



Genomic insight into the origins and dispersal of the Brazilian coastal natives

Marcos Araújo Castro e Silva^a, Kelly Nunes^a, Renan Barbosa Lemes^a, Àlex Mas-Sandoval^{b,c}, Carlos Eduardo Guerra Amorim^d, Jose Eduardo Krieger^e, José Geraldo Mill^f, Francisco Mauro Salzano^{b,1}, Maria Cátira Bortolini^b, Alexandre da Costa Pereira^e, David Comas^c, and Tábita Hünemeier^{a,2} 

^aDepartamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, SP, Brazil 05508-090; ^bDepartamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil 91501-970; ^cInstitut de Biologia Evolutiva (CSIC-UPF), Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, 08003 Barcelona, Spain; ^dDepartment of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095; ^eInstituto do Coração, Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil 05403-000; and ^fDepartamento de Fisiologia, Universidade Federal do Espírito Santo, Espírito Santo, Brazil 29040-090

Edited by Anne C. Stone, Arizona State University, Tempe, AZ, and approved December 12, 2019 (received for review May 25, 2019)

In the 15th century, ~900,000 Native Americans, mostly Tupí speakers, lived on the Brazilian coast. By the end of the 18th century, the coastal native populations were declared extinct. The Tupí arrived on the east coast after leaving the Amazonian basin ~2,000 y before present; however, there is no consensus on how this migration occurred: toward the northern Amazon and then directly to the Atlantic coast, or heading south into the continent and then migrating to the coast. Here we leveraged genomic data from one of the last remaining putative representatives of the Tupí coastal branch, a small, admixed, self-reported Tupiniquim community, as well as data of a Guaraní Mbyá native population from Southern Brazil and of three other native populations from the Amazonian region. We demonstrated that the Tupiniquim Native American ancestry is not related to any extant Brazilian Native American population already studied, and thus they could be considered the only living representatives of the extinct Tupí branch that used to settle the Atlantic Coast of Brazil. Furthermore, these data show evidence of a direct migration from Amazon to the Northeast Coast in pre-Columbian time, giving rise to the Tupí Coastal populations, and a single distinct migration southward that originated the Guaraní people from Brazil and Paraguay. This study elucidates the population dynamics and diversification of the Brazilian natives at a genomic level, which was made possible by recovering data from the Brazilian coastal population through the genomes of mestizo individuals.

Native Americans | peopling of South America | Tupí speakers | Brazilian natives | genetics

In the 15th century, the Brazilian coast was densely populated by Native American populations. At that time, a total of 3 million indigenous individuals lived in the territory currently corresponding to Brazil, with about a third inhabiting its coast (1). The conquest of the Brazilian territory by the Portuguese (circa 1500) led to a rapid decline of the coastal native populations, culminating in their extinction by the end of the 18th century (2). This massive depopulation completely changed the distribution of the Native American populations within Brazil, delimiting their territory to the Amazon region and the inland. At present there are just two small admixed communities self-reported as coastal Tupí (Tupiniquim and Tupinambá) living in Brazil; however, they do not speak any indigenous language.

When the Portuguese first arrived in South America, the Tupiniquim and Tupinambá, both originally Tupí speakers, were the dominant groups in the Brazilian Atlantic Coast (2). It is not clear how the Tupí speakers arrived on the east coast after they left the Amazonian basin. The origins of the Proto-Tupí (Amazonian, southern, and coastal Tupí ancestrals) dates back to possibly 5,000 y before present (YBP) in the Northwest Amazon (ref. 3 and references therein). More than 2,000 YBP, different Tupí populations expanded from this region over 4,000 km eastward and southward, respectively peopling the Atlantic coast and the western

Brazilian inland. They expanded to most of the South American lowlands during the late Holocene epoch, becoming one of the most populous and diverse linguistic families (with >35 languages still spoken). The Tupí expansion is comparable in importance to the Bantu expansion in Africa; however, relatively little is known about the event. There is no consensus in the literature regarding linguistic expansion models for the Tupí family (4, 5). Genetic studies based on uniparental markers are consistent with linguistic data indicating that northwestern Amazon was the center of diversification of the Tupí (3, 6), but they do not define any clear route of expansion, mainly due to lack of data from coastal populations. The causes of expansion are also unknown, and could have involved ecological adaptation or cultural issues (7). The Tupí-Guaraní branch (which includes coastal and southern Tupí groups) has assumed an expansionist character over the last 2,000 to 3,000 y, populating the Brazilian southwest, northeast, and entire coast, distinguishing them from the other Tupí speakers. On the basis primarily of archeological and linguistic evidence (2, 8), two main broad and contrasting hypotheses

Significance

The indigenous populations of the Brazilian coast were decimated by European conquerors and declared extinct by the 18th century. The disappearance of these populations created a gap in the understanding of South American settlement. The present study rescues the genome of an extinct coastal lineage of the Tupí branch through the examination of a small, admixed, self-reported Native American community. Our results suggest that genetic lineages representative of the Tupí peoples who inhabited the coast survived in this specific extant population. We also show the relationships among Coastal, Amazonian, and ancient Brazilian populations and elucidate the putative migratory routes used by Amazonian peoples between the Amazon and the Atlantic coast ~2,000 y ago.

Author contributions: T.H. designed research; M.A.C.e.S., D.C., and T.H. performed research; A.M.-S., J.E.K., J.G.M., F.M.S., M.C.B., A.d.C.P., and T.H. contributed new reagents/analytic tools; J.G.M. and F.M.S. collected the biological data; M.A.C.e.S., K.N., and R.B.L. analyzed data; and M.A.C.e.S. and T.H. wrote the paper with contributions from C.E.G.A., M.C.B., A.d.C.P., and D.C.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Published under the PNAS license.

