Genetic Diversity of Cambuci (*Campomanesia Phaea)* Revealed by Microsatellite Markers

Rafael Oliveira Moreira
Universidade de Sao Paulo Escola Superior de Agricultura Luiz de Queiroz
Eduardo de Andrade Bressan
Universidade de Sao Paulo Centro de Energia Nuclear na Agricultura
Horst Bremer Neto
Universidade de Sao Paulo Escola Superior de Agricultura Luiz de Queiroz
Angelo Pedro Jacomino
Universidade de Sao Paulo Escola Superior de Agricultura Luiz de Queiroz
Antonio Figueira
Universidade de Sao Paulo Centro de Energia Nuclear na Agricultura
Francisco de Assis Alves Mourão Filho (francisco.mourao@usp.br)
Universidade de São Paulo Escola Superior de Agricultura Luiz de Queiroz https://orcid.org/0000-0001-8950-9513

Research Article

Keywords: Atlantic Forest, core collection, Myrtaceae, population structure, SSR markers.

Posted Date: July 13th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-669769/v1

Abstract

Campomanesia phaea (Myrtaceae), known as cambuci, is a native species from the Brazilian Atlantic Forest with great potential to be developed as a new fruit crop. Microsatellite markers were developed for cambuci to characterize the genetic diversity and to investigate the genetic structure of a group of accessions originally collected at the presumed center of diversity of the species. The work involved the collection of 145 accessions from five regional groups (Juquitiba, Paraibuna, Mogi das Cruzes, Ribeirão Pires, and Salesópolis) in São Paulo state, Brazil. Fourteen loci were identified in an enriched genomic library developed from one of these accessions. Six out of 14 loci revealed to be polymorphic, disclosing 26 alleles. Based on the allele frequencies, the calculated genetic parameters of the five groups indicated an average allele number per loci (A) of 3.83, with the expected heterozygosity (He) of 0.57 and the observed heterozygosity (Ho) of 0.54. The analysis of the genetic structure indicated that most of the genetic diversity is found within each population (H_S = 0.57), whereas the genetic diversity among populations was low (G_{ST} = 0.19). The genetic diversity parameter of Nei was considered low for the cambuci analyzed populations, with no evidence of inbreeding. Based on Darwin analysis, we chose 18 accessions from the five regional populations to compose a core collection that includes most of the genetic diversity found in this study. Our findings may contribute to define better conservation strategies and genetic breeding approaches for this native species in Brazil.

Introduction

Myrtaceae is one of the most species-rich families found in Brazil, with over 1,000 native species (Kawasaki and Landrum 1997; Lucas and Bünger 2015; de Araújo et al. 2019). It has been reported as one of the dominant woody-species family in the Atlantic Forest of Brazil, particularly with endemic species belonging to the *Eugenia*, *Campomanesia*, *Psidium* and *Myrciaria* genera (Lucas and Bünger 2015; de Araújo et al. 2019). The Brazilian Atlantic Forest is considered one of the global priority-hot-spots for biodiversity, supporting *ca*. 14,000 plant species, half of which are endemic, and highly threatened by urbanization, agriculture, logging, and mining (Lucas and Bünger 2015).

Species from various Myrtaceae genera are cultivated in many parts of the world for their highly appreciated fruits, some of global importance, such as guava (*Psidium guajava*), or of regional importance in Brazil, such as ‘pitanga’ (*Eugenia uniflora*), ‘jabuticaba’ (*Myrciaria cauliflora*), and ‘jambo’ (*Syzygium jambos*) (Nogueira et al. 2016). Additional fruit-bearing species, such as ‘uvaia’ (*Eugenia pyriformis* Cambess.), ‘grumixama’ (*E. brasiliensis* Lam.), ‘Rio-Grande cherry’ (*E. involucrata* DC.) and ‘cambuci’ (*Campomanesia phaea* (O. Berg) Landrum) are locally available, particularly in the Brazilian Atlantic Forest biome, with great potential to be developed as new crops (Nogueira et al. 2016; de Araújo et al. 2019). These fruit species occur naturally, with commercial cultivation stimulated by an incipient but fast-growing industry.

