously shown, mtDNA haplogroup D has a high frequency in MGR Tupians and a low frequency in Tupians populations outside MGR. Hence, considering haplogroup D as an important marker for Tupí groups, we reviewed mtDNA HVR I sequences and estimated two Theta values from two genetic variability estimators (S and Pi), which are specific for DNA sequences. The pattern Theta S < Theta Pi is indicative of an ancient depopulation, that was observed in Tupians outside MGR.

On the other hand, an opposite pattern was observed in MGR (Theta S > Theta Pi), which is characteristic of an ancient population expansion (Figure 3). This approach is equivalent to the Tajima's D statistics that detected ancient expansion signals among all major haplogroups in South American Indians (Bonatto; Salzano, 1997). The present analysis restricted the analysis to haplogroups D and compared both regions. The ancient expansion signal could be observed in MGR but seems to be lost outside MGR. This trend agrees with other results of the present meta-analysis and those from other authors, as reviewed. The intense process of dispersion characterized by sequential founder effects and bottlenecks could explain the lack of expansion signal, supporting the hypothesis of MGR as the Tupí homeland.

MIGRATION PROCESSES DURING TUPI EXPANSION

Patrilocality is defined as lower male mobility when compared to female, meaning that females migrate (or are exchanged among distinct populations) in a high rate than males. This phenomenon is widely studied in the literature using uniparental genetic data at regional scale and even worldwide, as reviewed bellow. The most common approach to assess patrolocality is based on the comparison between interpopulation genetic differentiation estimated for matrilineal (mtDNA) and patrilineal (Y-chromosome) markers. The most common estimate of differentiation between populations is called F statistics or F_{ST}. Over time, two populations accumulate genetic differences by genetic drift, but this differentiation may be delayed by factors like migration. Since, each uniparental marker should inform separately of the history of females and males, and differences between F_{ST} estimates from both sets of markers should be attributable to differences in the intensity of migration between genders. This is because migration decreases the differentiation between populations, thus decreasing F_{ST}.

The first study comparing mtDNA and Y-chromosome markers was performed in a global scale (Seielstad *et al*., 1998),
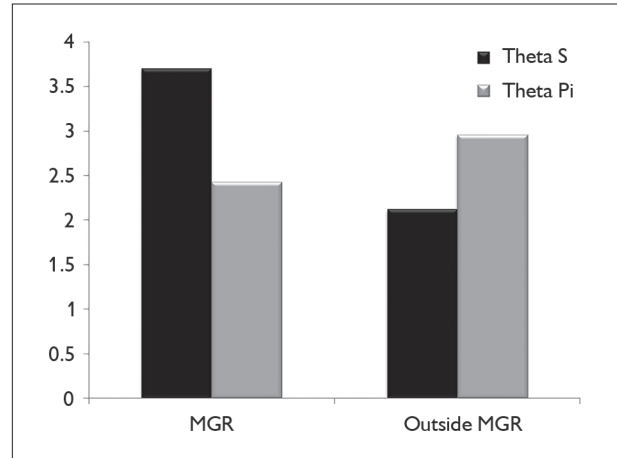

Figure 3. Theta estimators obtained from number polymorphic sites (Theta S) and Pi (Theta Pi). All statistics were estimated for haplogroup D mtDNA sequences in Tupí populations from MGR (Zoró, Gavião, Suruí) and outside MGR (Mundurukú, Asuriní do Trocará, Asuriní do Coatinemo, Araweté, Urubú-Kaapór, Parakanã, Awá-Guajá, Wayampi, Zoé; Ache; Guarani). Source: Ewerton, 2011.

suggesting that female migration was eight times higher than males ones. Jorde *et al*., (2000) demonstrated the same trend in Asia and Europe, but not in Africa. However, further studies, analyzing more populations and uniparental markers, showed some controversial results worldwide, but did not discard regional patrilocality (Wilder *et al*., 2004).