Data deposition: The newly genotyped datasets reported in this paper have been deposited in the European Genome-phenome Archive and are available for download under accession no. [EGAS00001004036](https://www.ebi.ac.uk/ena/browser/view/EGAS00001004036).

¹Deceased September 28, 2018.

²To whom correspondence may be addressed. Email: hunemeier@usp.br.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1909075117/-DCSupplemental>.

First published January 13, 2020.

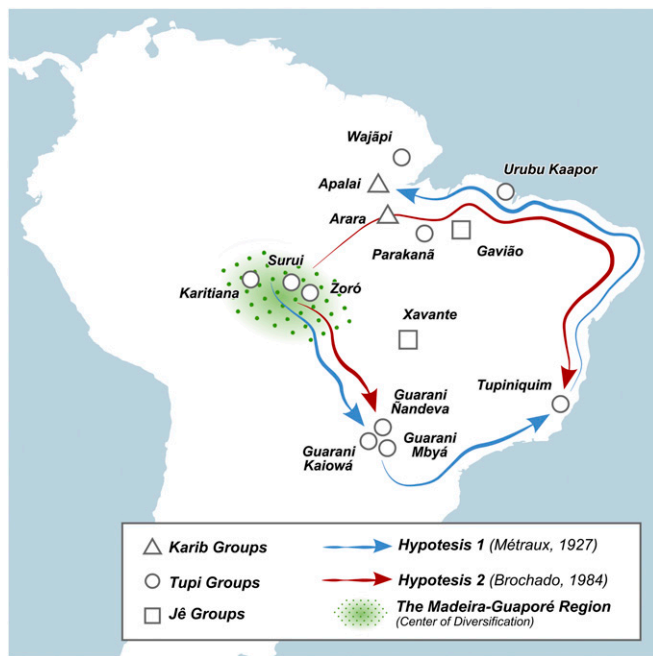


Fig. 1. Tupí Expansion hypotheses. Two main contrasting broad hypotheses can be recognized from literature (2, 8), which try to explain the Tupí Expansion. In hypothesis 1, the coastal Tupí would have derived from Guaraní populations in the south, which would have arrived there expanding southward from the Amazon Basin, here represented by the blue arrow. Conversely, hypothesis 2 postulates that the coastal Tupí and the Guaraní populations would have been originated in two separated expansions, with the former expanding eastward along the coast from the Amazon River mouth, the latter southward from the Amazon, here indicated by the two red arrows.

regarding the settlement of the Brazilian coast by the Tupí groups can be distinguished in the literature (Fig. 1). The first proposes that the Tupí from the Brazilian coast reached this region after coming from southwest Brazil, deriving from the same Tupí-Guaraní branch of Guaraní populations (9, 10) (blue arrow in Fig. 1). This hypothesis (10, 11) is based on archaeological data, linguistic analysis, and paleoenvironmental data, and associates the Tupí expansion with forest reductions that would have occurred during the Holocene. In this context, changes in vegetation would have forced these nonceramicist, preagriculturalist populations to seek new subsistence niches. Although these forest refuges were located both to the south and the east, linguistic data suggest that the most likely migration route to the Atlantic coast would have been through Brazil's western border, and then to the east shore. The alternative hypothesis assumes that one branch of Tupí moved first eastward, reaching the coast, and then southward along the coast, originating the coastal Tupí, whereas the other branch went southward, originating the Guaraní people (12) (red arrows in Fig. 1). According to this interpretation, the Proto-Tupí were already agriculturalists and ceramists, and the reason for their expansion was likely the demographic pressure caused by a continuous increase in population, which forced them to disperse in search of new lands to cultivate. This proposition (12) is motivated by the independent mode and evolution of Guaraní and Tupinambá potteries from the Amazonian Polychrome Tradition of Proto-Tupí speakers (characterized by the use of red and black paint on a white engobe). Tupinambá pottery is only found in the northeast Amazon and along the Brazilian coast to the Tropic of Capricorn, while Guaraní pottery has been found from southern Amazon to northern Argentina, Paraguay, and southern Brazil.

To reconstruct the history of the Tupí, we generated genomic data for the last remaining putative representatives of the Tupí

coastal branch, a small admixed self-reported community of Tupiniquim people; for a Guaraní native population from Southern Brazil; and for three native populations from the Amazonian region. We investigated their genetic origins and demonstrated that the Tupiniquim Native American ancestry is not related to any extant Brazilian Native American population for which genetic data have been generated to date. Therefore, we infer that the Tupiniquim are the only living representatives of this extinct Tupí branch that was settled along the Brazilian Atlantic Coast at the arrival of the Europeans. Leveraging genomic information of the coastal Tupí branch retrieved from these admixed individuals, we elucidated the pre-Columbian dispersion of the Tupí-stock from the Amazon to Southern Brazil and to the coast, finding evidence of two migrations: there was a direct migration from the Amazon to the coast, which originated the Tupí coastal populations, and a single distinct migration to south that originated the Guaraní people from Brazil and Paraguay. We further showed the existence of genetic continuity within Brazil when comparing ancient and modern individuals. The intensity of this continuity changed when the linguistic groups split and became structured, around 6,000 YBP, producing specific patterns of shared ancestry.

Results and Discussion

Overview of the Data. We generated data for more than 600,000 SNPs in 102 Native Americans from Brazil (47 Tupiniquim, 48 Guaraní Mbyá, 2 Wajãpi, 3 Parakanã and 2 Gavião; *SI Appendix, Table S1*). The following public datasets were also used in the analysis: 48 Native Americans (13), Human Genome Diversity Project dataset 11 (<http://www.cephb.fr/hgdp/>), 1000 Genomes Project (<https://www.internationalgenome.org/>), Anzick-1 (14), and 15 ancient DNA samples from Brazil (15) (*SI Appendix, Table S2*). *SI Appendix, Fig. S1* shows the geographic position of the analyzed samples.

