

## Maize dispersal patterns associated with different types of endosperm and migration of indigenous groups in lowland South America

Flaviane Malaquias Costa<sup>1,\*</sup>, Natalia Carolina de Almeida Silva<sup>2</sup>, Rafael Vidal<sup>3</sup>, Charles Roland Clement<sup>4</sup>, Fabio de Oliveira Freitas<sup>5</sup>, Alessandro Alves-Pereira<sup>6</sup>, César Daniel Petrolí<sup>7</sup>, Maria Imaculada Zucchi<sup>8</sup> and Elizabeth Ann Veasey<sup>1,\*</sup>

<sup>1</sup>Departamento de Genética, Universidade de São Paulo, Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, SP, 13418-900, Brazil, <sup>2</sup>Universidad Tecnológica del Uruguay, Durazno 97000, Uruguay, <sup>3</sup>Facultad de Agronomía, Universidad de la República, Montevideo, 12900, Uruguay, <sup>4</sup>Instituto Nacional de Pesquisas da Amazônia, Manaus, AM, 69067-375, Brazil, <sup>5</sup>Embrapa Recursos Genéticos e Biotecnologia, Brasília, DF, CEP 70770-901, Brazil, <sup>6</sup>Universidade Estadual de Campinas, Campinas, SP, 13083-875, Brazil, <sup>7</sup>Centro Internacional de Mejoramiento de Maíz y Trigo, Texcoco, 56237, México and <sup>8</sup>Secretaria de Agricultura e Abastecimento do Estado de São Paulo, Piracicaba, SP, 13400-900, Brazil

\* For correspondence. E-mail [flavianemcosta@hotmail.com](mailto:flavianemcosta@hotmail.com) or [eaveasey@usp.br](mailto:eaveasey@usp.br)

Received: 17 March 2022 Returned for revision: 20 September 2021 Editorial decision: 29 March 2022 Accepted: 31 March 2022  
Electronically published: 27 April 2022

- **Background and Aims** The lowlands of South America appear to be remarkably important in the evolutionary history of maize, due to new evidence that suggests that maize dispersed from Mexico and arrived in this region in a state of partial domestication. This study aimed to identify dispersal patterns of maize genetic diversity in this part of the continent.
- **Methods** A total of 170 maize accessions were characterized with 4398 single nucleotide polymorphisms (SNPs) and analysed to determine if maize dispersal was associated with types of endosperm and indigenous language families.
- **Key Results** Four genetic groups were identified in the discriminant analysis of principal components and five groups in the cluster analysis (neighbour-joining method). The groups were structured according to the predominance of endosperm types (popcorn, floury, flint/semi-flint). Spatial principal component analysis of genetic variation suggests different dispersal patterns for each endosperm type and can be associated with hypotheses of expansions of different indigenous groups.
- **Conclusions** From a possible origin in Southwestern Amazonia, different maize dispersal routes emerged: (1) towards Northern Amazonia, which continued towards the Caatinga and south-eastern Atlantic Forest (Floury); (2) towards Southern Brazil, passing through the Cerrado and Southern Atlantic Forest reaching the Pampa region (Floury); and (3) along the Atlantic Coast, following Tupi movements originating from two separate expansions: one (Tupinamba) from north to south, and the other (Guarani) in the opposite direction, from south to north (flint, floury and popcorn).

**Key Words:** Genetic diversity, genetic structure, landraces, SNP markers, *Zea mays* subsp. *mays*, diversification.

### INTRODUCTION

Maize (*Zea mays* subsp. *mays* L.) is a crop of considerable economic and socio-cultural importance worldwide (FAO, 2018). Recent research has shown that its domestication was more complex than previously thought (Vallebuena-Estrada *et al.*, 2016; Ramos-Madrigal *et al.*, 2016), and suggested that maize was dispersed from Mexico to lowland South America in a state of partial domestication (Kistler *et al.*, 2018, 2020). Domestication of maize started ~9000 years before the present (BP) in Southern Mexico from populations of *Zea mays* ssp. *parviglumis* (Matsuoka *et al.*, 2002; Piperno *et al.*, 2009; Moreno-Letelier *et al.*, 2020). Archaeological records show that maize had already arrived in South America at least ~7150 years BP in Ecuador (Stoehert, 1985; Pearsall and Piperno, 1990), ~6850 years BP in the Amazonian lowlands

of Bolivia (Lombardo *et al.*, 2020), ~6700 years BP in Peru (Grobman *et al.*, 2012), ~5300 years BP in Rondônia, Brazil (Hilbert *et al.*, 2017), and ~4190 years BP in Uruguay (Iriarte *et al.*, 2004). In South America, the dispersal of maize in the lowlands occurred independently of its dispersal throughout the Andes (Freitas *et al.*, 2003; Freitas and Bustamante, 2013). Numerous studies have identified two large genetic groups, the Andean group and the lowland group (Matsuoka *et al.*, 2002; Freitas *et al.*, 2003; Lia *et al.*, 2007; Vigouroux *et al.*, 2008; van Heerwaarden *et al.*, 2011; Freitas and Bustamante, 2013; Bedoya *et al.*, 2017). These studies suggested that maize races from lowland South America are more closely related to Central American races than to Andean races (Freitas *et al.*, 2003; Vigouroux *et al.*, 2008; van Heerwaarden *et al.*, 2011; Freitas and Bustamante, 2013; Bedoya *et al.*, 2017). However, the old maize races in Mexico are not directly associated with

races considered primitive in South America, which suggests a long temporal separation between the races in these regions (Bedoya *et al.*, 2017).

The dispersal of maize across the Americas is associated with many adaptive changes and occurred through exchanges among human groups throughout the continent (Brieger *et al.*, 1958). Lowland South America, which encompasses areas with altitudes below 1500 m (72 % of the South American continent), is considered a secondary centre of genetic diversity for maize (Brieger *et al.*, 1958; Paterniani and Goodman, 1977; Kistler *et al.*, 2018), within which Kistler *et al.* (2018) suggested Southwestern Amazonia as a secondary improvement centre for the partially domesticated crop. Indigenous landraces of lowland South America (*Caingang*, *Entrelaçado*, *Lenha*, *Moroti* and *Guarani Popcorns*) are considered to best represent pre-European-conquest maize (Brieger *et al.*, 1958; Paterniani and Goodman, 1977). The *Entrelaçado* landrace, typical of Amazonia, may have been widely distributed (Grobman *et al.*, 1961), but collections in Amazonia are scarce (Brieger *et al.*, 1958; Paterniani and Goodman, 1977; Costa *et al.*, 2021). The interlacing character of the ears of this race was mentioned as an ancestral character, which originated in Southwestern Amazonia and dispersed to Eastern Amazonia (Brieger *et al.*, 1958). *Entrelaçado* has floury endosperm and continues to be used by traditional farmers and indigenous peoples from pre-conquest times to the present day in Southwestern Amazonia (Costa *et al.*, 2021).

According to Goodman (1995), the different types of endosperm, their uses and the preferences of farmers are related to different stages of the maize domestication process. Some of the earliest maize kernels resemble those of the popcorn type, which suggests that the oldest maize was popcorn, with other endosperm types, such as flint, floury, dent and sweet, appearing later. Brieger *et al.* (1958) suggested that popcorn would correspond to the first and oldest stage of domestication, since popcorn has some 'wild type' characters, such as very small and stiff grains, which are often pointed. These characteristics would require stones for grinding or heat to roast or explode the grains. Flint endosperm could represent the second stage of domestication. The main difference between the flint and popcorn types would be the size and, consequently, the proportional increase in the volume of the inner core, containing soft endosperm within the stiff shell. Floury endosperm was suggested as the third stage of domestication and was of high value to indigenous people because it is easily ground and cooked. Dent endosperm would be a more modern maize type and has a good yield for animal feed, which was not important for indigenous people. It is believed that the different maize endosperms may have originated independently at different times and in different locations (Brieger *et al.*, 1958; Goodman, 1995; Staller *et al.*, 2006). Goodman (1995) presented the geographical distribution of the principal maize races and endosperm types in the Americas.

Studies of ancient DNA samples reveal that the *sugary1* (*su1*) allele related to the sweet endosperm type was dispersed from Mexico and fixed in the Southwestern USA (Jaenicke-Després *et al.*, 2003; Fonseca *et al.*, 2015). Maize's initial dispersal in this region was about 4000 years BP and is likely to have occurred along a highland route, followed by gene flow from a lowland coastal maize beginning at least 2000 years BP (Fonseca *et al.*, 2015). Other studies on the dispersal of different

maize endosperm types are scarce in the contemporary scientific literature. Bracco *et al.* (2016) evaluated the genetic structure of lowland maize landraces in America and identified two groups exclusive to lowland South America associated with the floury and popcorn endosperm types. However, their study did not present a representative sample involving the different biomes of the region and these genetic groups were restricted to Northeastern Argentina.

Based on linguistic data, maize became important to different prehistoric human groups at different times in the Americas (Brown *et al.*, 2014). In general, these findings agree with previous information regarding the time of domestication and routes of dispersal of the species (Matsuoka *et al.*, 2002; Piperno *et al.*, 2009; Vigouroux *et al.*, 2008; van Heerwaarden *et al.*, 2011; Bedoya *et al.*, 2017; Kistler *et al.*, 2018, 2020). In lowland South America, the linguistic reconstruction suggested that by 4400 years BP maize became important to the southern Arawak speakers, in what is today southern Peru and adjacent Brazil and Bolivia. Further east, in the upper Madeira River and upper Tapajós River basins, linguistic reconstructions suggest that speakers of other languages, including Tupi, considered maize to be important between 3000 and 2000 years BP. To speakers of southern Macro-Jê, maize became important about 2000 years BP, in what is today central and southern Brazil (Brown *et al.*, 2014). These reconstructions provide suggestions of when maize became important enough for a name to be handed down across generations, but maize arrived in these regions earlier, as shown by the archaeological finds.

The Arawak linguistic family is the largest in South America; it appears to have originated in Southwestern Amazonia and spread both into northern South America and the Caribbean, and into southern South America (Walker and Ribeiro, 2011; Aikhenvald, 2013). The Tupi linguistic family also originated in Southwestern Amazonia, and its territorial expansion was primarily within Brazil (Migliazza, 1982; Schmitz, 1991; Walker *et al.*, 2012; Aikhenvald, 2013), as also pointed out by other archaeological (Iriarte *et al.*, 2016), linguistic (Ramallo *et al.*, 2013) and genetic (Ramallo *et al.*, 2013; Castro e Silva *et al.*, 2020) studies. Maize became important to Tupi-Guarani speakers at least 2000 years BP (Brown *et al.*, 2014), and may have been dispersed with them in what is today the Brazilian territory. Like the Arawak expansion, the Tupi-Guarani represents one of the largest expansions of lowland South America, extending over 4000 km, from Southwestern Amazonia to the South American Atlantic coast (Migliazza, 1982; Schmitz, 1991; Walker *et al.*, 2012; Iriarte *et al.*, 2016). Genetic evidence suggested an origin of the Macro-Jê group in the Northern Cerrado, a region between the São Francisco and Tocantins rivers, which expanded towards Southern Brazil (Ramallo *et al.*, 2013; Castro e Silva *et al.*, 2020). Along the Atlantic Coast, there appear to have been two separate expansions, one from north to south by the Tupinamba, and the other in the opposite direction, from south to north by the Guarani (Iriarte *et al.*, 2016; Castro e Silva *et al.*, 2020).