The genus *Campomanesia* is of exclusive South American occurrence, and contains around 25 species (Landrum 1986). *Campomanesia phaea* yields fruits with fleshy pulp, with an intense and unique aroma

(Vallilo et al. 2005; Sanches Azevedo et al. 2017). The low amount of carbohydrates and high acidity make the fruit little attractive for *in natura* consumption, however, they are valued for juices and other downstream uses (Vallilo et al. 2005). The leaves and bark are locally used to prepare infusions and syrups for alleged medical purposes (Tokairin et al. 2018). The species offers great commercial potential as an exotic fruit, and a small production-chain, focused on social-environmental sustainability, composed by small family growers and non-governmental organizations, has been established in cities located around São Paulo city (Tokairin et al. 2018). Cambuci trees occur spontaneously at low density in small groups in the slopes of remains of the Atlantic Forest in the states of Sao Paulo, Rio de Janeiro, and Minas Gerais (Landrum 1986; Lorenzi 1992; Kawasaki and Landrum 1997).

Little is known about the biology and lifestyle of *C. phaea*, its genetic diversity, and the species occurrence. Therefore, more information is required to develop a strategic program to conserve genetic resources and to sustainably explore this species (Bianchini et al. 2017). Floral biology and pollination in *C. phaea* have been investigated (Cordeiro et al. 2017). Morphological and quality differences among cambuci accessions (Bianchini et al. 2016), and the first attempts to start a breeding program have also been reported (Bianchini et al. 2017). On the other hand, there is no *in situ* or *ex situ* germplasm collection for cambuci, and some small growers currently cultivate and maintain a few accessions. *Campomanesia phaea* is considered a vulnerable species by the International Union for Conservation of Nature (IUCN, 2020).

We are interest in investigating the cambuci fruit quality and post-harvest attributes from available accessions to develop this species into a new fruit crop. To determine the accessions origin and genotype, we developed *C. phaea*-specific microsatellite markers and used these loci to characterize the genetic diversity and investigate the structure of five cambuci regional group of accession from São Paulo state, from sites around the presumed center of diversity of the species. We used this information to identify and estimate the minimum number of accessions in an attempt to establish a core-collection that encompasses the whole detected genetic diversity in these collections. The availability of genetic markers will also allow us to determine the cambuci reproductive system, and to estimate gene flow and mating system in natural stands. Together, this information will be instrumental to develop a breeding program and a strategy for species conservation.

Material And Methods

Plant material: Samples from 145 accessions were collected from cultivated stands (plant collections) from five regional group of accessions (Juquitiba, Paraibuna, Mogi das Cruzes, Ribeirão Pires, and Salesópolis; Figure 1), most of them from the remaining of the Atlantic Forest in São Paulo state.

DNA extraction: Leaf samples were collected in the field and stored in 96% ethanol until DNA extraction (Bressan et al. 2014). Total genomic DNA was extracted using a CTAB protocol described by Sereno et al. (2006). DNA was quantified by fluorimetry in a DyNA Quant 2000 fluorometer (Amersham Biosciences, Buckinghamshire, UK), and the quality was verified by gel electrophoresis.

Test of transferability: Five pairs of primers from *P. guajava*, developed by Risterucci et al. (2005), were selected for the test of transferability, as follows: mPgCIR11, mPgCIR13, mPgCIR16, mPgCIR25, and mPgCIR26. Five cambuci selections (from Paraibuna collection) were used in this test. PCR reactions had 25 ng of DNA, Taq Buffer 10X, 100 µM of each dNTP, $MgCl_2$ 1.5 mM, 0.2 µM of each initiator, and 1 U of Taq polimerase (Fermentas) in a total volume of 20 µL. The reactions of amplification were performed by an initial denaturation at 94ºC for 3 min, followed by 35 cycles of 30 s at 94ºC, 1 min in a gradient temperature of 44ºC to 54°C, 1 min at 72ºC, and a final step at 72ºC for 7 min. PCR products were separated in an 1.5% agarose gel at 5 V cm^{-1}, stained with Ethidium bromate.

Microsatellite library construction: The genomic library enriched for microsatellites was developed based on Billotte et al. (1999), with a few modifications. Total genomic DNA (5 µg) from accession 52 (Paraibuna, SP, Brazil) was digested with 60 U *Rsa* I (Fermentas, Burlington, Canada), plus 400 mM spermidine in a final volume of 100 µL overnight at 37 °C. Specific adapters *Rsa* 21 and *Rsa* 25 (Billotte et al. 1999) were ligated to the digested DNA (1 µg) using 5 U T4 DNA ligase (Promega, Madison, WI, USA), with 0.2 µM of each adapter in 200 µL at 20 °C for 2 h. For enrichment, biotinylated oligonucleotide [$(CT)_8$ and $(GT)_8$] probes were used to capture amplified genomic fragments with complementary microsatellite sequences (Kijas et al. 1994). Amplified fragments were then cloned into pGEM-T Easy Vector (Promega), and the ligation products were used to transform electrocompetent *Escherichia coli* (DH10B) cells. Positive white colonies were picked and transferred to 96-well plates containing selective Luria–Bertani media plus 8% glycerol, grown for 24 h at 37°C, and stored at −80°C. Plasmids from one 96-well plate were purified using alkaline lysis followed by filtration with a Millipore (MAGV N22; EMD Millipore; Billerica, MA, USA) filter. Plasmids were sequenced using BigDye Terminator v.3.1 sequencing kit (Thermo Fisher Scientific; Waltham, MA, USA), and analyzed in an ABI 3500 Genetic Analyzer (Thermo Fisher Scientific). Sequence files were evaluated for quality using Phred/Phrap/Consed (http://www.phrap.org/index.html), and microsatellite repeat motifs were identified using WebSat (Martins et al. 2009). Specific primers were designed using Primer3 (Rozen and Skaletsky 2000) with default parameters (length 18 to 30 bp; 40-60% GC%; melting temperature T_m 50-60 °C; maximum T_m difference between primers <4 °C) to amplify fragments from 150 to 400 bp.