Postcontact History. We performed Global Ancestry Inferences of the Tupiniquim and Guaraní Mbyá with ADMIXTURE (16). The Tupiniquim community exhibits a greater proportion of European and African (25.9% and 22.54%, respectively) admixture (*SI Appendix, Fig. S2A* and *Dataset S1*) in comparison with the Guaraní Mbyá (15.62% and 7.1%) (*SI Appendix, Fig. S2B* and *Dataset S1*). The Tupiniquim presented higher Native American ancestry (51.55%) when compared with the general Brazilian population (~7%; refs. 17 and 18). The Wajãpi, Parakanã, and Gavião populations presented no evidence of admixture with Africans and Europeans (*SI Appendix, Figs. S3–S5*).

To establish a timeline for these admixture events, we analyzed the decay of linkage disequilibrium between markers with Rolloff (19). Using the Guaraní Mbyá in addition to the Iberian and Yoruba populations from the 1000 Genomes Project as Tupiniquim parental populations, the last intense gene flow between the Native American and the European components was dated to seven generations ago (*SI Appendix, Fig. S6A* and *Dataset S3*), and between the Native American and African components to 5.5 generations ago (*SI Appendix, Fig. S6B* and *Dataset S3*), which is also approximately the same date estimated for the African and European components (*SI Appendix, Fig. S6C* and *Dataset S3*). We also performed an analysis of time and admixture dynamics inferred from TRACTS (20, 21). The results suggest that the admixture process in the Tupiniquim population was complex and continuous, involving two pulses of admixture followed by a continuous migratory flow (*SI Appendix, Fig. S7* and *Dataset S4*). The results indicated the beginning of the process of admixture with Europeans ~11.2 generations ago and with Africans ~8.3 generations ago. This initial admixing was followed by a second major pulse that started ~5.2 generations ago and with a continuous flow of Africans and Europeans for subsequent generations. Eleven generations ago (the time of the first European pulse) coincides with the height of the Brazilian Gold Cycle (1690–1750);

ref. 22), a period in which the Brazilian population increased from 300,000 to almost 3 million (1), forcing King João V of Portugal to restrict free Portuguese access to Brazilian lands. The exploitation of gold led to the enslavement of the indigenous people by the Portuguese to work in the mines (23), which decimated a great part of the coastal and central Brazilian native populations. Interestingly, also approximately eight generations ago (1807; time of the first African pulse), the Portuguese Royal family moved the court to Rio de Janeiro to escape the invasion of the Kingdom of Portugal by Napoleon Bonaparte (24), which quickly intensified colonization of the Brazilian coast and promoted rapid population growth. The transfer of the Portuguese Court to Brazil also intensified the slave trade, and between 1806 and 1830 alone, more than 850,000 Africans were forcibly brought to Brazil (1). Approximately five generations ago (1888; the time of the second pulse of admixture with Africans and Europeans), slavery was abolished in Brazil (24), which resulted in increases in the African-derived populations in all regions. During the same period, a massive migration of ~1.5 million Italians to the Brazilian southeastern region was encouraged by the government (1) to replace slave labor.

We also analyzed the distribution of runs of homozygosity (ROH) in the Tupiniqum in comparison with modern Native Americans (newly genotyped Guaraní Mbyá, Wajãpi, Parakanã, and Gavião; ref. 13), Africans, Europeans, and East Asians (HGDP). The distribution of ROH reflects demographic processes and mating patterns occurring throughout the population's history, since the longer tracts represent recent events, such as inbreeding, while shorter segments were formed by older demographic processes (e.g., bottlenecks and founder effects). Our results (Fig. 2 and *SI Appendix, Fig. S8*) showed that Amazonian and non-Amazonian Native Americans present, on average, larger amounts of short/intermediate ROH (0.5 to 8 Mb), while Tupiniqum exhibits an ROH pattern similar to that observed for Mesoamericans,

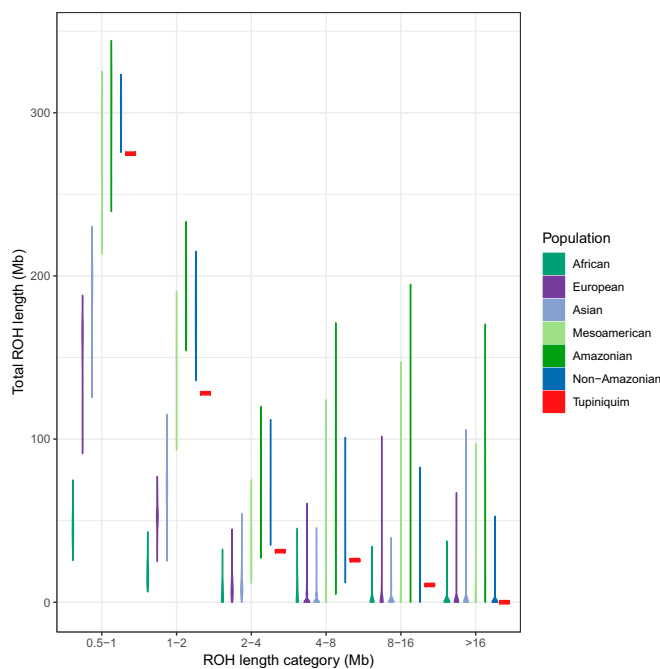


Fig. 2. ROH distribution of Native American, Asian, European, and African individuals. ROH identification was performed using PLINK v1.9 software (35) on a set of 395,840 SNP markers obtained from 609 individuals including the Tupiniqum, other modern Native Americans [newly genotyped and public (8)], and African, European, and Asian populations (Human Genome Diversity Project). The average total ROH lengths obtained per population are presented, binned by the ROH length.

with a higher amount of genetic diversity. This could be a result of the greater effective population size of the Tupiniqum population during the time of the Conquest, estimated at ~90,000 individuals living along the Brazilian coast (1).