Hypotheses of domestication and dispersal of maize have been addressed with different molecular markers, such as microsatellites (simple-sequence repeats, SSRs) (Matsuoka *et al.*, 2002; Vigouroux *et al.*, 2008; Mir *et al.*, 2013; Bracco *et al.*, 2016; Bedoya *et al.*, 2017), the alcohol dehydrogenase 2 (*Adh2*) gene (Freitas *et al.*, 2003; Freitas and Bustamante, 2013; Bustamante *et al.*, 2014) and single nucleotide polymorphisms (SNPs) (van

Heerwaarden *et al.*, 2011; Hufford *et al.*, 2013; Fonseca *et al.*, 2015; Kistler *et al.*, 2018, 2020; Moreno-Letelier *et al.*, 2020). The present study analysed the genetic diversity and structure of native landraces of maize from lowland South America in a larger sample, not explored in previous studies using SNP markers, aiming to understand the patterns of dispersal in this part of the continent, and to contribute to a better understanding of the history of maize diversification. The hypotheses considered in this study are associated with the Arawak, Tupi (especially the Tupi-Guarani) and Macro-Jê expansions at different times. From a possible origin in Southwestern Amazonia, we hypothesize that different maize dispersal routes emerged: one from Southern Amazonia towards Northern Amazonia, and the other from Southern Amazonia towards Southern Brazil, and later along the Atlantic Coast in both directions. We analysed the geographical distribution of genetic variation aiming to detect a greater genetic relationship between samples from the

regions considered in each dispersal event included in our hypotheses. In addition, we gathered historical and archaeological evidence of maize in lowland South America to support these hypotheses, providing a basis for comparison for future historical investigations and other genetic studies of local maize populations.

## MATERIAL AND METHODS

### Plant material

In total, 170 maize accessions were used in this study (Fig. 1.; Supplementary Data Table S1), of which 126 were from Brazil (78 obtained from the germplasm bank of Embrapa Milho e Sorgo; 48 from our recent collections), 42 from Uruguay (our recent collections) and two commercial hybrids for comparison purposes (H1: DKB390PRO and H2: DOW2B587).

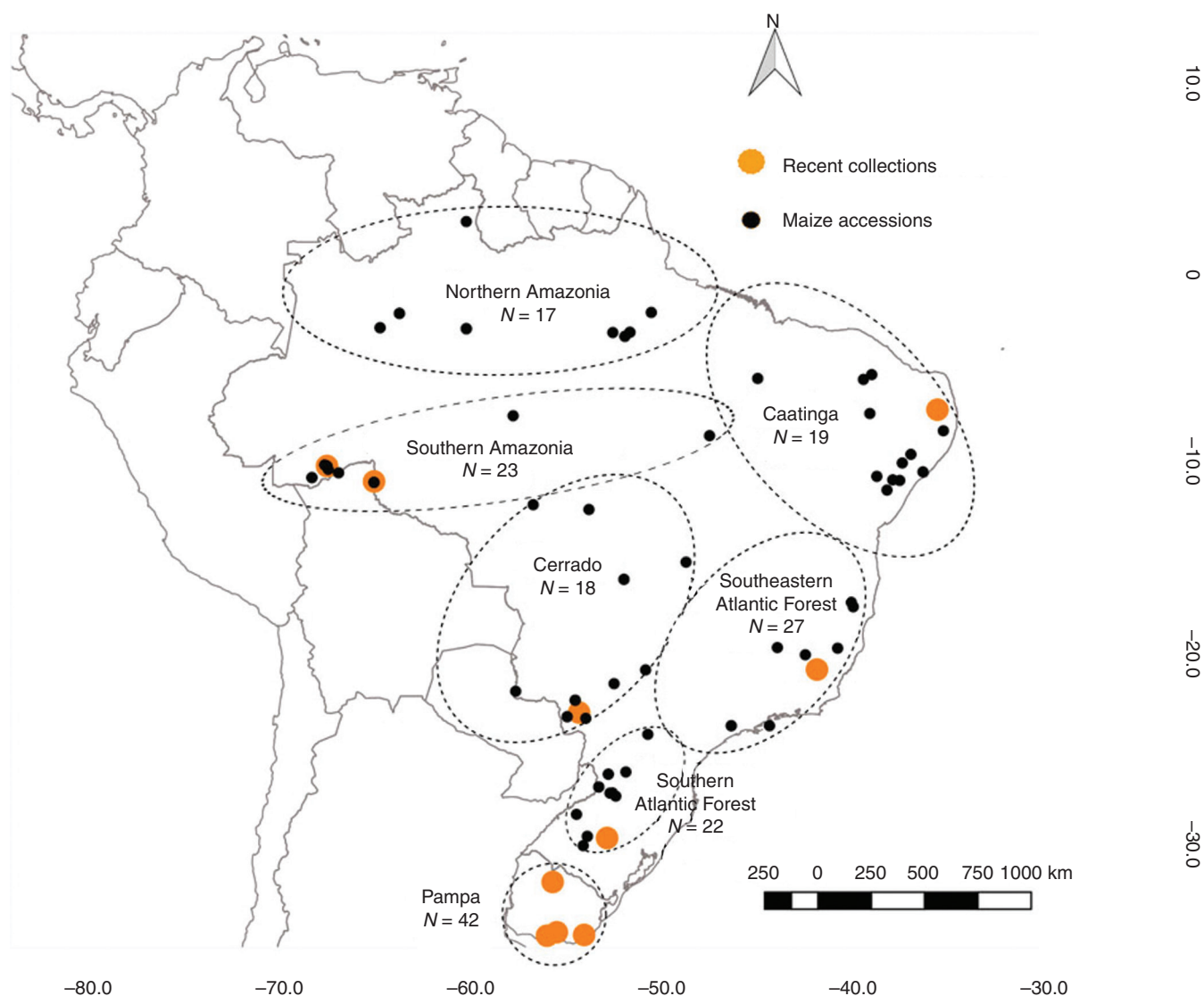


FIG. 1. Geographical distribution of the 170 maize accessions from lowland South America used in this study organized by region (dashed ellipses), including 90 new collections (orange) and 78 accessions from the Embrapa Milho e Sorgo germplasm bank (black). Molecular characterization also included two commercial hybrids. See Supplementary Data Table S1 for passport data.

The Embrapa accessions were selected from passport data that suggested that they are native races (Table S1) that represent the types of endosperm expected in lowland South America (flint, flourey and popcorn). Dent maize was not included in this study because it was introduced from North America in the 19th century (Paterniani and Goodman, 1977). An accession corresponds to a distinct, uniquely identifiable sample of seeds representing a cultivar or a population, which is maintained in storage for conservation and use (FAO, 2021). The term ‘maize races’ refers to groups of related populations with sufficient morphological characteristics in common to allow their recognition as a group (Anderson and Cutler, 1942), which occupy a defined geographical region (Bird and Goodman, 1977) and are associated with specific uses, many of which depend upon endosperm type.

Field collections were conducted in six states in Brazil and in five departments in Uruguay, including parts of five lowland South American biomes (Amazonia, Cerrado, Caatinga, Atlantic Forest and Pampa) (Supplementary Data Table S1). The accessions were classified into regional groups (Fig. 1; Table S1): Northern Amazonia, Southern Amazonia, Caatinga, Cerrado, Southeastern Atlantic Forest, Southern Atlantic Forest and Pampa.

In total, 133 accessions used in this study were characterized by type of endosperm, based on crop descriptors (IBPGR, 1991) for our recent collections, and by consulting the passport data of the accessions obtained from Embrapa’s germplasm bank (Supplementary Data Table S1): 39 flint and Semi-Flint, 40 Flourey, 44 popcorn, 10 Flourey – *Entrelaçado*. Thirty-seven accessions were not characterized because they did not have this information in their passport data.

This research was conducted between March 2016 and August 2020 and was approved by the Research Ethics Committee of ESALQ/USP, by the National Research Ethics Commission (CONEP), through CAAE process 60382016.2.0000.5395, the National System of Biodiversity Information (SisBio) (registration no. 61447-1), and was registered in the National System for the Management of Genetic Heritage and Associated Traditional Knowledge (SisGen) (registration no. AD2EF0B).

### Molecular characterization

Collected leaves were lyophilized and sent to the Genetic Analysis Service for Agriculture (SAGA) laboratory at the International Maize and Wheat Improvement Center (CIMMYT), located in Mexico, where DNA extraction and genotyping were performed. Genomic variance in maize is higher within than between populations, which is common in allogamous species (Dubreuil and Charcosset, 1998; Reyes-Valdés et al., 2013; Gouda et al., 2020). Therefore, a bulk sample of multiple individuals of a maize landrace can be used to estimate genetic diversity more accurately, rather than selecting and genotyping different individuals separately. Studies with SSR markers showed that a bulk of 30 individuals per accession represents an effective sample to estimate diversity parameters and is highly accurate for estimating genetic distances among the maize landraces, in addition to having a much lower cost (Reif et al., 2005; Reyes-Valdés et al., 2013). This sampling strategy has been evaluated and tested by the

SeeD project (<https://seedsofdiscovery.org/>) using SNPs–allele frequencies for the accessions belonging to the CIMMYT Maize Germplasm Bank (unpublished), as well as by Silva et al. (2020) and Arca et al. (2020, 2021) using SNP data and a bulk of 30 individuals per accession.

Thus, for this study the DNA was extracted from a bulk of 30 individuals per accession using a modified CTAB method (Doyle and Doyle, 1990); a fragment of fresh young leaf collected from each individual in the third week after planting was used in the bulk. Molecular characterization involved the use of DArTseq (Sansaloni et al., 2011; Al-Beyroutióv et al., 2016), a sequencing-based technology, which has been adopted by CIMMYT to genotype the accessions included in its international maize germplasm bank. This method uses a digestion process through a combination of two enzymes, a frequent cut (*PstI*) and another rare cut (*NspI*) enzyme; subsequently, primers were added for fragment amplification, as well as adaptors (barcodes) to allow the multiplexing of all samples in a single sequencing run. After a PCR (polymerase chain reaction) process, equimolar amounts of amplification products from each sample were pooled by plate and multiplexing in a flow cell, followed by sequencing of fragments using an Illumina HiSeq 2500 ([www.illumina.com](http://www.illumina.com)). SNP markers were identified *de novo* by comparing the sequences of fragments present in genomic representations (libraries) obtained in this assay and processed later in the analytical pipeline (DArTsoft14). Thus, SNP calling was done completely independently of any reference genome, which makes this strategy an excellent resource to evaluate the genetic diversity of any species (Sansaloni et al., 2020).

A detailed description of the sequence capture, SNP genotyping pipeline and its quality control can be found in Petroli and Kilian (2019). The FASTQ files were quality filtered using a Phred quality score of 30, which represents a 99.9 % base call accuracy. Also, the average of reads per marker (depth) was 1 893 270. A set of filtering parameters was applied to select high-quality markers for this study: Call Rate  $\geq 0.95$  (proportion of samples for which the corresponding marker information does not contain missing data); RepAvg (proportion of replicated technical assay pairs for which the marker score is consistent)  $\geq 0.85$ ; AvgPIC (average of the polymorphism information content of the reference and SNP alleles)  $> 0$ .