Microsatellite loci validation and analysis: Validation of primer-pairs was conducted using PCR reactions containing 25 ng DNA, 50 mM KCl, 10 mM Tris-HCl (pH 8.8), 0.8% Nonidet P40, 1.5 mM $MgCl_2$, 100 µM of each dNTP, 0.12 µM of each primer (Table 1) and 1 U *Taq* polymerase (Fermentas) in 20 µL. The PCR reactions were performed in a MyCycler thermocycler (Bio-Rad, Hercules, CA, USA), starting at 94 °C for 3 min, followed by 10 cycles of 40 s at 94°C; 40 s at temperatures decreasing from 60°C to 50°C by 1°C every cycle; and 60 s at 72°C, followed by 35 cycles of 40 s at 94°C, 40 s at 50°C, and 60 s at 72°C. Amplification products were separated and visualized in denaturing polyacrylamide gels (7% polyacrylamide; 7 M urea) run in 1× TBE (89 mM Tris base; 89 mM boric acid; 2 mM EDTA) at 50 W for 2.5 h, and visualized by silver staining (Creste et al. 2001).

ITS amplification: To amplify the Internal Transcribed Spacer (ITS) region from the ribosomal RNA gene, we used the universal primers *ITS1-18S* (CGTAACAAGGTTTCCGTAGG) and *ITS4* (TCCTCCGCTTATTGATATGC) (White et al., 1990). The amplification reactions contained 25 ng of DNA, 10X *Taq* buffer, 100 µM of each dNTP, 1.5 mM $MgCl_2$, 0.2 µM of each primer and 1 U de *Taq* polymerase (Fermentas) in 20 µL. The amplification reactions were run with an initial denaturation at 94ºC for 3 min, followed by 35 cycles of 30 s at 94ºC, 1 min at 58ºC, 1 min at 72ºC, and a final step at 72ºC for 7 min. Products were separated in 1.5% agarose gel at 5 V cm^{-1} stained with Ethidium Bromide. Amplified products were cut and purified from the gel using Illustra GFX PCR and Gel Band Purification Kit (GE Healthcare) following the recommendation of the manufacturer. The purified products were cloned into pGEM-T Easy (Promega). Purified plasmids were sequenced as described above. Chromatograms were analyzed for quality using the Sequencing Analysis Software (Applied Biosystems) and contigs were assembled using SeqScape® Software (Applied Biosystems).

Statistical analysis: The allelic data were analyzed using Genetic Data Analysis (GDA) (Lewis and Zaykin 2001). Allelic frequencies and distribution of the genetic diversity among and within populations were analyzed according to Nei (1973) and were estimated by FSTAT (Goudet 1995). The polymorphism information content (PIC) was estimated based on the number of detected alleles, distribution, and frequency in the structured population (Anderson et al. 1993). The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura et al. 2004) and are in the units of the number of base substitutions per site. The proportion of sites where at least 1 unambiguous base is present in at least 1 sequence for each descendent clade is shown next to each internal node in the tree. This analysis involved 18 nucleotide sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There was a total of 826 positions in the final dataset. Evolutionary analyses were conducted in MEGA X (Kumar et al. 2018).

Choice of accessions for core collection: To develop a germplasm collection that encompasses all the genetic diversity sampled in the five regional group of accessions, we analyzed the genotyping data with Darwin 6.0.14 (Perrier et al. 2003; Perrier and Jacqemoud-Collet, 2006). The dissimilarities were estimated using 10,000 bootstraps and transformed into Euclidean distances. After data transformation, the UnWeighted Neighbor-Joining approach was adopted to build a diversity tree for all genotypes. Then, we applied the maximum length subtree function to identify the core collection (Campoy et al. 2016). The core collection represents the maximum diversity of all evaluated populations, with minimum redundancy. The core collection contains the minimum number of genotypes that represent the maximum diversity of all populations (Egbadzor et al. 2014).