This controversy has been well discussed by Wilkins and Marlowe (2006). For these authors, inferences about male and female demographic histories taken from uniparental markers are not simple and need more complex modeling. Sampling of close populations reveals recent and local population dynamics, while worldwide sampling reflects more ancient events. In this context, studies using worldwide sampling will reveal earlier sex biased migration patterns (Wilkins; Morlowe, 2006).

When focusing on Amerindian populations, initial studies were also controversial (Mesa *et al*., 2000; Goicoechea *et al*., 2001; Bortolini *et al*., 2002; Fagundes *et al*., 2002; Hunley *et al*., 2008), likely due to sampling limitations, low number or inadequacy of markers and even the use of populations with heterogeneous degree of agriculture development. A more recent study, using

larger sets of populations and uniparental markers, clearly demonstrates evidences of patrilocality across many regions of the American Landmass (Yang *et al.*, 2010).

When looking specifically for patterns of uniparental marker differentiation in Tupí populations, Mazieres *et al.*, (2011) have found a higher differentiation in Y-chromosome markers than in mtDNA in Tupí groups from French Guiana, which indicates patrilocality (Table 2). It is important to test if this pattern has also been observed outside French Guiana, and therefore, we performed a meta-analysis comparing Y-chromosome and mtDNA markers from other regions of America populations and have found the same evidence (Table 2). In other words, independent of the region the genetic differentiation between populations is always higher for Y-chromosome markers than for mtDNA. The trend showed in the Table is statistically significant if we compare both columns using paired Wilcoxon test (Z(U)=2.2; p=0.028). Interestingly, the highest differences could be observed among the Tupians.

As pointed previously, development of agriculture is highly associated with both population size expansion and dispersion in many regions of the world, including that of Tupí and Arawak linguistic groups (Epps, 2009), in addition to patrilocality (Wilkins; Morlowe, 2006). The integrated view of meta-analysis data and literature review strongly supports the idea that Tupí expansion dynamics was associated with patrilocality practices. Since ethnohistorical and genetic data show contact between Tupí and non-Tupí groups (Aguiar, 1991), it is plausible to suppose that preferential female exchange would not be restricted to Tupí tribes.

CONCLUDING REMARKS

The present study shows a comprehensive set of genetic evidence from meta-analysis data and literature review,

Table 2. F_{ST}-based estimates in Y-chromosome and mtDNA markers for tupians and several regions of America.

Region	Y-STR	mtDNA
America[6]	0.390	0.256
Mesoamerica[6]	0.506	0.144
Northwest South America[6]	0.402	0.300
Andes[6]	0.225	0.160
East South America[6]	0.585	0.548
French Guiana Tupi[7]	0.651	0.177
Amazonian Tupi[8]	0.787	0.254

leading to a proposed scenario (Figure 4) where MGR is Tupí homeland. Moreover, in this scenario, Tupí groups from MGR and outside MGR showed different patterns of demographic dynamics, which are in accordance with the hypothesis of an ancient Tupí expansion in MGR, followed by dispersion of Tupí-Guarani branch to other South America regions. This dispersion was accompanied by sequential founder effect bottleneck events. Furthermore, depopulation could be detected in both regions but in different periods (ancient and recent). Finally, the dispersion seems to be related to patrilocality practices. This genetic model of dispersion dynamics may have an important impact in the interpretation of archeological and linguistic data. If some technologies and their products are more associated to the female gender, like ceramic and other artisanal skills, these technologies would be shared more extensively among different populations than those not associated with females, because they would be carried along with the female migration. This hypothesis might be tested with a reappraisal of archeological materials. The same trend would be also observed in linguistics, because the borrowing of words related to these female-associated technologies might also be more extensively observed.