### Analytical methods

The selected SNPs are presented in Supplementary Data Table S2, in which the genotypic scoring represents allelic frequencies per accession. The genetic structure of the molecular data was assessed using discriminant analysis of principal components (DAPC) with adegenet (Jombart, 2008) for R (R Development Core Team, 2019). DAPC defines a number of clusters and presents the genetic structure so that variation is minimized within groups and maximized between groups (Jombart et al., 2010). The optimal number of groups was calculated using the *K*-means method and the Bayesian Information Criterion (BIC), without considering the number of groups a priori and not assuming that the groups meet the assumptions of Hardy–Weinberg equilibrium between loci, as

required by Structure (Pritchard *et al.*, 2000). The relationship between the groups identified by DAPC and the type of endosperm was analysed with Fisher's exact test (Fisher, 1935), at 5 % significance, using PAST v.4.0 (Hammer *et al.*, 2001). Cluster analysis with the neighbour-joining method (Saitou and Nei, 1987) was used to show the genetic relationships among 170 maize accessions, based on Nei's genetic distances (Nei, 1972), using ape (Paradis and Schliep, 2018) for R (R Development Core Team, 2019). The dendrogram was edited in FigTree v.1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree/>). For each  $k$ -means group used in DAPC, we performed a spatial principal component analysis (sPCA) (Jombart *et al.*, 2008) to investigate the spatial distributions of genetic variation, using adegenet (Jombart, 2008) for R (R Development Core Team, 2019). This analysis uses a matrix with allele frequencies and a spatial weighting matrix containing measurements of spatial proximity among entities based on a connection network to produce scores that summarize the spatial structure and the genetic variability across geographical space (Jombart *et al.*, 2008). The connection networks were built with the  $K$  nearest neighbours method considering  $K = 20$ , which allowed possible connections on a regional scale. The spatial structure is based on Moran's Index that generates global (genetic similarity is greater among samples from close regions) or local (genetic similarity is greater among samples from distant regions) patterns of spatial genetic structure (Jombart *et al.*, 2008). The significance of global and local structures was tested with 1000 permutations using the functions `global.rtest()` and `local.rtest()`, respectively. Each significant component was represented independently using squares in greyscale, in which the colour of squares is related to the sample's position on the specific component, using the function `s.value()`. Interpolation of the coordinates of each significant component was performed with the `interp()` function from the `akima` package (Akima and Gebhardt, 2021). In addition, for the synthetic projection of all significant principal components that were retained in each sPCA, we used the `colorplot()` function to summarize the spatial gradient of genetic diversity based on the RGB (red green blue) system.

The distribution of major linguistic families (Eriksen, 2011) was projected on the map to associate the indigenous groups with the maize groups, with the assumption that 400 years after European conquest there is still an association between geographical location of the sampling and the indigenous language from these locations. The accessions were plotted on the map using QGIS software (<http://qgisbrasil.org>) and were classified according to the language family corresponding to their geographical location (Supplementary Data Table S1). We emphasize that the frequencies of the observed language families of our maize accessions may present bias due to a greater number of collections carried out in specific areas.

The maize dispersal hypothesis map was prepared using QGIS software (<http://qgisbrasil.org>), demonstrating the associations of the genetic groups, linguistic data and archaeological records. The frequencies of the DAPC groups were projected on sector charts in the geographical regions, and the hypotheses of maize dispersal suggested by the neighbour-joining dendrogram and sPCA plots were projected on the map. The archaeological records were based on the literature listed in Table 1.

We also compared our data with those of Kistler *et al.* (2018, 2020), who used whole genome sequencing of maize samples to define five ancestral groups distributed throughout America, with the aim of analysing which endosperm type predominated in each ancestral group defined by the authors. Five accessions characterized by Kistler *et al.* (2018, 2020) were also included in our study (Supplementary Data Table S3), which allowed us to verify to which ancestral groups these samples were related to.

## RESULTS

### *Genetic diversity and structure of maize in lowland South America*

The molecular characterization resulted in 50 696 SNPs, among which 4398 were selected after a strict quality control process (Supplementary Data Table S2). The  $K$ -means method suggested four genetic groups that were clearly separated in DAPC

TABLE 1. Date (years before present – BP), region and type of sample in the archaeological records of maize of lowland South America.

Date (years BP)	Region of origin	Type of sample	Reference(s)
~7150	Las Vegas, Ecuador (Pacific Coast)	Phytoliths	Stohtert (1985), Pearsall and Piperno (1990)
~7000	Panamá (Pacific Coast)	Starch	Piperno <i>et al.</i> (2000)
~6850	Llanos de Moxos, Bolivia (Southwestern Amazonia)	Phytoliths	Lombardo <i>et al.</i> (2020)
~6700	Llanos de Moxos, Bolivia (Southwestern Amazonia)	Phytoliths	Lombardo <i>et al.</i> (2020)
~6700	Paredones and Huaca Prieta, Peru (Pacific Coast)	Ear, grain and straw	Grobman <i>et al.</i> (2012)
~6500	Rogaguado Lake, Bolivia (Southwestern Amazonia)	Pollen	Brugger <i>et al.</i> (2016)
~6320	Lake Sauce, Peru (Peruvian Amazonia)	Pollen	Bush <i>et al.</i> (2016)
~6000	Ayauch Lake, Ecuador (Western Amazonia)	Pollen and phytoliths	Bush <i>et al.</i> (1989)
~5760	Pará, Brazil (Northern Amazonia)	Pollen and phytoliths	Piperno (2011)
~5300	Ayauch Lake, Ecuador (Western Amazonia)	Phytoliths	Piperno (1990)
~5300	Upper Madeira River, Rondônia, Brazil (Southwestern Amazonia)	Phytoliths	Hilbert <i>et al.</i> (2017)
~4690 ± 40	Peña Roja and Abejas, Colombia (Western Amazonia)	Pollen	Piperno (2011)
~4645 ± 40	Araracuara, Colombia (Western Amazonia)	Pollen	Herrera <i>et al.</i> (1992)
~4190 ± 40	Los Ajos, Rocha, Uruguay (Southeast Uruguay)	Phytoliths and starch	Iriarte <i>et al.</i> (2004)
~3350	Pará, Brazil (Northern Amazonia)	Pollen	Bush <i>et al.</i> (2000)
~3000 to ~1500	Upper Madeira River (Southwestern Amazonia)	Phytoliths and starch	Watling <i>et al.</i> (2020)
~1390 ± 40	São Francisco do Sul, Santa Catarina, Brazil (Southern Brazil)	Starch	Wesolowski <i>et al.</i> (2010)
~990 ± 60	Januária, Northern Minas Gerais, Brazil (Southeast Brazil)	Ear, grain and straw	Freitas <i>et al.</i> (2003)

(Fig. 2A). The first two principal components explained 35.2 % of the total genetic variation.

The DAPC groups were associated with the type of endosperm (Table 2): Group G1 was dominated by popcorn; G2 by Flint/Semi-Flint; G3 by Flourey; and G4 by Flourey, 62.5 % of which belong to the *Entrelaçado* race (Costa et al., 2021). The two commercial hybrid cultivars were attributed to G2. Groups G1 and G4 were more homogeneous with respect to the type

of endosperm (96 % popcorn endosperm type and 94 % flourey endosperm type, respectively), while G2 and G3 were more heterogeneous. This degree of homogeneity may explain the isolation of these groups (G1 and G4) in relation to the others in the DAPC (Fig. 2A). The predominant biomes in each group suggest a connection between these regions (Fig. 2B). G1 was present in the Pampa, Southern Atlantic Forest, Southeastern Atlantic Forest and Cerrado; G2 in the Pampa, Southeastern

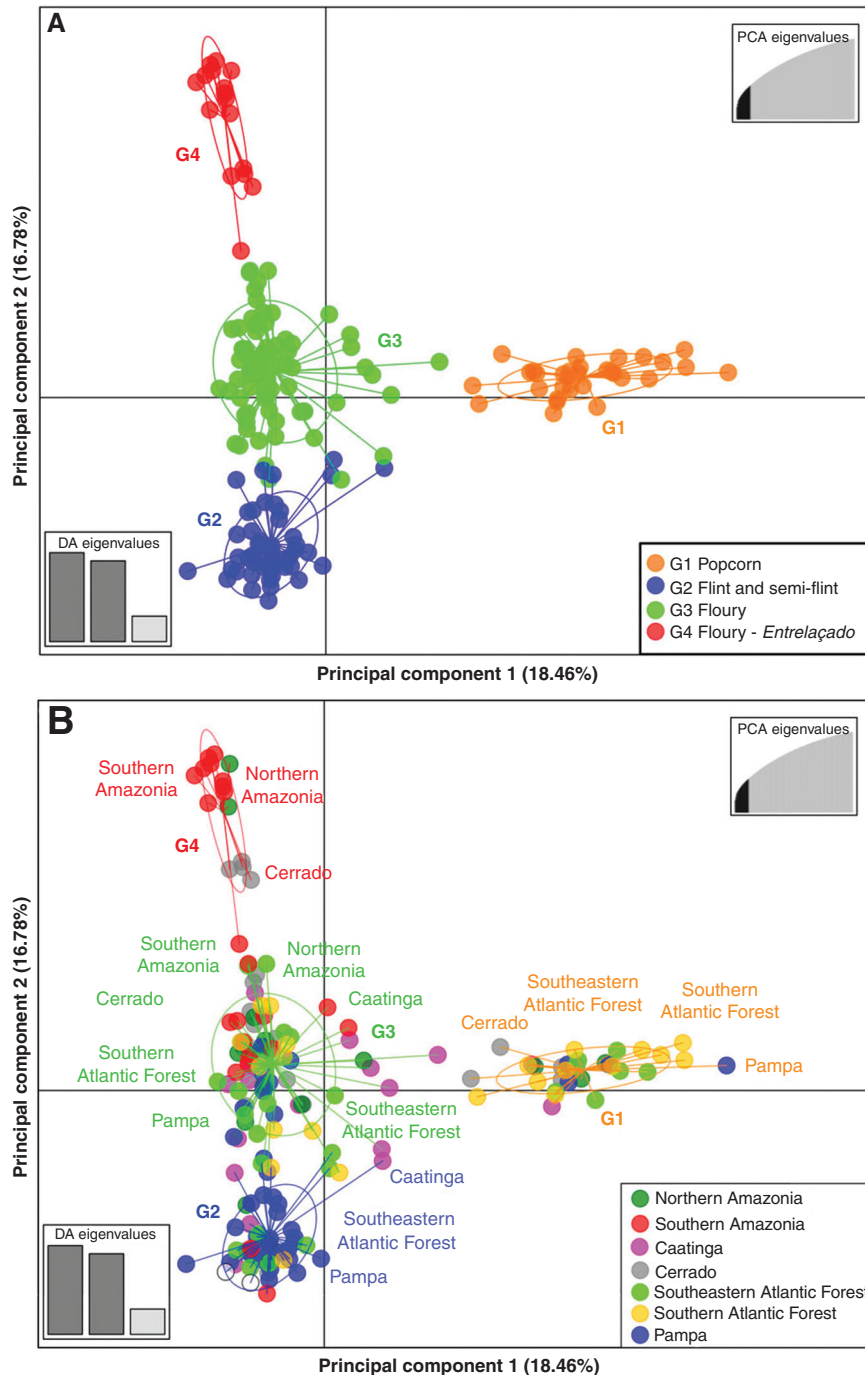


FIG. 2. Discriminant analysis of principal components (DAPC) based on 4398 SNPs from 170 maize accessions. Dispersion graphs of the groupings formed by the *K*-means method. (A) Groups identified by DAPC associated with the type of endosperm (Table 2), using Fisher's exact test at 5 % significance. (B) The predominant regional groups highlighted around each group identified in the DAPC.