To refine the estimation of the effective population size (N_e), we used a method recently introduced by Browning et al. (25), based on identity by descent (IBD) and local ancestry, to estimate the effective population size histories of the analyzed samples. The ancestry-specific (African, European, and Native American), historical N_e estimates are somewhat similar to and within the variation observed for Caribbean and Central American admixed populations (25), although the bottleneck seems to have been stronger in the Tupiniqum, with a minimum N_e of $\sim 10^2$ (Fig. 3), while the same estimate for most of the former populations is at least an order of magnitude higher. Interestingly, the minimum value of N_e in the Tupiniqum was caused by a bottleneck 7 generations ago, which coincides with the estimated date for the admixture between the Native American and European components (*SI Appendix, Fig. S6A* and *Dataset S3*). There are few historical data available regarding the Tupiniqum census; however, the last count was 55 individuals in 1876 (26), which is ~7 generations ago, so the genetic data seem to recover the collapse time of this population.

Precontact History. Individuals from the Tupiniqum population showed large variance in proportions of admixture (*SI Appendix, Fig. S2C* and *Dataset S1*), with one of them being almost entirely of Native American descent, whereas others showed less than 5% Native American ancestry. We leveraged this feature in each individual to look in more detail at the Native American ancestry component of the admixed genome and make inferences about deeper timescales in two ways: 1) by performing a Local Ancestry Inference on every Tupiniqum and masking the non-Native American markers as missing data, after which the individuals presented no or negligible evidence of non-Native American ancestry (*SI Appendix, Fig. S9*), and selecting individuals with more than 80% estimated Native American ancestry; and 2) by using exclusively the individual with the highest proportion of Native American ancestry (94.06%) and performing genotyping with high-density SNPs array Axiom Human Origins (19); yielding 1) ~70,000 and 2) ~600,000 SNPs overlapping those of the reference panels used (*SI Appendix*).

Principal component analysis (17) of the Native American populations clustered the Tupiniqum with Amazonian Tupi-Guaraní populations (the Parakanã, Urubu Kaapor, and Wajãpi), as well as with Karib speakers (the Apalai and Arara; *SI Appendix, Fig. S32*). However, principal component analysis of the Tupi populations placed the Tupiniqum next to the Parakanã and Urubu Kaapor (*SI Appendix, Fig. S32*), providing evidence for closer genetic relationships between these groups.

We then used the F_3 -statistics, as implemented by AdmixTools (19), to investigate the Tupiniqum Native American ancestry component and its relationship with other modern Native American populations. We calculated the F_3 -statistics in the form F_3 (Tupiniqum, Y, Z) for every pair (Y and Z) of modern Native American populations, and found no evidence of admixture between Tupiniqum and other Native American populations (*SI Appendix, Fig. S10A* and *B*), as no significant negative F_3 estimate was observed. Furthermore, using Treemix (27), Maximum Likelihood trees were inferred for all Native American, and separately for all Tupi, populations. Then we allowed the algorithm to fit up to 5 gene flow events between branches of the trees (*SI Appendix*); no gene flow was detected from any Native American population toward the Tupiniqum branch (*SI Appendix, Fig. S11A-D*).

Considering then the absence of admixture, we investigated the patterns of ancestral allele sharing among these groups. To assess this, the outgroup- F_3 was calculated in the form F_3 (Mbuti Pygmy; Y, Z) for every pair (Y and Z) of modern Native American populations, and the estimated F_3 values were plotted

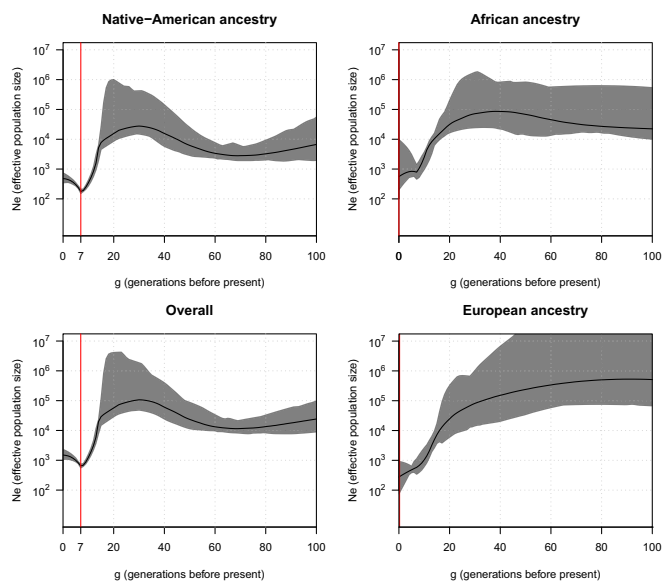


Fig. 3. Ancestry-specific effective population size (N_e) history estimated for the Tupiniquim. Related Tupiniquim samples ($k < 0.0625$) were removed from the data, which was then phased with Beagle v.5 (32). On the basis of the phased data, IBD segments were estimated with RefinedIBD (33) and Local Ancestry Inference with RFMix (34). Using IBDNe (25) ancestry specific and overall N_e were estimated. The ancestry-specific N_e values are coded in the y axis and indicated by the line for each generation before present depicted in the x axis. The gray areas show a 95% bootstrap confidence interval. Results for Native American, African, and European ancestry are shown in different panels, along with the N_e estimates obtained using all IBD segments (Overall). The red line indicates the generation with the minimum estimated value of N_e .

as heat map points on the map of the American continent (*SI Appendix, Fig. S12*). The Tupiniquim share more alleles with South Native Americans; however, they display no linguistic, geographic, or any other specific pattern of allele sharing among the latter, and they are not genetically close to the Guaraní Mbyá, who are currently settled near them (*SI Appendix, Fig. S12A*). In contrast, Guaraní Mbyá are related to the other Guaraní groups (*SI Appendix, Fig. S12B*). This pattern of geographic-genomic relationships was also present in some populations located next to the Madeira-Guaporé region in the Amazon basin (*SI Appendix, Figs. S13 and S14*).