---

6   Yang *et al.*, 2010.
7   Mazieres *et al.*, 2011.
8   Present Study. Data from four Y-chromosome STR and mtDNA (Tarazona-Santos *et al.*, 2001; Palha *et al.*, 2010; Ramallo *et al.*, 2013). F_{ST} estimated from eight Tupi populations: Zoró, Suruí, Gavião, Wayampi, Zoé, Urubú-Kaapór, Awá-Guajá and Parakanã. Estimates of both columns were compared by paired Wilcoxon test (Z(U)=2.2; p=0.028).
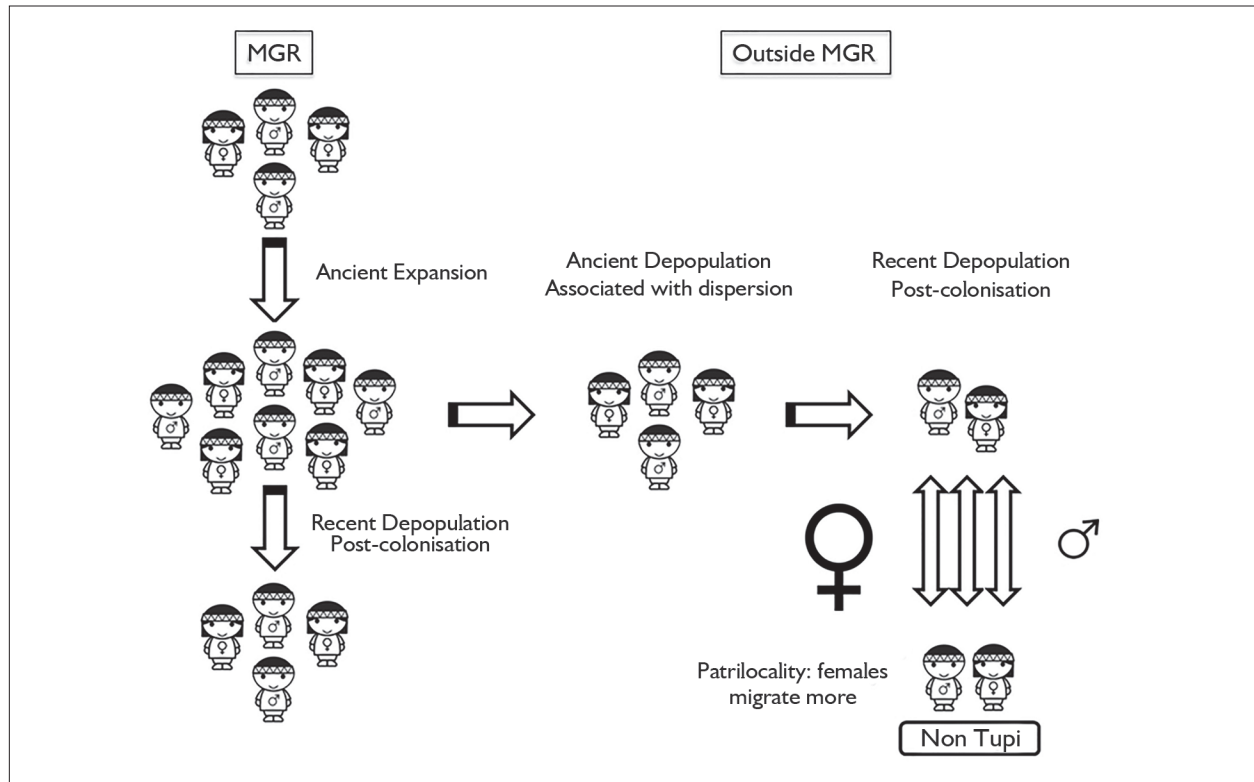
Figure 4. Schema representing the putative dynamics of Tupí dispersion. MRG is proposed as homeland. After an ancient population expansion, Tupí dispersed outside MGR, experiencing sequential founder effects and/or depopulation. Afterward, signatures of recent depopulation could be detected in both MGR and outside MGR Tupí populations.

Regarding the statistical analysis, it should be considered that more robust tests were not applied in the present study as most of the data was gathered from several studies, not allowing for more sophisticated approaches, because more detailed data, like locus estimates of variability and their variances, are not available. Most importantly, many different and independent genetic evidences agree with the main conclusions about origin, demographic sequence of events and patrilocality cultural practices. Additionally, the main conclusions are open to future testing by archeological and linguistic approaches.