TABLE 2. Frequency of endosperm types within the K-means groups identified by the discriminant analysis of principal components (DAPC) based on 4398 SNPs in 133<sup>a</sup> maize accessions.

Endosperm type	Group 1		Group 2		Group 3		Group 4	
	No.	%	No.	%	No.	%	No.	%
Popcorn	<b>27</b>	<b>96</b>	3	9	14	25	0	0
Flint/Semi-Flint	0	0	<b>26</b>	<b>76</b>	12	22	1	6
Floury	1	4	5	15	<b>29</b>	<b>53</b>	<b>15</b>	<b>94</b>
Total	28	100	34	100	55	100	16	100

Values in bold refer to the predominant endosperm types in the DAPC groups.

<sup>a</sup>The lower number of accessions is because no information on the type of endosperm was available for 35 accessions and two accessions are commercial hybrids.

Atlantic Forest and Caatinga; G3 in all biomes; and G4 in Southern Amazonia, Northern Amazonia and Cerrado.

The relationship between the 170 maize accessions is presented in an unrooted neighbour-joining dendrogram (Fig. 3), considering that an unrooted dendrogram specifies the relationships among accessions and not their evolutionary origin (Ridley, 2003; Barton *et al.*, 2007). The analysis suggests the existence of four clear clusters and one less clear (I) cluster; the latter is difficult to interpret, although dominated by floury endosperm (Fig. 3A). Three clusters (II, IV, V) are strongly dominated by only one endosperm type, popcorn, floury and flint, respectively. Cluster III is also dominated by floury, of which the *Entrelaçado* race is a particular type.

The relationship among endosperm clusters and regional geographical groups (Fig. 3B) suggests that in lowland South America no regional groups evolved in isolation. Nonetheless, some clustering patterns between regions can be interpreted as the result of dispersal processes. Specific regional groups predominated in these clusters, forming regional blocks, within which groups with frequencies >10 % were highlighted (Supplementary Data Table S4).

The regional connections are also supported by the sPCA (Fig. 4). In G1 (popcorn) (Fig. 4A and B), the sPCA shows relationships between the Pampa, Southern Atlantic Forest, Southeastern Atlantic Forest and Cerrado, diverging from the Southern Amazonia, Northern Amazonia and Caatinga. In G2 (flint and semi-flint) (Fig. 4C–E), a greater similarity between Southern Amazonia and Pampa is observed, diverging from the flint maize distributed along the Atlantic Coast and Northern Amazonia. The analysis carried out jointly for the groups dominated by floury endosperm (G3 + G4) shows a significantly greater relationship between Southern Amazonia and Northern Amazonia in the first principal component (PC) (Fig. 4F), differing from the maize distributed along the Atlantic Coast; the second PC (Fig. 4G) shows a greater similarity between Southern Amazonia and Pampa. The synthetic projection (Fig. 4H) demonstrates stronger relationships among the following biomes: (1) Southern Amazonia and Northern Amazonia; (2) Caatinga and Southeastern Atlantic Forest; (3) Southern Atlantic Forest and Pampa; and (4) Cerrado, which appears as a transition area between biomes. Interpolations of the first to the third significant principal components of these analyses are presented in Supplementary Data Fig. S1.

The sPCAs considering the two groups dominated by Floury endosperm separately (Supplementary Data Fig. S2) agreed with the patterns described above. However, the spatial synthetic projection of the PCs for G3 (floury) (Fig. 5A) identified the largest number of distinct groups (at least six colours), within which four were concentrated in the Atlantic Forest (Southern and Southeastern) and the southern Cerrado, which may suggest a diversification zone for this endosperm type. In this area, the presence of three language families is also observed (Arawak, Macro-Jê and Tupi), with one maize group (pink/purple) found exclusively where these three language families converge (Fig. 5). The samples from this region are in a Cerrado area and are apparently intermediate between the Northern (Amazonia) and Southern (Atlantic Forest and Pampa) biomes. In the DAPC, these samples are more closely related to Southern Amazonia. On the other hand, in the neighbour-joining analysis they are in group IV and are more closely related to the Atlantic Forest and Pampa, reinforcing their admixed ancestry.

#### Geographical distribution of maize genetic diversity in lowland South America

In our sample, floury endosperm maize (G3) was the most widely distributed type. The *Entrelaçado* floury type (G4) dominates in Amazonia and, somewhat less so, in the Cerrado. Curiously, popcorn (G1) is absent from Southern Amazonia and becomes very common in Southern Brazil, a region that encompasses the Atlantic Forest and Pampa biomes. Flint endosperm maize (G2) dominates in the Pampa and Southern Atlantic Forest and is common in the Caatinga.

The maize accessions in this study were mapped to indigenous language families (Eriksen, 2011), assuming that 400 years after European conquest there is still an association with these ethnic groups and maize. Taking into account that our sampling was not representative of linguistic families, we found that 38 % of the accessions were associated with the Tupi language family, 24 % with Macro-Jê, 11 % with Arawak and 1 % with Pano. The G3 (Floury) group presented the greatest richness of linguistic families (four). Groups G1, G2 and G4 were associated with the Arawak, Macro-Jê and Tupi families.

These analyses suggest several routes of maize dispersal that appear to be associated with the Arawak, Tupi (especially the Tupi-Guarani) and Macro-Jê expansions. Assuming an early presence of maize in Southwestern Amazonia, different migratory routes dispersed from this area in different directions (Fig. 6): (1) there was a dispersal towards northern and eastern Amazonia, which continued along the Atlantic coast, entered the Caatinga and south-eastern Atlantic Forest; (2) another dispersal occurred towards southern Brazil, passing through the Cerrado, the southern Atlantic Forest and reaching the Pampa region; and (3) along the Atlantic Coast, following Tupi movements originating from two separate expansions: one from north to south (Tupinamba), and the other in the opposite direction, from south to north (Guarani).

Regarding the comparison of our data with those of Kistler *et al.* (2018, 2020) (Supplementary Data Table S3), these authors classified the accessions into five ancestral groups: (1) Andes and the Pacific coast, (2) Lowland South America, (3)

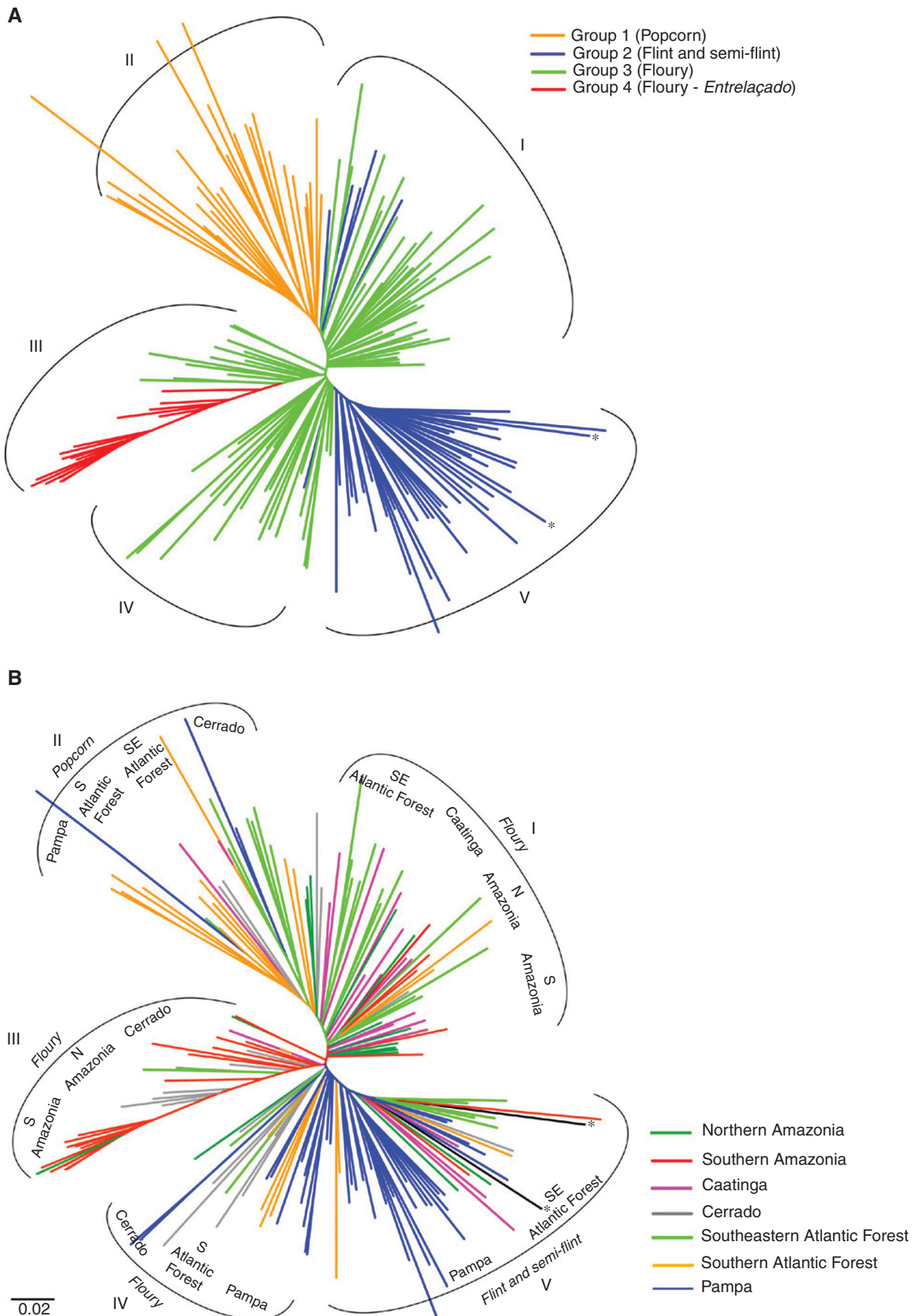


FIG. 3. Neighbour-joining dendrogram of 170 maize accessions characterized with 4398 SNPs using Nei's (1972) genetic distance. (A) Colours according to the *K*-means groups identified by the DAPC (Fig. 2). (B) Colours according to the regional geographical groups (Fig. 1). In each cluster the predominant endosperm types and regional groups are highlighted around the dendrogram. The following regional groups were defined: Northern (N) Amazonia, Southern (S) Amazonia, Caatinga, Cerrado, Southeastern (SE) Atlantic Forest, Southern (S) Atlantic Forest and Pampa. \*Commercial hybrid accessions.