We also performed the outgroup- F_3 analysis to investigate the relationships between modern-day Native American populations and ancient individuals from various periods of time located in current Brazilian territory; namely, Lapa do Santo (9,600 YBP) and 3 Sambaquis (Brazilian coastal and fluvial shell mounds, which are cultural deposits with diverse sizes and stratigraphies, mainly composed of shells: Laranjal [6,700 YBP], Moraes [5,800 YBP], and Jabuticabeira [2,000-2,100 YBP]), along with the Anzick-1 (12,900 to 12,700 YBP; a Clovis Culture-associated sample). In each comparison, ancient Brazilians were more closely related to the modern Brazilian native populations than to the modern Mesoamerican natives (*SI Appendix, Figs. S15 and S16*). More specifically, the most ancient individuals (Lagoa Santa and Laranjal) seemed to be more broadly related to the modern Brazilian natives (*SI Appendix, Figs. S15 A and B and S16 B and C*), indicating that the genetic similarity between modern populations and these ancient populations is independent of linguistic affiliation or geographic location. Most interestingly, the same was observed for the Anzick-1 individual (*SI Appendix, Fig. S16A*), implying some level of a distinctive contribution of Anzick-1-related lineages to modern South Americans (at least for the Brazilian populations represented

here) when compared with Mesoamericans (Maya and Pima). Beginning around 5,800 YBP (Moraes), these patterns of shared ancestry between paleo and modern individuals become progressively more distinctive (i.e., specific to some populations, such as Xavante [Jê-speaker population] and Arara [Karib-speaker population; *SI Appendix, Figs. S15 C and D and S16 D and E*]), replicating the long-standing continuity inside South American regions described by Posth et al. (15). Here we also detected some level of genetic continuity inside South America between the modern and the most ancient South Americans (Lagoa Santa and Laranjal), as well as with the Anzick-1, given that they share significantly more alleles than are shared between modern Mesoamericans and ancient Americans (*SI Appendix, Figs. S17–S20*). These most recent paleo individuals, Moraes and Jabuticabeira, and particularly the latter, presented differential high affinity with some populations (mainly Jê and Karib groups), but showed less close relationships with the Tupí groups (*SI Appendix, Figs. S15 C and D and Fig. S16 D and E*).

The scenario that emerges is one of increasing differentiation between Native American populations since the initial settlement of South America. Posth et al. (15) provide evidence for the existence of different migrations to this continent and subsequent replacement of the initial populations to a large extent. We add to this model, specifically in the case of eastern South America, the idea of the effects of demographic movements that occurred after the linguistic split (i.e., Tupí-Jê split), which involved several fission-fusion events (nonrandom migration processes that affect the structure of hunter-gatherer populations; ref. 28 and references therein) that genetically differentiated modern native populations from each other over time (28); this is more pronounced in recent samples (i.e., Jabuticabeira ~2,000 YBP).

Furthermore, we examined the relationship between ancient Brazilian samples and modern populations to see if we could detect any patterns of specific shared ancestry among them. With this purpose, we calculated F_4 (Mbuti Pygmy, aDNA; Y, Z), with aDNA iterating over all groups of samples according to archeological sites and Y and Z over all modern populations. For comparisons involving Pima and Maya, virtually any modern native South American population exhibited higher levels of allele sharing with all ancient samples, including the Anzick-1 (*SI Appendix, Figs. S19 and S20*). A similar pattern was also present in the outgroup- F_3 results (*SI Appendix, Figs. S15 and S16*). In addition, analysis of variance (Dataset S5) and Tukey's honestly significant difference test demonstrated that these differences were significant (*SI Appendix, Figs. S17 and S18*). In general, Xavante (Jê-speaker population) is the population that shares the most alleles with ancient samples, whereas the Mesoamericans (Maya and Pima), along with Wajápi, Guaraní Mbyá, and Parakanã, are the populations that share the least alleles with ancient samples (*SI Appendix, Figs. S19 and S20*). This may indicate distinctive demographic processes acting in these southern Native American populations, such as higher genetic drift or a more complex demographic history involving differential gene flow among the populations.

The Tupí Expansion. Our results thus far suggest that the Tupiniquim Native American ancestry is of Tupí origin, and they therefore may be used as proxies for the Tupí populations extinct from the Brazilian coast in investigations into the process of their expansion toward the coast. Thus, we tried to shed light on this question of the Tupí expansion, using two approaches.

First, we used an unsupervised approach, in which no prior expectations about the Tupí population history would be assumed. In this sense, we tried to produce trees depicting the evolutionary relationship between all Tupí populations using three methods: pairwise F_{ST} and pairwise F_2 Neighbor Joining Trees and Treemix (27). Using qpGraph (19), we tested these trees (*SI Appendix, Figs. S21–S24*), and only two of them showed a good fit to the data