ACKNOWLEDGMENTS

REFERENCES

AGUIAR, G. F. S. 1991. Ethnohistory, intertribal relationships and genetic diversity among Amazonian Indians. **Human Biology** 63:743-762.

AMORIM, C. E. G., BISSO-MACHADO, R., RAMALLO, V., BORTOLINI, M. C., BONATTO, S.L., SALZANO, F. M., HUNEMEIER, T. 2013. A Bayesian Approach to Genome/Linguistic Relationships in Native South Americans. **Plos One** 8 (5): 1-8.

BATTILANA, J., FAGUNDES, N.J., HELLER, A.H., GOLDANI, A., FREITAS, L.B., TARAZONA-SANTOS, E., MUNKHBAT, B., MUNKHTUVSHIN, N., KRYLOV, M., BENEVOLENSKAIA, L., ARNETT, F.C., BATZER, M.A., DEININGER, P.L., SALZANO, F.M., BONATTO, S.L. 2006. Alu insertion polymorphisms in Native Americans and related Asian populations. **Annals of Human Biology** 33(2):142-160.

BONATTO, S.L., SALZANO, F.M. Diversity and age of the four major mtDNA haplogroups, and their implications for the peopling of the New World. **American Journal of Human Genetics**, v. 61, n. 6, p. 1413-1423. 1997.

BORTOLINI, M.-C., SALZANO, F. M., BAU, C. H. D., LAYRISSE, Z., PETZL-ERLER, M. L., TSUNETO, L. T., HILL, K., HURTADO, A.M., CASTRO-DE-GUERRA, D., BEDOYA, G., RUIZ-LINARES, A. Y-chromosome biallelic polymorphisms and Native American population structure. **Annals of Human Genetics**, v. 66, n. 4, p. 255-259. 2002.

CALLEGARI-JACQUES, S.M., TARAZONA-SANTOS, E.M., GILMAN, R.H., HERRERA, P., CABRERA, L., DOS SANTOS, S.E., MORÉS, L., HUTZ, M.H., SALZANO, F.M. Autosome STR in native South America-Testing models of association with geography and language. **American Journal of Physical Anthropology**, 145(3):371-381. 2011.

CHAKRABORTY, R., SMOUSE, P.E., NEEL, J.V. Population amalgamation and genetic variation: observations on artificially agglomerated tribal populations of Central and South America. **American Journal of Human Genetics**, v. 43, n. 5, p. 709-25. 1988.

CORNUET, J.M. & LUIKART, G., Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. **Genetics**, 144:2001-2014. 1997.

EPPS, P. Language Classification, Language Contact, and Amazonian Prehistory. **Language and Linguistics Compass**, v. 3, n. 2, p. 581-606. 2009.

EWERTON, P.D. **História demográfica de Sulameríndios inferida a partir de marcadores mtDNA e microssatélites conjuntamente com simulações por coalescência e o uso de Redes Neurais Artificiais**. 2011, 119 f. Dissertação (Mestrado em genética e biologia molecular) – Fundação Amazônia Paraense de Amparo à pesquisa, Universidade Federal do Pará, Belém, 2011.

FAGUNDES, N. J. R., BONATTO, S. L., CALLEGARI-JACQUES, S. M., & F.M. SALZANO. Genetic, geographic, and linguistic variation among South American Indians: possible sex influence. **American Journal of Physical Anthropology**, v.117, n. 1, p. 68-78. 2002.

GABAS JÚNIOR, N. Tupian Languages. In: BROWN, K. (Ed): **Encyclopedia of Language & Linguistics**. Oxford: Elsevier, 2006. p. 146-150.

GARRIGAN, D.; HEDRICK, P. W. Perspective: detecting adaptive molecular polymorphism: lessons from the MHC. **Evolution International Journal of Organic Evolution**, v. 57, n. 8, p. 1707–1722. 2003.