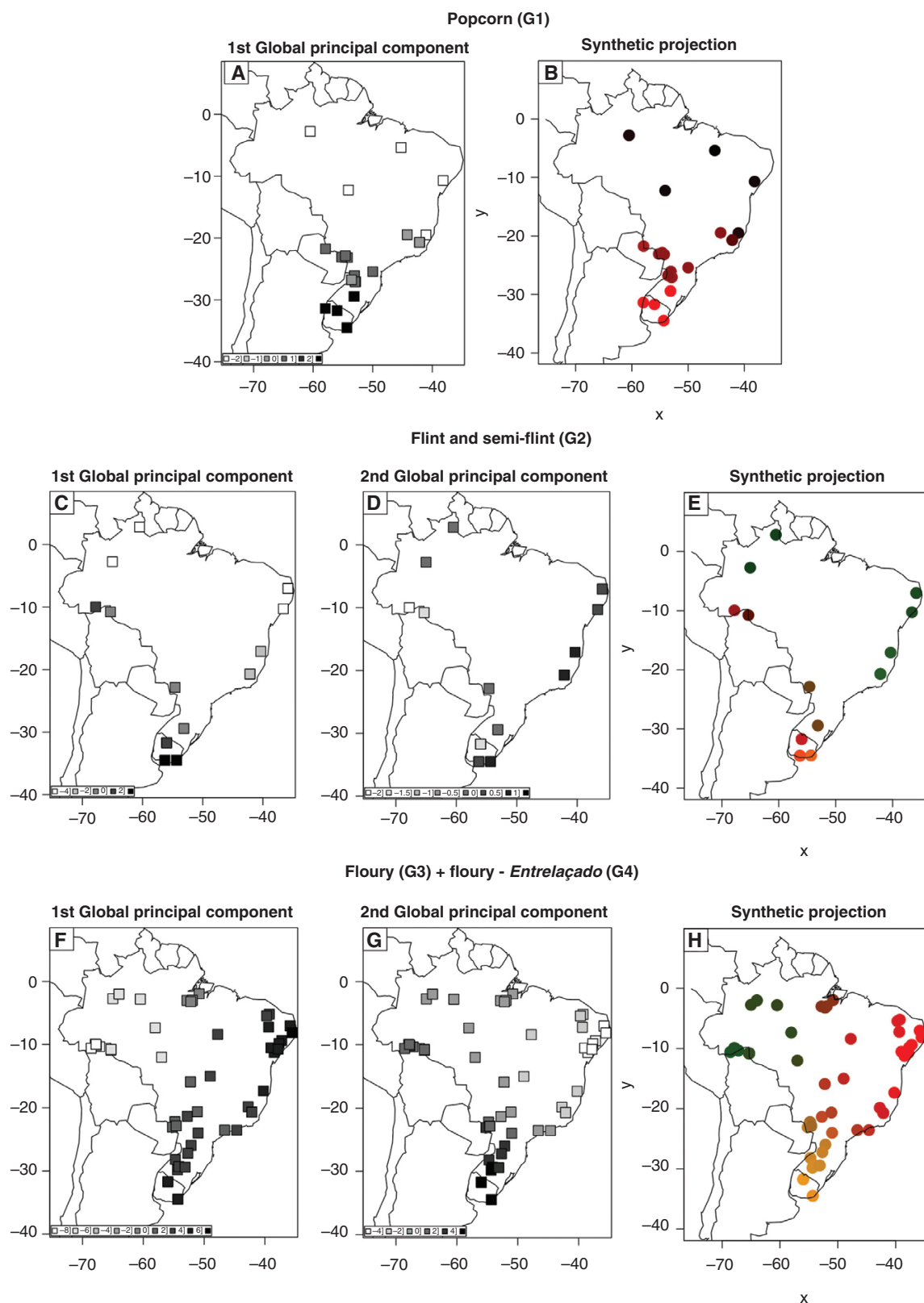


FIG. 4. Spatial principal component analysis (sPCA) of groups identified by the discriminant analysis of principal components (DAPC) based on 4398 SNPs in 170 maize accessions. Plots of significant global principal components of the genomic dataset were analysed. Each significant component is represented independently using squares in greyscale; the intensity of grey is related to the sample's position on the specific component. Synthetic projections summarize the representation of the values along all the significant principal components for each group, using circles with an RGB (red green blue) colour gradient; similar shades of colour indicate close genetic relationships. Popcorn (G1): projections of accessions and their position on the first (A) significant global component; synthetic projection of the first (B) significant component, using a colour gradient. Flint and Semi-Flint (G2): projections of accessions and their positions on the first (C) and the second (D) significant global components; synthetic projection of the first two (E) components. Flourey (G3) + Flourey - *Entrelaçado* (G4): projections of accessions and their positions on the first (F) and the second (G) significant global components; synthetic projection of the first two (H) components.

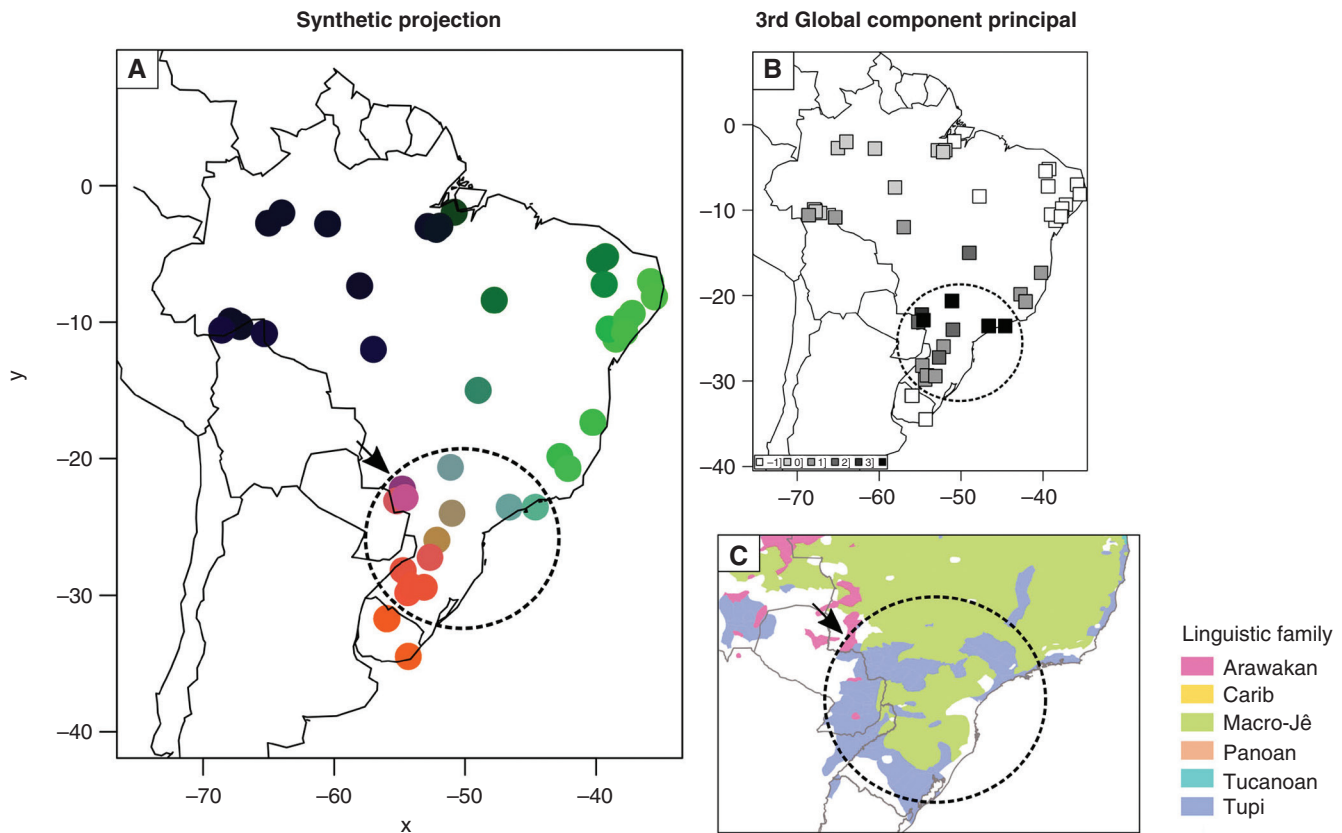


FIG. 5. Spatial principal component analysis (sPCA) of the G3 (Floury) group identified by the discriminant analysis of principal components (DAPC) based on 4398 SNPs in 55 maize accessions. Synthetic projections summarize the representation of the values along all the significant components for each group, using circles with an RGB colour gradient. Similar shades of colour indicate close genetic relationships. Conversely, different colours indicate more divergent accessions. Synthetic projection of the three (A) significant global components. Projections of accessions and their position on the third (B) significant global component. Distributions of the most important indigenous linguistic families (Eriksen, 2011) in the highlighted region (C). The dotted circle suggests a diversification zone for G3 and the arrow points to a zone of convergence of the Arawak, Macro-Jê and Tupi indigenous language families.

North America, (4) Highland Mexico and Central America, and (5) Pan-American, spanning from northern Mexico into lowland South America. From the five samples included in both studies, three were 100 % classified in the Lowland South America ancestral group and were dominated by floury endosperm. These samples originated from Southern Amazonia (two samples) and Northern Amazonia (one sample). Another sample, of popcorn endosperm, originating from the Cerrado biome, was classified into the Lowland South America (82 %) and North American (18 %) ancestral groups. Finally, a fifth accession of flint or semi-flint endosperm, from Northern Amazonia, was classified in the Pan-American (67 %) and Lowland South America (33 %) groups. Kistler *et al.* (2018, 2020) suggested that the Lowland South America ancestral group was an older introduction from South America (~6000 years BP) and the Pan-American group was a more recent introduction (~2000 years BP). The importance of the comparison of these studies is to show that floury and popcorn endosperm types were predominant in Lowland South America, while the flint endosperm predominated in the Pan-American ancestral group, suggesting that the floury and popcorn types were part of the older introduction in South America, while the flint type arrived in a more recent introduction.

## DISCUSSION

Population dynamics and structure are key aspects for understanding the distribution of crop genetic diversity, shaped by a complex set of evolutionary events involving ancestral human populations and modern communities (Kistler *et al.*, 2018, 2020). The present study expands our knowledge of the dispersal history of maize in lowland South America, involving a broader sample than previously available, and proposed hypotheses of dispersal of the species in the region. We have gathered genetic, archaeological and historical evidence to support our hypotheses, integrating them into a coherent history of maize dispersal. Our analyses suggested that maize dispersal was associated with different types of endosperm. The interregional connection patterns presented by the analyses suggest dispersal by different indigenous groups (Walker and Ribeiro, 2011; Walker *et al.*, 2012; Ramallo *et al.*, 2013; Iriarte *et al.*, 2016; Castro e Silva *et al.*, 2020). The geographical distribution of the linguistic families shows the main indigenous peoples reportedly involved in the processes of maize dispersal in this region: Tupi, Macro-Jê and Arawak. Furthermore, our data suggested a probable zone of diversification, involving the Cerrado and Atlantic Forest, in an area related to the Guarani (Tupi family) expansion (Bustamante *et al.*, 2014; Iriarte *et al.*, 2016).