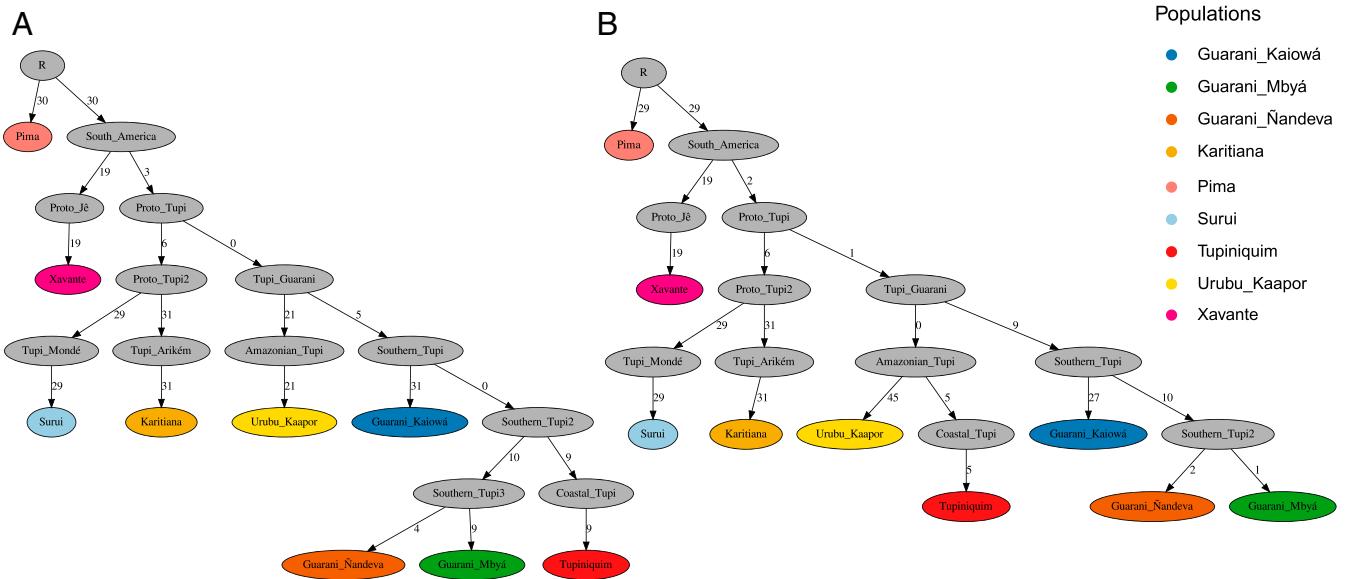


Fig. 4. Modeling Tupí expansion. The two main concurrent Tupí expansion hypotheses were modeled and assessed with qpGraph (19) to test the fit between all expected and observed F-statistics. A good fit to the data are indicated by absence of $|Z| > 3$ values. (A) Example of Tupí expansion hypothesis 1 model, with maximum $|Z|$ equal to 6.087. (B) Example of Tupí expansion hypothesis 2 model, where the maximum $|Z|$ is 2.763. Gray circles represent internal nodes of the tree for which there are no data, while colored circles stand for the modern Native American populations, in the same color scheme of other figures. The branch lengths are presented as units of FST, multiplied by 1,000.

(SI Appendix, Fig. S24 B and C). Both trees have very similar topologies, differing only in the relationships inside the Tupí-Guaraní, also presenting very small branches (values of 1 or less; SI Appendix, Fig. S24 B and C) inside this group, excepting the Guaraní populations, which are likely a monophyletic group (also evident in all other produced trees; SI Appendix, Figs. S21 and S22). This pattern of star-like radiation of the Tupí-Guaraní suggests that the common ancestor populations had relatively large effective population sizes and/or that the expansion happened in a short time. Thus, a polytomy appears in the root of Tupí-Guaraní populations, obscuring their relationships with one another, and with the Guaraní cluster being the only more easily discernible group. Interestingly, evidence for an excess of Native Mesoamerican-related ancestry was detected in the Guaraní, as a gene flow event from a Mesoamerican population to the Guaraní cluster was inferred with Treemix (27) (SI Appendix, Fig. S11 A and C). This pattern has been previously described between Andean and Native Mexican populations (29), but here we showed evidence of Mesoamerican gene introgression in a lowland population. This could indicate that the barrier between Andeans and non-Andeans was not so strict in the past and that the division observed in modern populations is likely related to both the establishment of agriculture in the highlands and strong drift in the lowlands.

Our second approach was to model population history scenarios (also with qpGraph) (19) to test the two main broad Tupí Expansion hypotheses (2, 8): hypothesis 1, the Tupí reaching the coast through a single expansion direction initially going through the south and then moving upward, deriving from the Guaraní people (9–11) (Fig. 1, blue arrow); and hypothesis 2, in which the Tupí occupied the coast after originally expanding eastward from the mouth of the Amazon and the Guaraní spread from the Amazon to the Paraná basin (12) (Fig. 1, red arrows). Essentially, we tried to differentiate between two hypotheses: 1) that the Tupiniquim would have reached the coast from the south and would, therefore, be genetically closer to the Guaraní populations, as they would share an exclusive most recent common ancestor with them; 2) that they would be more closely related to Tupí-Guaraní populations from the north (e.g., the Parakanã), based on the same rationale. Hence, we used Pima as an outgroup

for all South Native Americans, and Xavante (Jê-speaker population) as an outgroup for all Tupí. Several models were produced for each hypothesis, including different sets of populations and switching the Tupiniquim position (SI Appendix, Figs. S25–S28). Models of the second hypothesis consistently presented a better fit to the data in comparison with those from the first hypothesis, as inferred with qpGraph (19) (Fig. 4 A and B and SI Appendix, Figs. S25–S28). According to hypothesis 2, the Atlantic coast was peopled by Amazonian Tupí-speakers that probably reached this region ~2,000 YBP through a route along the northeast coast of Brazil (Fig. 1).

In this context, our results support the notion that expansion was caused by a search for new lands to cultivate by incipient Amazonian agriculturalists. Pottery found in the south Amazon (Guaraní Tradition) and the east Amazon and coast (Tupinambá Tradition) are separated from each other by as much as 4,000 km and present features that reveal the distinct evolution of these two groups after they left the Amazon basin.