GOICOECHEA, A.S.; CARNESE, F.R.; DEJEAN, C.; AVENA, S.A.; WEIMER, T.A.; FRANCO, M.H.; CALLEGARI-JACQUES, S.M.; ESTALOTE, A.C.; SIMÕES, M.L.; PALATNIK, M.; SALZANO, F.M. Genetic relationships between Amerindian populations of Argentina. **American Journal of Physical Anthropology**, v.115, n. 2, p.133-143. 2001.

GUERREIRO, J.F.; DOS SANTOS, E.J.M; DOS SANTOS, S.E. Effect of average heterozygosity on the genetic distance of several Indian tribes from the Amazon region. **Annals of Human Biology**, v. 21, n. 6, p. 589-595. 1994.

HECKENBERGER, M.J.; NEVES, E.G.; PETERSEN, J.B. De onde surgem os modelos? As origens e expansões Tupi na Amazônia Central. **Revista de Antropologia**, n. 41, p. 69-96. 1998.

HUNLEY, K.L.; SPENCE, J.E.; MERRIWETHER, D.A. The impact of group fissions on genetic structure in Native South America and implications for human evolution. **American Journal of Physical Anthropology**, v. 135, n. 2, p. 195-205. 2008.

JORDE, L. B.; WATKINS, W. S.; BAMSHAD, M. J.; DIXON, M. E.; RICKER, C. E.; SEIELSTAD, M. T.; BATZER, M. A. The Distribution of Human Genetic Diversity: A Comparison of Mitochondrial, Autosomal, and Y-Chromosome Data. **American Journal of Human Genetics**, v. 66, p. 979-988. 2000.

KING, J. P.; KIMMEL, M.; CHAKRABORTY, R. A Power Analysis of Microsatellite-Based Statistics for Inferring Past Population Growth. **Molecular Biology and Evolution**, v. 17, n. 12, p. 1859-1868. 2000.

MARRERO, A.R.; SILVA-JUNIOR, W.A.; BRAVI, C.M.; HUTZ, M.H.; PETZL-ERLER, M.L.; RUIZ-LINARES, A.; SALZANO, F.M.; M. C. BORTOLINI. Demographic and Evolutionary Trajectories of the Guarani and Kaingang Natives of Brazil. **American Journal of Physical Anthropology**, v. 132, p. 301-310. 2007.

MAZIERES, S.; CALLEGARI-JACQUES, S. M.; CROSSETTI, S. G.; DUGOUJON, J.-M.; LARROUY, G.; BOIS, E.; CRUBEZY, E.; HUTZ, M.H.; SALZANO, F.M. French Guiana Amerindian demographic history as revealed by autosomal and Y-chromosome STRs. **Annals of Human Biology**, v. 38, n.1, p. 76-83. 2011.

MESA, N. R.; MONDRAGÓN, M. C.; SOTO, I. D.; PARRA, M. V.; DUQUE, C.; ORTÍZ-BARRIENTOS, D.; GARCÍA, L. F.; VELEZ, I.D.; BRAVO, M.L.; MÚNERA, J.G.; BEDOYA, G.; BORTOLINI, M.C.; RUIZ-LINARES, A. Autosomal, mtDNA, and Y-chromosome diversity in Amerinds: pre- and post-Columbian patterns of gene flow in South America. **American Journal of Human Genetics**, v. 67, n. 5, p. 1277-1286. 2000.

MIGLIAZZA, E. C. Linguistic prehistory and the refuge model in Amazonia. In: PRANCE, G. T. (Ed). **Biological diversification in the tropics**. New York: Columbia University Press, 1982. p. 497-519.

MOORE D. Brazil: language situation. In: BROWN, K. (Ed). **Encyclopedia of languages and linguistics**. Amsterdam: Elsevier, 2005. p. 117–127

MULLIGAN, C.J.; HUNLEY, K.; COLE, S.; LONG, J.C. Population Genetics, History and Health Patterns in Native Americans. **Annual Review of Genomics Human Genetics**, v. 5, p. 295-315. 2004.

NEI, M; LIVSHITS, G. Genetic relationships of Europeans, Asians and Africans and the origin of modern Homo sapiens. **Human Heredity**, v. 39, n. 5, p. 276-281. 1989.