Results And Discussion

Enriched library and microsatellite loci identification

Transferability of microsatellite primers between species or congeners is possible (Fagundes et al. 2016). However, the transferability rate decreases proportionally to the phylogenetic divergence, and the

transference of primer-pairs tends to reduce the disclosure of polymorphisms (Alves et al. 2007). Since there was no microsatellite loci sequence available for *C. phaea* or any congener at the beginning of this work, we started by testing loci available for *P. guajava* (Risterucci et al. 2005). We tested a few loci for transferability for cambuci, but most of them did not amplify any product or amplified non-specific products despite attempts to optimize amplification conditions; only one locus (*mPgCIR11*) amplified a monomorphic allele with the expected size. Nonetheless, to be successful in primer transferability between congeners or more distantly related taxa as in this case, a large number of primers needs be evaluated. Nogueira et al. (2016) evaluated 158 microsatellite primer-pairs from *P. guajava* in 18 Myrtaceae fruit species, including *C. phaea* and another two *Campomanesia* (*C. guaviroba* and *C. hirsute*). Around 50% of the primer-pairs showed some amplification, with 64 loci considered to exhibit a good quality amplification in *C. phaea* (Nogueira et al. 2016). However, the transferability screening was based on detection of amplified products in 1.5% agarose gels stained with ethidium bromide, an insufficient condition to discriminate alleles or polymorphism. Primers originally developed for *Eucalyptus urophylla* and *E. grandis* (Grattapaglia et al. 2015) were tested in *C. adamantinum* and *C. pubescens,* two species from the Brazilian savanna biome ('Cerrado') (Miranda et al. 2016). From the 120 primer-pairs tested, 12 loci were successfully transferred to both *Campomanesia* species, and used to analyze the genetic diversity of two populations of each species.

Accordingly, we proceeded to develop a library enriched for microsatellite sequences of *C. phaea* using the protocol proposed by Billotte et al. (1999), with a few modifications. We obtained 192 clones, from which 96 were sequenced. Fourteen clones contained microsatellite sequences (14% enrichment), from which seven contained perfect dinucleotide and one perfect trinucleotide repeats (Table 1). From the 14 loci identified, we developed primers specific for each locus, and six revealed polymorphism (Table 1) for the five cambuci regional group of accessions analyzed. The loci were denominated *Cam.ph03*, *Cam.ph04*, *Camp.ph05*, *Cam.ph06*, *Cam.ph09*, and *Cam.ph13*. The remaining eight loci were monomorphic. The six loci disclosed 26 alleles for the 145 accessions tested, ranging from two (*Cam.ph05*) to six alleles (*Cam.ph03* and Cam.*ph06*) per locus (Table 1). The number of alleles and the profile of polymorphism can be associated with the composition of the repeat motif, the number of repeats, and the size of the microsatellite sequence (Kelkar et al. 2008). The interruption of the repeat tends to reduce the DNA polymerase slippage, which affects the level of polymorphism disclosed (Estoup and Cornuet 1999; Lia et al. 2007). The monomorphic nature of the locus *Camp.ph12* appears to derive from the type of the repeat motif (Table 1). On the other hand, loci with perfect di- and tri-nucleotide repeat motifs, with longer repeats, tend to reveal more and polymorphic alleles (Zalapa et al. 2012), such as *Camp.ph03* and *Camp.ph06* (Table 1), whereas loci with single nucleotide motif or di-nucleotides with fewer repeats (< 9) present monomorphic profiles.

The average PIC (Polymorphism Information Content) for the all the loci was 0.499, ranging from 0.341 (*Cam.ph05*) to 0.633 (*Cam.ph06*) (Table 2). Thus, all loci developed for *C. phaea* are considered informative, with *Cam.ph04*, *Cam.ph05* and *Cam.ph13* considered moderately informative (0.25 < PIC < 0.50), and loci *Cam.ph03*, *Cam.ph06*, and *Cam.ph09* as highly informative (PIC > 0.50), according to the classification of PIC proposed by (Botstein et al. 1980). No linkage disequilibrium was detected among

the six loci after the Bonferrroni correction for multiple tests (95%; α = 0.05). The genetic diversity of Nei (1973) is defined as the probability of two random gametes from a population having different alleles for a certain locus, which corresponds to H_e. H_o represents the real rate of heterozygous individuals from a specific population. The loci *Cam.ph03, Cam.ph05, Cam.ph06* and *Cam.ph13* revealed H_o lower than H_e, whereas, for *Cam.ph04* and *Cam.ph09*, H_o was superior to H_e (Table 2).