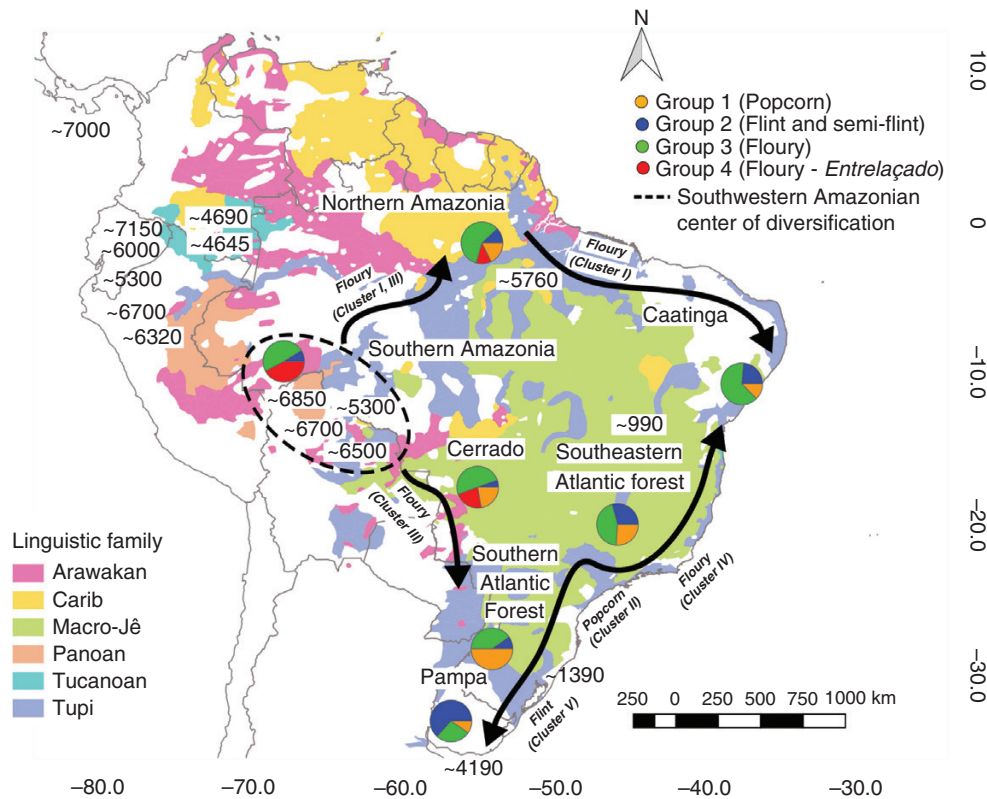


FIG. 6. Frequencies of the maize genetic groups defined by discriminant analysis of principal components (DAPC; pie charts by regional group); hypotheses of maize dispersal (arrows) suggested by the neighbour-joining dendrogram (interregional connections suggested by groups I, II, III, IV and V) and by the sPCA, associated with the history of Arawak (Walker & Ribeiro, 2011), Tupi (Walker et al., 2012; Iriarte et al., 2016; Castro e Silva et al., 2020) and Macro-Jê (Ramallo et al., 2013; Castro e Silva et al., 2020) expansions on map of the distributions of the six most important indigenous linguistic families (Eriksen, 2011); and the archaeological records of maize (in years BP).

DAPC and cluster analyses suggest that maize dispersal was associated with different types of endosperm. The G1 (popcorn) DAPC appeared in all regions of the research, except in Southern Amazonia. Although the present study did not sample any accessions of popcorn in Southwestern Amazonia, popcorn was present in Peru by 6200 years BP (Grobman et al., 2012). In the South of Brazil, popcorn was important to the indigenous Guarani (Brieger et al., 1958; Paterniani and Goodman, 1977). In Santa Catarina, a microcentre of maize diversity was identified, with many popcorn landraces (Costa et al., 2017; Silva et al., 2017). A recent study showed that the expansion capacity or popping expansion pattern (the formation of large flakes after the kernels explode in response to heating) of popcorn landraces displayed a north–south geographical continuum, in which the average values increased with geographical distance from the centre of origin in Mexico (Silva et al., 2020). The highest averages were obtained for the lowland groups of South America (Silva et al., 2020). This evidence suggests that when popcorn arrived in South America, it gained importance among the people who lived there and was widely cultivated and, consequently, diversified. The G2 group (Flint and Semi-flint) showed higher frequencies along the Atlantic Coast and lower frequencies in the interior, which corroborates the hypothesis of dispersal of the *Cateto* flint maize race along the coast (Brieger et al., 1958; Paterniani and Goodman, 1977; Goodman, 1995). G3 (predominantly Floury) had a wider dispersal and was distributed

in all regions of the study. Our findings agree with the proposal for the geographical distribution of the endosperm types of maize in South America presented by Goodman (1995), within which the presence of floury maize was pointed out in the lowland areas, with typical characteristics of the *Entrelaçado* race, in both Amazonia and the Cerrado. All indigenous maize races (*Caingang*, *Entrelaçado*, *Lenha*, *Moroti* and *Guarani Popcorns*) are probably very similar to maize grown in the Pre-Columbian era, and in the case of the *Entrelaçado* race, their occurrence exclusively in Amazonia was noted in previous studies (Brieger et al., 1958; Paterniani and Goodman, 1977; Costa et al., 2021). However, G4 (*Entrelaçado*) also dispersed out of Amazonia and arrived in the Cerrado.

Interregional connection patterns suggested by cluster analysis and sPCA supported our hypothesis of maize dispersal. Cluster analysis showed a greater relationship between Southern Amazonia, Northern Amazonia and the Atlantic Coast (cluster I) of floury maize (G3). This same path was followed by the expansion of Arawak (Walker and Ribeiro, 2011) and Tupi (Tupinamba) (Iriarte et al., 2016; Castro e Silva et al., 2020). sPCA also suggested a dispersal pattern for Flint (G2) and Floury (G3 + G4) passing through the Cerrado, towards the Pampa. Group V of the cluster analysis, with the highest frequency of the flint and semi-flint grain types, showed a relationship between the Pampa and Atlantic Forest. Brieger et al. (1958) and Paterniani and Goodman (1977) suggested a coastal migratory movement in the south–north direction by

the peoples of the Tupi linguistic family. The authors assumed that the *Cateto* maize race, which has a flint-type endosperm, is native to the Brazil–Argentina coastal region (Pampa biome), where this group showed its maximum development, having been cultivated by the indigenous peoples of the Atlantic Coast of Argentina up to the Guianas (with dispersal in the south–north direction).

Archaeological evidence shows that maize arrived in lowland South America thousands of years ago and dispersed across the region (Table 1; Fig. 6). The oldest records identified were 6850 years BP in Southwestern Amazonia (Lombardo *et al.*, 2020), which indicates that before the appearance of the Arawak and Tupi groups there were people cultivating maize in the region. Considering the oldest presence in this region, we assume that this area corresponds to the origin of the species' dispersal routes, which also overlaps with the origin of indigenous expansions considered in our hypotheses (Arawak, Tupi). Archaeological records also point to the presence of maize in the Pampa region, at least 4190 years BP (Iriarte *et al.*, 2004), which also suggests that an ancient movement occurred from Southwestern Amazonia to the south. Considering these linguistic families, the Tupi, Macro-Jê and Arawak were the main indigenous peoples reportedly involved in the processes of maize dispersal. Among these three groups, the Tupi language family predominated (Supplementary Data Fig. S3). Less frequently, the Macro-Jê and Arawak presence was also observed in all clusters, which may indicate only traces of Macro-Jê and Arawak dispersal in these areas, since the regions that involve the Macro-Jê, Arawak and Tupi routes overlap. These associations must be interpreted cautiously because few germplasm collection expeditions use Indigenous language as a priority descriptor and Tupi is the most common Indigenous language in Brazil. Nonetheless, we suggest that the Tupi language family was important in maize dispersal, as well as the Arawak and Macro-Jê families, as also observed by Kistler *et al.* (2018). According to these authors, several landraces and two archaeogenomes (~700 years BP) located near Southwestern Amazonia share alleles with Cerrado and Atlantic Forest maize. This study also analysed the linguistic distribution of the word 'maize' for Arawak and Macro-Jê groups, showing that the genetic pattern observed is consistent with the Arawak and Macro-Jê linguistic patterns that link maize cultivation in these regions (Kistler *et al.*, 2018).

In the synthetic projection of G3 (floury endosperm) in the sPCA (Fig. 5), a probable zone of diversification was suggested in an area that presents high biocultural diversity, characterized by the presence of three indigenous linguistic families and two biomes. This area involves a Guaraní expansion zone, according to Bustamante *et al.* (2014) and Iriarte *et al.* (2016). Bracco *et al.* (2016) identified two genetic groups in lowland South America associated with popcorn and floury endosperm types, with high divergence between them, as observed in our study (Fig. 2A). The authors assumed that these maize groups were grown by the Guaraní people (from the Tupi–Guaraní linguistic family) (Bracco *et al.*, 2016), where floury maize and popcorn were known as *Moroti* and *Guaraní Popcorns*, respectively (Paterniani and Goodman, 1977). The present study, with a broader sampling, demonstrated the presence and distribution of floury and popcorn genetic groups in different biomes

of the region. The data that we have compared to Kistler *et al.* (2018, 2020) (Supplementary Data Table S3) showed that the floury and popcorn endosperm types were related to the older 'Lowland South America' ancestral group, and the flint type to the 'Pan-American', the most recent group that supposedly arrived later. This suggests that older maize dispersal routes probably involved floury and popcorn, and did not include flint. In addition, the fact that the G3 samples belong to the Lowland South America group (100%), one of the oldest ancestral groups in South America (Kistler *et al.*, 2018, 2020), may be related to the greater variation in spatial diversity observed within this group (Fig. 5). Kistler *et al.* (2018, 2020) suggested a secondary improvement centre of maize in Southwestern Amazonia, recognized as a centre of domestication and diversification of cultivated species already reported in other studies (Piperno, 2011; Clement *et al.*, 2016; Dequigiovanni *et al.*, 2018; Lombardo *et al.*, 2020). Archaeobotanical studies found phytoliths and starch in ceramics at 3000–1500 years BP in the upper Madeira River region, Southwestern Amazonia (Watling *et al.*, 2020), possibly involved with the Tupi expansion (Iriarte *et al.*, 2016; Castro e Silva *et al.*, 2020), where the most represented species in ceramics was maize. Starch grains produced by soft (floury) and stiff (popcorn) endosperm maize were found in these residues, implying that landraces with different types of endosperm were being used in this period. The damage present in the starch grains found in the ceramics indicates that the plants underwent different mechanical processes (e.g. milling) and heating (toasted and boiled), which indicates the production and processing of maize flour (Watling *et al.*, 2020). Ethnobotanical research also conducted in Southwestern Amazonia (Costa *et al.*, 2021) identified different typical uses of maize, such as fermented drink (called *chicha* by the native peoples), in indigenous and riverside communities up to the present. Bracco *et al.* (2016) noted that maize from central South America constitutes a unique and locally adapted gene pool, and does not represent just a region of contact between Andean and Eastern South American races. These authors suggest that this region played an important role in structuring the genetic diversity of local races.