NEVES, W.A.; BERNARDO, D.V.; OKUMURA, M.; ALMEIDA, T.F.; STRAUSS, A.M. Origin and dispersion of the Tupiguarani: What does cranial morphology say? **Boletim do Museu Paraense Emílio Goeldi Ciências Humanas**, v. 6, n. 1, p. 95-122. 2011.

NOELLI, F.S. The Tupi: explaining origin and expansions in terms of archeology and of historical linguistics. **Antiquity**, v. 72, p. 648-663. 1998.

NOVICK, G.E.; NOVICK, C.C.; YUNIS, J.; YUNIS, E.; MARTINEZ, K.; DUNCAN, G.G.; TROUP, G.M.; DEININGER, P.L.; STONEKING, M.; BATZER, M.A.; HERRERA, R.J. Polymorphic human specific Alu insertions as markers for human identification. **Electrophoresis**, v. 16, n.9, p. 1596-1601. 1995.

PALHA, T.J.; RODRIGUES, E.M.; DOS SANTOS, S.E. Y-STR haplotypes of Native American populations from the Brazilian Amazon region. **Forensic Science International Genetics**, v. 4, n.5, p. 121-123. 2010.

RAMALLO, V.; B ISSO-MACHADO, R.; BRAVI, C.; COBLE, M.D.; SALZANO, F. M.; HUNEMEIER, T.; BORTOLINI, M. C. Demographic Expansions in South America: Enlightening a Complex Scenario with Genetic and Linguistic Data. **American Journal of Physical Anthropology**, v. 150, p. 453-463. 2013.

RODRIGUES, A. D. Classificação do tronco linguístico Tupi. **Revista de Antropologia**, v. 12, p. 99-104. 1964.

ROGERS, A.R.; L.B. JORDE. Ascertainment bias in estimates of average heterozygosity. **American Journal of Human Genetics**, v. 58, n. 5, p. 1033-1041. 1996.

SANTOS, E. J. M.; TAKESHITA, L. Y. C.; MAIA, M. H. T.; EWERTON, P. D. Evidências Genéticas da Dispersão Tupi Centrada em Rondônia Envolvendo Patrilocalidade. **Cuadernos del Instituto Nacional de Antropología y Pensamiento Latinoamericano.** 2013. *In press*.

SEIELSTAD, M.T.; MINCH, E.; CAVALLI-SFORZA, L.L. Genetic evidence for a higher female migration rate in humans. **Nature Genetics**, v. 20, n. 3, p. 278-280. 1998.

SHI, W.; AYUB, Q.; VERMEULEN, M.; SHAO, R.G.; ZUNIGA, S.; VAN DER GAAG, K.; DE KNIJFF, P.; KAYSER, M.; XUE, Y.; TYLER-SMITH, C. A worldwide survey of human male demographic history based on Y-SNP and Y-STR data from the HGDP-CEPH populations. **Molecular Biology and Evolution**, v. 27, n. 2, p. 385-393. 2010.

SHRIVER, M.D.; JIN, L.; FERRELL, R.E.; DEKA, R. Microsatellite data support an early population expansion in Africa. **Genome Research**, v. 7, n. 6, p. 586-591. 1997.

SILVA, D.F.L.; SANTOS, A. K. C. R. dos; SANTOS, S.E.B. Diversidade genética de populações humanas na Amazônia. In: I. VIEIRA; D.C. OREN; J.M. CARDOSO (Ed.). **Biodiversidade Biológica e Cultural da Amazônia**, Belém: Museu Goeldi. 2001, p. 167-194.

SLATKIN, M. Linkage disequilibrium in growing and stable populations. **Genetics**, v. 137, n. 1, p. 331-336. 1994.

SLATKIN, M. A measure of population subdivision based on microsatellite allele frequencies. **Genetics**, v. 139, n. 1, p. 457-462. 1995.