Conclusion. Our study rescued part of the Native American history that had been concealed by European colonization. First, we recovered genomes from extinct coastal populations through the genomes of admixed people historically related to the Tupiniquim. Then, using this information of the Coastal Tupí populations, combined with data from natives of other regions, we managed to retrace how the occupation of the Brazilian territory by the Tupí occurred before the arrival of the Europeans. Notably, we reveal how the Atlantic coast was occupied by Amazonian peoples through a migratory wave from the northwest of the Amazon, and we further show that the Guaraní peoples of southern Brazil and Paraguay came from a separate migration but share a common ancestor. We also detected a subsequent migratory wave coming from Mesoamerica that may have influenced the formation of the southern Tupí groups (Guaraní branch). In addition, we found genomic evidence of the collapse of the coastal population, with an extreme bottleneck effect on the admixed Tupiniquim population. Last, when comparing modern and ancient individuals, we see that originating 6,000 y ago, there is genetic structure between populations that is most likely

generated by the strong drift events caused by the language diversification in South America.

Materials and Methods

To investigate the admixture events, we applied Roloff (19) and TRACTS software (20, 21). Using AdmixTools (19) we computed F_{ST} , F_{ST} , D -statistics, and F_{ST} , clustering the Tupiniquim as a population and treating them as separate individuals in the calculation in some of the analyses. For these analyses, datasets v [47 Tupiniquim (masked) + 48 Guaraní Mbyá + 48 Native Americans (13) + 7 newly genotyped Native Americans + HGDP], vi [1 Tupiniquim (ID: 2004) + 4 Guaraní Mbyá (IDs: 3001, 3036, 3038, 3051) + 48 Native Americans (13) + 7 newly genotyped Native Americans + HGDP], ix [47 Tupiniquim (masked) + 48 Guaraní Mbyá + 48 Native Americans (13) + 7 newly genotyped Native Americans + HGDP + 15 Ancient DNA samples (15)], and x [1 Tupiniquim (ID: 2004) + 4 Guaraní Mbyá (IDs: 3001, 3036, 3038, 3051) + 48 Native Americans (13) + 7 newly genotyped Native Americans + HGDP + 15 Ancient DNA samples (15) + Anzick-1 Clovis Culture associated ancient DNA (14); *SI Appendix, Table S3*] were used. We calculated F_{ST} and F_{ST} for all pairs of populations (*SI Appendix, Table S3*: datasets v and vi), to shed light on the relations between these populations and to pinpoint where the Tupiniquim fit within the Native American groups. Matrices containing pairwise genetic distance values were produced using R scripts (https://github.com/BenjaminPeter/cph_course/blob/master/scripts/analysis.R) and plotted as Neighbor-Joining trees using R packages *ape* and *ggtree* (30, 31) to provide models for the history of population splits between these populations. We also used Treemix (27) to estimate the Maximum Likelihood tree and fit putative admixture events. For a subset of populations that included all Tupí (*SI Appendix, Table S3*: datasets v and vi), we tested the fit between empirical data and the pairwise F_{ST} and F_{ST} NJ trees, along with the Maximum Likelihood trees produced with Treemix, using AdmixTools (19). Finally, we tried to explicitly model the two main Tupí Expansion hypotheses (2, 8), producing several models for each hypothesis with different populations, repositioning the Tupiniquim in the trees (again using datasets v and vi; *SI Appendix, Table S3*). Model fit was assessed by the differences between estimated and expected F-statistics values. Models with $|Z| < 3$ for all (or almost all) differences were considered to present a good fit to the data.

Ancestry-specific Effective Population Size (N_e) history was reconstructed for both the Tupiniquim and the Guaraní Mbyá (*SI Appendix, Table S3*; datasets vii [47 Tupiniquim (unmasked) + 48 Guaraní Mbyá + Sub-Saharan Africans, Europeans, and East Asians (1000 Genomes Project)] and xi [48

Guaraní Mbyá + Peruvians from Lima, Sub-Saharan Africans and Europeans (1000 Genomes Project)], respectively). First phasing was done with Beagle v.5 (32), and IBD segment estimation with RefinedIBD (33) and Local Ancestry Inference implemented with RFMix (34). Finally, IBDNe (25) was used to estimate ancestry-specific N_e from the estimated IBD segments and the ancestry blocks identified through the Local Ancestry Inference. ROH were identified using PLINK v1.9 (35) with a minimum length of 500 Kb, using a sliding window of 50 SNPs, a maximum gap of 100 Kb between consecutive SNPs, a proportion of 5% overlapping windows forming homozygous segments, and an SNP density of at least one per 50 Kb. A complete description of sampling, genotyping strategies, dataset assembly, quality control procedures, and methods is included in the *SI Appendix*.

Ethical approval for sample collection was provided by the Brazilian National Ethics Commission (CONEP Resolution no. 123 and 4599). CONEP also approved the oral consent procedure and the use of these samples in studies of population history and human evolution. Individual and/or tribal informed oral consents were obtained from participants who were not able to read or write. All sampling was coordinated by coauthors of this study (F.M.S. and J.G.M.) and their collaborators, in a manner consistent with the Helsinki Declaration and Brazilian laws and regulations applicable at the time of sampling. Logistical support for the sample collection was provided by the Fundação Nacional do Índio. The results of this study were discussed with the participating communities. A description of the sampling and genotyping strategies, along with the dataset assembly and quality control procedures is included in the *SI Appendix*.

Our dataset has been deposited at the European Genome-phenome Archive, which is hosted by the European Bioinformatics Institute (EBI) and the Centre for Genomic Regulation (CRG), under accession number EGAS00001004036. The informed consent associated with these samples is restricted to population history/evolutionary analyses. The data will be available to researchers who sign the Data Access Agreement with the Data Access Committee on the European Genome-phenome Archive website.