## CONCLUSIONS

Our study expands knowledge of the evolutionary history of maize in lowland South America and proposes hypotheses of dispersal of the species associated with different types of endosperm. Interregional patterns suggest dispersal in agreement with the expansion of the Tupi, Macro-Jê and Arawak indigenous groups. From a possible origin in Southwestern Amazonia, different maize dispersal routes emerged: (1) towards Northern Amazonia, which continued towards the Caatinga and south-eastern Atlantic Forest (Floury); (2) towards Southern Brazil, passing through the Cerrado and Southern Atlantic Forest reaching the Pampa region (Floury); and (3) along the Atlantic Coast, following Tupi migratory movements and originating from two separate expansions: one from north to south, and the other in the opposite direction, from south to north (Flint, Floury and popcorn).

## SUPPLEMENTARY DATA

Supplementary data are available online at <https://academic.oup.com/aob> and consist of the following. **Table S1**: Passport data of 170 maize accessions used in this study. **Table S2**: Allele frequencies from SNP markers for 170 maize accessions. **Table S3**: Accessions included in the study of Kistler *et al.*, with five of them also characterized in our study. **Table S4**: Frequency of regional groups within the clusters identified by the neighbour-joining dendrogram of 170 maize accessions characterized by 4398 SNPs, using Nei's genetic distance. **Fig. S1**: Spatial principal component analysis of groups identified by the discriminant analysis of principal components based on 4398 SNPs in 170 maize accessions. **Fig. S2**: Spatial principal component analysis of groups identified by the discriminant analysis of principal components based on 4398 SNPs in 55 and 16 maize accessions. **Fig. S3**: Cluster analysis of 170 maize accessions performed by the neighbour-joining method, based on 4398 SNP markers and Nei's genetic distance.

## ACKNOWLEDGEMENTS

We thank especially all local and indigenous farmers who collaborated with this study. We also thank the Collaborative Research Network of the Interdisciplinary Group of Agrobiodiversity Study – InterABio (<https://interabiogrupo.wixsite.com/interabio>) – for the support provided in the maize collections; the Brazilian Agricultural Research Corporation – Embrapa Milho e Sorgo – for kindly donating samples from its maize collection; and Logan Kistler, National Museum of Natural History, Smithsonian Institution, and Robin Allaby, University of Warwick, for sharing the data of Kistler *et al.* (2018). F.M.C. thanks CNPq for a doctoral (142371/2016-5) and a post-doctoral (162792/2020-4) scholarship. A.A-P. thanks FAPESP (2018/00036-9) for a post-doctoral scholarship. CNPq awarded scholarships and research fellowships to N.C.A.S., E.A.V. and C.R.C.

## FUNDING

This work was supported by São Paulo Research Foundation (FAPESP; process 2015/26837-0, Brazil); Brazilian National Council for Scientific and Technological Development (CNPq; process 421045/2016-7, 142371/2016-5 and 162792/2020-4, Brazil); and Sectoral Commission of Scientific Research (CSIC; process id2016/400, Uruguay).

## LITERATURE CITED

- Aikhenvald AY. 2013. Amazonia: linguistic history. In: Ness I, Bellwood P, eds. *The encyclopedia of global human migration*. Hoboken: Wiley-Blackwell, 384–391.
- Akima H, Gebhardt A. 2021. Interpolation of irregularly and regularly spaced data. R package version 0.6-2.3. <https://cran.r-project.org/web/packages/akima/akima.pdf>.
- Al-Beyroutiová M, Sabo M, Sleziak P, *et al.* 2016. Evolutionary relationships in the genus *Secale* revealed by DArTseq DNA polymorphism. *Plant Systematics and Evolution* 302: 1083–1091. doi:10.1007/s00606-016-1318-2.
- Anderson E, Cutler H. 1942. Races of *Zea mays*: I. Their recognition and classification. *Annals of the Missouri Botanical Garden* 29: 69–89.
- Arca M, Gouesnard B, Mary-Huard T, *et al.* 2020. Genome-wide SNP genotyping of DNA pools identifies untapped landraces and genomic regions that could enrich the maize breeding pool. *BioRxiv*. doi:10.1101/2020.09.30.321018.
- Arca M, Mary-Huard T, Gouesnard B, *et al.* 2021. Deciphering the genetic diversity of landraces with high-throughput SNP genotyping of DNA bulks: methodology and application to the maize 50k Array. *Frontiers in Plant Science* 11: 568699. doi:10.3389/fpls.2020.568699.
- Barton, NH, Briggs D, Eisen JA, Goldstein DB, Patel NH. 2007. *Evolution*. New York: Cold Spring Harbor Laboratory Press.
- Bedoya CA, Dreisigacker S, Hearne S, *et al.* 2017. Genetic diversity and population structure of native maize populations in Latin America and the Caribbean. *PLoS One* 12: e0173488. doi:10.1371/journal.pone.0173488.
- Bird RM, Goodman MM. 1977. The races of maize V: grouping maize races on the basis of ear morphology. *Economic Botany* 31: 471–481. doi:10.1007/bf02912560.
- Bracco M, Cascales J, Hernández JC, Poggio L, Gottlieb AM, Lia VV. 2016. Dissecting maize diversity in lowland South America: genetic structure and geographic distribution models. *BMC Plant Biology* 16: 1–13. doi:10.1186/s12870-016-0874-5.
- Brieger FG, Gurgel JTA, Paterniani E, Blumenchein A, Alleoni MR. 1958. *Races of maize in Brazil and other eastern South America Countries*. Washington: National Academic of Sciences.
- Brown CH, Clement CR, Epps P, Luedeling E, Wichmann S. 2014. The paleobiogeography of maize (*Zea mays* L.). *Ethnobiology* 5: 52–64. doi:10.14237/eb1.5.2014.130.
- Brugger SO, Gobet E, van Leeuwen JFN, *et al.* 2016. Longterm man–environment interactions in the Bolivian Amazon: 8000 years of vegetation dynamics. *Quaternary Science Reviews* 132: 114–128. doi:10.1016/j.quascirev.2015.11.001.
- Bush MB, Piperno DR, Colinvaux PA. 1989. A 6,000 year history of Amazonian maize cultivation. *Nature* 340: 303–305. doi:10.1038/340303a0.
- Bush MB, Correa-Metrio A, McMichael CH, *et al.* 2016. A 6900-year history of landscape modification by humans in lowland Amazonia. *Quaternary Science Reviews* 141: 52–64. doi:10.1016/j.quascirev.2016.03.022.
- Bush MB, Miller MC, De Oliveira PE, Colinvaux PA. 2000. Two histories of environmental change and human disturbance in eastern lowland Amazonia. *The Holocene* 10: 543–553. doi:10.1191/095968300672647521.
- Bustamante PG, Pádua V, LM, Maia VH, Ferreira PCG. 2014. Evidence of maize (*Zea mays* L.) exploitation around an ancient crossroad linking different aboriginal American civilizations. *Acta Scientiae et Technicae* 29: 2317–8957. doi:10.17648/uezo-ast-v2i1.56.
- Castro e Silva M, Nunes K, Lemes RB, *et al.* 2020. Genomic insight into the origins and dispersal of the Brazilian coastal natives. *Proceedings of the National Academy of Sciences* 117: 2372–2377. doi:10.1073/pnas.1909075117.
- Clement CR, Rodrigues DP, Alves-Pereira A, *et al.* 2016. Crop domestication in the upper Madeira River basin. *Boletim do Museu Paraense Emílio Goeldi* 11: 193–205. doi:10.1590/1981.81222016000100010.
- Costa FM, Silva NCDA, Ogliari JB. 2017. Maize diversity in southern Brazil: indication of a microcenter of *Zea mays* L. *Genetic Resources and Crop Evolution* 64: 681–700. doi:10.1007/s10722-016-0391-2.
- Costa FM, Silva NCA, Vidal R, *et al.* 2021. *Entrelaçado*, a rare maize race conserved in Southwestern Amazonia. *Genetic Resources and Crop Evolution* 68: 51–58. doi:10.1007/s10722-020-01008-0.
- Dequigiovanni G, Ramos SLF, Alves-Pereira A, *et al.* 2018. Highly structured genetic diversity of *Bixa orellana* var. *urucurana*, the wild ancestor of annatto, in Brazilian Amazonia. *PLoS One* 13: e0198593. doi:10.1371/journal.pone.0198593.
- Doyle JJ, Doyle JL. 1990. Isolation of plant DNA from fresh tissue. *Focus* 12: 13–15.
- Dubreuil P, Charcosset, A. 1998. Genetic diversity within and among maize populations: a comparison between isozyme and nuclear RFLP loci. *Theoretical and Applied Genetics* 96: 577–587.
- Eriksen L. 2011. *Nature and culture in prehistoric Amazonia: using GIS to reconstruct ancient ethnogenetic processes from archaeology, linguistics, geography, and ethnohistory*. PhD thesis, Lund University, Lund.
- FAO. 2018. *FAOSTAT statistics database*. Rome: Food and Agriculture Organization of the United Nations, Italy. <http://www.fao.org/faostat/>. (accessed 3 March 2020).
- FAO. 2021. *World Information and Early Warning System on Plant Genetic Resources for Food and Agriculture*. Rome: Food and Agriculture