TAJIMA, F. The effect of change in population size on DNA polymorphism. **Genetics**, v. 123, n. 3, p. 597-601. 1989.

TARAZONA-SANTOS, E.; CARVALHO-SILVA, D.R.; PETTENER, D.; LUISELLI, D.; DE STEFANO, G.F.; LABARGA, C.M.; RICKARDS, O.; TYLER-SMITH, C.; PENA, S.D.; SANTOS, F.R. Genetic differentiation in South Amerindians is related to environmental and cultural diversity: evidence from the Y chromosome. **American Journal of Human Genetics**, v. 68, n. 6, p. 1485-1496. 2001.

TAUTZ, D. Notes on the definition and nomenclatures of tandemly repetitive DNA sequences. In: PENA, S.D.J.; CHAKRABORTY, R.; EPPLEN, J.T.; JEFFREYS, A.J., (Ed.). **DNA fingerprinting:** State of the science. Birkhäuser: Basel, 1993. p. 21-28.

URBAN, G. On the geographical origins and dispersion of Tupian languages. **Revista de Antropologia**, v. 39, p. 61-104. 1996.

URBAN, G. A história da cultura brasileira segundo as línguas nativas. In: CUNHA, M. C. da. (Ed.). **História dos Índios no Brasil**. São Paulo: Companhia das Letras. 1998, p. 87-102.

WALKER, R. S., WICHMANN, S., MAILUND, T., ATKISSON, C.J. Cultural Phylogenetics of the Tupi Language Family in Lowland South America. **Plos One**, v. 7, n. 4, p. 1-9. 2012.

WANG, S.; LEWIS, C.M.; JAKOBSSON, M.; RAMACHANDRAN, S.; RAY, N.; BEDOYA, G.; ROJAS, W.; PARRA, M.V.; MOLINA, J.A.; GALLO, C.; MAZZOTTI, G.; POLETTI, G.; HILL, K.; HURTADO, A.M.; LABUDA, D.; KLITZ, W.; BARRANTES, R.; BORTOLINI, M.C.; SALZANO, F.M.; PETZL-ERLER, M.L.; TSUNETO, L.T.; LOOP, E.; ROTHHAMMER, F.; EXCOFFIER, L.; FELDMAN, M.W.; ROSENBERG, N.A.; RUIZ-LINARES, A. Genetic variation and population structure in native Americans. **PLoS Genetics**, v. 3, n. 11, p. 185. 2007.

WILDER, J.A; KINGAN, S.B.; MOBASHER, Z.; PILKINGTON, M.M.; M.F. HAMMER. Global patterns of human mitochondrial DNA and Y-chromosome structure are not influenced by higher migration rates of females versus males. **Nature Genetics**, n. 36, n. 10, p. 1122-1125. 2004.

WILKINS, J.F & MARLOWE, F.W. Sex-biased migration in humans: what should we expect from genetic data? **Bioessays**, v. 28, n. 3, p. 290-300. 2006.

YANG, N.N.; MAZIÈRES, S.; BRAVI, C.; RAY, N.; WANG, S.; BURLEY, M.W.; BEDOYA, G.; ROJAS, W.; PARRA, M.V.; MOLINA, J.A.; GALLO, C.; POLETTI, G.; HILL, K.; HURTADO, A.M.; PETZL-ERLER, M.L.; TSUNETO, L.T.; KLITZ, W.; BARRANTES, R.; LLOP, E.; ROTHHAMMER, F.; LABUDA, D.; SALZANO, F.M.; BORTOLINI, M.C.; EXCOFFIER, L.; DUGOUJON, J.M.; RUIZ-LINARES, A. Contrasting patterns of nuclear and mtDNA diversity in Native American populations. **Annals Human Genetics**, v. 74, n. 6, p.525-538. 2010.

ZIMPEL NETO, C.A. **Na direção das periferias extremas da Amazônia: arqueologia na bacia do Rio Jiparaná, Rondônia**. Dissertação (Mestrado em Arqueologia) – Museu de Arqueologia e Etnologia, Universidade de São Paulo, São Paulo, 2009.