ACKNOWLEDGMENTS. We thank Rui Sérgio Sereni Murrieta and André Menezes Strauss for their helpful comments on the historical and archeological data. We are also grateful to Regina Célia Mingroni Netto and Lilian Kimura for laboratory assistance and technical support. Finally, we would like to thank all the native communities who participated in the study without whom this work would not have been possible. M.A.C.e.S was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (2018/013716; 2015/26875-9) and K.N was funded by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) (PNPD/1645581); NIH (R01 GM075091).

- IBGE, Instituto Brasileiro de Geografia e Estatística. <https://brasil500anos.ibge.gov.br/estatisticas-do-povoamento.html>. Accessed 1 December 2019.
- M. C. da Cunha, *História dos índios no Brasil* (Editora Companhia das Letras, 1992).
- V. Ramallo *et al.*, Demographic expansions in South America: Enlightening a complex scenario with genetic and linguistic data. *Am. J. Phys. Anthropol.* **150**, 453–463 (2013).
- R. S. Walker, S. Wichmann, T. Mailund, C. J. Atkinson, Cultural phylogenetics of the Tupi language family in lowland South America. *PLoS One* **7**, e35025 (2012).
- A. V. Galucio *et al.*, Genealogical relations and lexical distances within the Tupian linguistic family. *Bol. Mus. Para. Emílio Goeldi Ciênc. Hum.* **10**, 229–274 (2015).
- E. J. M. dos Santos, A. L. S. da Silva, P. D. Ewerton, L. Y. Takeshita, M. H. T. Maia, Origins and demographic dynamics of Tupi expansion: A genetic tale. *Bol. Mus. Para. Emílio Goeldi Ciênc. Hum.* **10**, 217–228 (2015).
- H. Silverman, W. Isbell, *Handbook of South American Archaeology* (Springer Science & Business Media, 2008).
- F. S. Noelli, "The Tupi expansion" in *The Handbook of South American Archaeology* (Springer, 2008), 659–670.
- A. Métraux, Migrations historiques des Tupi-Guarani. *J. Soc. Am.* **19**, 1–45 (1927).
- B. J. Meggers, C. Evans, A reconstrução da pré-história amazônica: Algumas considerações teóricas (1973).
- B. J. Meggers, C. Evans, "Lowland South America and the Antilles" in *Ancient Native Americans*, J. D. Jennings, ed. (W. H. Freeman and Company, CA: San Francisco, 1978), pp 543–591.
- J. P. Brochado, "An ecological model of the spread of pottery and agriculture into Eastern South America," PhD dissertation, University of Illinois at Urbana-Champaign (1984).
- P. Skoglund *et al.*, Genetic evidence for two founding populations of the Americas. *Nature* **525**, 104–108 (2015).
- M. Rasmussen *et al.*, The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* **506**, 225–229 (2014).
- C. Posth *et al.*, Reconstructing the deep population history of central and south America. *Cell* **175**, 1185–1197.e22 (2018).
- D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
- A. Ruiz-Linares *et al.*, Admixture in Latin America: Geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals. *PLoS Genet.* **10**, e1004572 (2014).
- F. S. G. Kehdy *et al.*, Brazilian EPIGEN Project Consortium, Origin and dynamics of admixture in Brazilians and its effect on the pattern of deleterious mutations. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 8696–8701 (2015).
- N. Patterson *et al.*, Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
- S. Gravel, Population genetics models of local ancestry. *Genetics* **191**, 607–619 (2012).
- S. Gravel *et al.*, 1000 Genomes Project, Reconstructing Native American migrations from whole-genome and whole-exome data. *PLoS Genet.* **9**, e1004023 (2013).
- T. E. Skidmore, Brazil: Five Centuries of Change. *OUP Catalogue* (2009). <https://ideas.repec.org/b/oxp/obooks/9780195374551.html>. Accessed 9 August 2019.
- H. Langfur, The return of the bandeira: Economic calamity, historical memory, and armed expeditions to the sertão in Minas Gerais, Brazil, 1750–1808. *Americas* **61**, 429–461 (2005).
- H. M. Starling, L. M. Schwarcz, *Brazil: A Biography* (Penguin, UK, 2018).
- S. R. Browning *et al.*, Ancestry-specific recent effective population size in the Americas. *PLoS Genet.* **14**, e1007385 (2018).
- B. Ricardo, F. Ricardo, *Povos indígenas no Brasil, 2011/2016* (Instituto Socioambiental, 2017).
- J. K. Pickrell, J. K. Pritchard, Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**, e1002967 (2012).
- F. M. Salzano, The fission-fusion concept. *Curr. Anthropol.* **50**, 959 (2009).
- G. A. Gnecci-Ruscione *et al.*, Dissecting the pre-Columbian genomic ancestry of native Americans along the Andes-Amazonia divide. *Mol. Biol. Evol.* **36**, 1254–1269 (2019).
- E. Paradis, K. Schliep, *ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics* **35**, 526–528 (2019).
- G. Yu, D. K. Smith, H. Zhu, Y. Guan, T. T.-Y. Lam, *ggtree: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol. Evol.* **8**, 28–36 (2017).
- B. L. Browning, Y. Zhou, S. R. Browning, A one-penny imputed genome from next-generation reference panels. *Am. J. Hum. Genet.* **103**, 338–348 (2018).
- B. L. Browning, S. R. Browning, Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics* **194**, 459–471 (2013).
- B. K. Maples, S. Gravel, E. E. Kenny, C. D. Bustamante, RFMix: A discriminative modeling approach for rapid and robust local-ancestry inference. *Am. J. Hum. Genet.* **93**, 278–288 (2013).
- C. C. Chang *et al.*, Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).