- Organization of the United Nations, Itay. <https://www.fao.org/wiews/glossary/en/#:~:text=Accession%3A%20A%20distinct%2C%20uniquely%20identifiable,storage%20for%20conservation%20and%20use>.
- Fisher RA. 1935.** The logic of inductive inference. *Journal of the Royal Statistical Society* **98**: 39–54. doi:10.2307/2342435.
- Fonseca RRD, Smith BD, Wales N, et al. 2015.** The origin and evolution of maize in the Southwestern United States. *Nature Plants* **1**: 14003. doi:10.1038/nplants.2014.3.
- Freitas FO, Bandel G, Allaby RG, Brown TA. 2003.** DNA from primitive maize landraces and archaeological remains: implications for the domestication of maize and its expansion into South America. *Journal of Archaeological Science* **30**: 901–908.
- Freitas FO, Bustamante PG. 2013.** Amazonian maize: diversity, spatial distribution and historical-cultural diffusion. *Tipiti: Journal of the Society for the Anthropology of Lowland South America* **11**: 60–65.
- Grobman A, Salhauana W, Sevilla R. 1961.** *Races of maize in Peru: their origins, evolution and classification*. Publication 915. Washington: National Academy of Sciences.
- Grobman A, Bonavia D, Dillehay TD, Piperno DR, Iriarte J, Holst I. 2012.** Pre-ceramic maize from Paredones and Huaca Prieta, Peru. *Proceedings of the National Academy of Sciences* **109**: 1755–1759. doi:10.1073/pnas.1120270109.
- Gouda, AC, Ndjondjop, MN, Djedatin, GL, et al. 2020.** Comparisons of sampling methods for assessing intra- and inter-accession genetic diversity in three rice species using genotyping by sequencing. *Scientific Reports* **10**: 13995. doi:10.1038/s41598-020-70842-0.
- Goodman M. 1995.** Maize. In: Smartt J, Simmonds N, eds. *Evolution of crop plants*. London: Longman, 192–202
- Hammer O, Harper D, Ryan PD. 2001.** PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica* **4**: 1–9.
- Herrera LF, Cavelier I, Rodriguez C, Mora S. 1992.** The technical transformation of an agricultural system in the Colombian Amazon. *World Archaeology* **24**: 98–113. doi:10.1080/00438243.1992.9980196.
- Hilbert L, Neves EG, Pugliese F, et al. 2017.** Evidence for mid-Holocene rice domestication in the Americas. *Nature Ecology & Evolution* **1**: 1693–1698. doi:10.1038/s41559-017-0322-4.
- Hufford MB, Lubinsky P, Pyhäjärvi T, Devengenzo MT, Ellstrand NC, Ross-Ibarra J. 2013.** The genomic signature of crop-wild introgression in maize. *PLoS Genetics* **9**: e1003477. doi:10.1371/journal.pgen.1003477.
- IBPGR. 1991.** *Descriptors for maize*. Rome: International Board for Plant Genetic Resources (IBPGR).
- Iriarte J, Holst I, Marozzi O, et al. 2004.** Evidence for cultivar adoption and emerging complexity during the mid-Holocene in the La Plata Basin, Uruguay. *Nature* **432**: 614–617. doi:10.1038/nature02983.
- Iriarte J, Smith R, Gregorio de Souza J, et al. 2016.** Out of Amazonia: Late-Holocene climate change and the Tupi-Guarani trans-continental expansion. *The Holocene* **27**: 967–975. doi:10.1177/09596836166678461.
- Jaenicke-Després V, Buckler ES, Smith BD, et al. 2003.** Early allelic selection in maize as revealed by ancient DNA. *Science* **302**: 1206–1208. doi:10.1126/science.1089056.
- Jombart T. 2008.** Adegnet: A R package for the multivariate analyses of genetic markers. *Bioinformatics* **24**: 1403–1405. doi:10.1093/bioinformatics/btn129.
- Jombart T, Devillard S, Balloux F. 2010.** Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* **11**: 94. doi:10.1186/1471-2156-11-94.
- Jombart T, Devillard S, Dufour AB, Pontier D. 2008.** Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity* **101**: 92–103. doi:10.1038/hdy.2008.34.
- Kistler LM, Maezumi SY, de Souza JG, et al. 2018.** Multi-proxy evidence highlights a complex evolutionary legacy of maize in South America. *Science* **362**: 1309–1313. doi:10.1126/science.aav0207.
- Kistler LM, Thakar HB, VanDerwarker AM, et al. 2020.** Archaeological Central American maize genomes suggest ancient gene flow from South America. *Proceedings of the National Academy of Sciences* **117**: 33124–33129. doi:10.1073/pnas.2015560117.
- Lia VV, Confalonieri VA, Ratto N, et al. 2007.** Microsatellite typing of ancient maize: insights into the history of agriculture in southern South America. *Proceedings Biological Sciences* **274**: 545–554. doi:10.1098/rspb.2006.3747.
- Lombardo U, Iriarte J, Hilbert L, Ruiz-Pérez J, Capriles JM, Veit H. 2020.** Early Holocene crop cultivation and landscape modification in Amazonia. *Nature* **581**: 190–193. doi:10.1038/s41586-020-2162-7.
- Matsuoka Y, Vigouroux Y, Goodman MM, Sánchez JJ, Buckler E, Doebley JF. 2002.** A single domestication for maize shown by multilocus microsatellite genotyping. *Proceedings of the National Academy of Sciences* **99**: 6080–6084. doi:10.1073/pnas.052125199.
- Migliazza EC. 1982.** Linguistic prehistory and the refuge model in Amazonia. In: Prance GT, ed. *Biological diversification in the tropics*. New York: Columbia University Press, 497–519.
- Mir C, Zerjal T, Combes V, et al. 2013.** Out of America: tracing the genetic footprints of the global diffusion of maize. *Theoretical and Applied Genetics* **126**: 2671–2682. doi:10.1007/s00122-013-2164-z.
- Moreno-Letelier A, AguirreLiguori JA, Piñero D, Vázquez-Lobo A, Eguarte LE. 2020.** The relevance of gene flow with wild relatives in understanding the domestication process. *Royal Society Open Science* **7**: 191545. doi:10.1098/rsos.191545.
- Nei M. 1972.** Genetic distance between populations. *The American Naturalist* **106**: 283–292.
- Paradis E, Schliep K. 2018.** ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**: 526–528. doi:10.1093/bioinformatics/bty633.
- Paterniani E, Goodman MM. 1977.** *Races of maize in Brazil and adjacent areas*. Mexico City: CIMMYT.
- Pearsall DM, Piperno DR. 1990.** Antiquity of maize cultivation in Ecuador: summary and reevaluation of the evidence source. *American Antiquity* **55**: 324–337. doi:10.2307/281650.
- Petroli C, Kilian A. 2019.** *Introduction to the DArTseq genotyping method and its data outputs*. CIMMYT Research Data & Software Repository Network. <https://doi.org/hdl:11529/10548358>
- Piperno DR. 1990.** Aboriginal agriculture and land usage in the Amazon Basin, Ecuador. *Journal of Archaeological Science* **17**: 665–677. doi:10.1016/0305-4403(90)90048-A.
- Piperno DR. 2011.** The origins of plant cultivation and domestication in the New World tropics: patterns, process, and new developments. *Current Anthropology* **52**: 453–470. doi:10.1086/659998.
- Piperno DR, Ranere AJ, Holst I, Hansell P. 2000.** Starch grains reveal early root crop horticulture in the Panamanian tropical forest. *Nature* **407**: 894–897. doi:10.1038/35038055.
- Piperno DR, Ranere AJ, Holst I, Iriarte J, Dickau R. 2009.** Starch grain and phytolith evidence for early ninth millennium b.p. maize from the Central Balsas River Valley, Mexico. *Proceedings of the National Academy of Sciences* **106**: 5019–5024. doi:10.1073/pnas.0812525106.
- Pritchard JK, Stephens M, Donnelly P. 2000.** Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.
- R Development Core Team. 2019.** *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. <http://www.R-project.org/>. (accessed 20 October 2019).
- Ramallo V, Bisso-Machado R, Bravi C, et al. 2013.** Demographic expansions in South America: Enlightening a complex scenario with genetic and linguistic data. *American Journal Physical Anthropology* **150**: 453–463. doi:10.1002/ajpa.22219.
- Ramos-Madrigal J, Smith BD, Moreno-Mayar JV, et al. 2016.** Genome sequence of a 5,310-year-old maize cob provides insights into the early stages of maize domestication. *Current Biology* **26**: 3195–3201. doi:10.1016/j.cub.2016.09.036.
- Reif JC, Hamrit S, Heckenberger M, et al. 2005.** Genetic structure and diversity of European flint maize populations determined with SSR analyses of individuals and bulks. *Theoretical and Applied Genetics* **111**: 906–913. doi:10.1007/s00122-005-0016-1.
- Reyes-Valdés, MH, Santacruz-Varela, A, Martínez, O, Simpson, J, Hayano-Kanashiro C, Cortés-Romero, C. 2013.** Analysis and optimization of Bulk DNA sampling with binary scoring for germplasm characterization. *PLoS One* **8**: e79936. doi:10.1371/journal.pone.0079936.
- Ridley M. 2003.** *Evolution*. Oxford: Wiley-Blackwell.
- Saitou N, Nei M. 1987.** The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**: 406–425. doi:10.1093/oxfordjournals.molbev.a040454.
- Sansaloni C, Petroli C, Jaccoud D, et al. 2011.** Diversity Arrays Technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of *Eucalyptus*. *BMC Proceedings* **5**: P54. doi:10.1186/1753-6561-5-S7-P54.
- Sansaloni C, Franco J, Santos B, et al. 2020.** Diversity analysis of 80,000 wheat accessions reveals consequences and opportunities of selection footprints. *Nature Communications* **11**: 4572. doi:10.1038/s41467-020-18404-w.

- Schmitz PI. 1991. *Migrantes da Amazônia: a tradição tupi-guarani. Arqueologia do Rio Grande do Sul, Brasil*. São Leopoldo: Instituto Anchieta de Pesquisas/UNISINOS.
- Silva NCA, Vidal R, Ogliari JB. 2017. New popcorn races in a diversity microcenter of *Zea mays* L. in the Far West of Santa Catarina, Southern Brazil. *Genetic Resources and Crop Evolution* 64: 1191–1204. doi:10.1007/s10722-016-0429-5.
- Silva NCA, Vidal R, Ogliari JB, Costich D, Chen J. 2020. Relationships among American popcorns and their links with landraces conserved in a micro center of diversity. *Genetic Resources and Crop Evolution* 67: 1733–1753. doi:10.1007/s10722-020-00935-2.
- Staller J, Tykot R, Benz B. 2006. *Histories of maize: multidisciplinary approaches to the prehistory, linguistics, biogeography, domestication, and evolution of maize*. Burlington: Elsevier Academic Press.
- Stothert KE. 1985. The preceramic Las Vegas culture of Coastal Ecuador. *American Antiquity* 50: 613–637. doi:10.2307/280325.
- Vallebuena-Estrada M, Rodríguez-Arévalo I, Rougon-Cardoso A, et al. 2016. The earliest maize from San Marcos Tehuacán is a partial domesticate with genomic evidence of inbreeding. *Proceedings of the National Academy of Sciences* 113: 14151–14156. doi:10.1073/pnas.1609701113.
- van Heerwaarden J, Doebley J, Briggs WH, Glaubitz JC, Goodman MM. 2011. Genetic signals of origin, spread, and introgression in a large sample of maize landraces. *Proceedings of the National Academy of Sciences* 108: 1088–1092. doi:10.1073/pnas.1013011108.
- Vigouroux Y, Glaubitz J, Matsuoka Y, Goodman MM, Sánchez GJ, Doebley JF. 2008. Population structure and genetic diversity of new world maize races assessed by DNA microsatellites. *American Journal of Botany* 95: 1240–1253. doi:10.3732/ajb.0800097.
- Walker RS, Ribeiro LA. 2011. Bayesian phylogeography of the Arawak expansion in lowland South America. *Proceedings Biological Sciences* 278: 2562–2567. doi:10.1098/rspb.2010.2579.
- Walker RS, Wichmann S, Mailund T, Atkisson CJ. 2012. Cultural phylogenetics of the Tupi language family in lowland South America. *PLoS One* 7: e35025. doi:10.1371/journal.pone.0035025.
- Watling J, Almeida F, Kater T, et al. 2020. Archaeobotanical data from ceramic occupations of the Cachoeira do Teotônio. *Boletim do Museu Paraense Emílio Goeldi* 15: e20190075. doi:10.1590/2178-2547-bgoeldi-2019-0075.
- Wesolowski V, de Souza SMFM, Reinhard KJ, Ceccantini G. 2010. Evaluating microfossil content of dental calculus from Brazilian sambaquis. *Journal of Archaeological Science* 37: 1326–1338. doi:10.1016/j.jas.2009.12.037.