Overall, in the *C. phaea* accessions, the mean observed heterozygosity (Ho) was 0.55 and the expected heterozygosity (H_e) was 0.64 (Table 2). The analysis of genetic diversity of *C. adamantium* and *C. pubescens* using 12 microsatellite loci from *Eucalyptus* disclosed 82 alleles in *C. adamantium* individuals, with an average of 6.8 alleles per locus, and 95 alleles in *C. pubescens* ones, with an average of 7.8 alleles per locus, ranging from 2 to 16 alleles per locus (Miranda et al. 2016). The average values of H_e and H_o were 0.517 and 0.504 for *C. adamantium*, and 0.579 and 0.503 for *C. pubescens*, respectively (*n*=80). Seven *Eucalyptus* microsatellite loci were also used to analyze *C. adamantium* individuals from six sites in Mato Grosso do Sul (MS) and Goias states (GO), and one from Paraguay (*n*=208). The analysis revealed 71 alleles, with an average of 10 allele per locus (ranging 3 to 21 alleles), and the mean H_e and H_o was 0.62 and 0.61 for the 207 individuals (Crispim et al. 2018). Both studies using microsatellite loci from *Eucalyptus* disclosed more alleles per locus than the number found here for *C. phaea* likely because the analyses of polymorphism were conducted in automatic sequencers using fluorescence, which can be more accurate in detecting alleles. In another study with *C. adamantium* based on 36 polymorphic microsatellite loci specifically developed for this species, the number of alleles varied from 2 to 14 per locus (mean 8.14), whereas mean values of H_e and H_o were 0.46 and 0.52, respectively, but based on a small sample (*n*=10) (Crispim et al. 2019). In general, the overall levels of expected and observed heterozygosity were comparable among the *Campomanesia* species, mostly with higher H_e than H_o, but not by a large margin. The values of H_e and H_o suggest an outcrossing mode of reproduction.

Population Genetic Structure

The five regional collections exhibited an acceptable level of genetic diversity estimated by the percent of polymorphic loci, mean number of alleles locus^{-1}, and heterozygosity. Among the five regional cambuci collections, the average number of alleles per locus ranged from 3.33 (Salesópolis) to 4.33 (Mogi das Cruzes), with an overall mean of 3.83 alleles per locus (Table 3). The percent of polymorphic loci varied from 53% (Juquitiba) to 61% (Mogi das Cruzes), whereas the average among the collections was 57% (Table 3). All accession collections displayed higher H_e than H_o, except for the one from Ribeirão Pires (Table 3). The fixation index among the populations ranged from -0.13 to 0.20. According to the parameters, the genetic diversity of the populations ranged from 0.53 to 0.61 (Table 3). The higher number of polymorphic loci were detected in populations of Paraibuna and Mogi das Cruzes (Table 3). In these populations, a higher number of samples were collected from wild plants, in the forest, especially in Mogi das Cruzes, where collections were realized mostly within preserved or in an advanced stage of

recovery fragments of the Atlantic Forest. These results clearly indicate the importance of a solid action of preservation and/or recuperation of this environment in a high scale, in order to conserve the biodiversity of this biome. On the other hand, in populations of Salesópolis, Ribeirão Pires, and Juquitiba, in spite of some plants were assessed in the wild, there was a more significant proportion of plants collected in backyards, what may explain the lower number of polymorphic loci. However, even with lower genetic diversity, it is important to emphasize the relevance of individual actions of the local resident populations, which with their economic and cultural value given to this species, also contributed for its conservation.

The studies conducted with populations of other *Campomanesia* species using microsatellite markers revealed similar levels of diversity among natural populations. The populations of *C. adamantium* from Mineiros and Três Ranchos (H_e=0.531 and 0.504; H_o=0.504 and 0.505, respectively), and the ones of *C. pubescens* from Santa Rita do Araguaia and Caiapônia (H_e = 0.629 and 0.529; H_o = 0.498 and 0.507), all exhibited comparable levels of genetic diversity (Miranda et al. 2016). In another study, seven populations of *C. adamantium* from the states MS (four populations) and GO (two populations), and one from Paraguay were analyzed for diversity; the values of H_e and H_o varied from 0.44 to 0.64 and from 0.44 to 0.73 (Crispim et al. 2018). Using a set of primers developed specifically for *C. adamantium* to analyze three populations from MS and Paraguay, the levels of H_e estimated ranged from 0.58 to 0.63 and of H_o from 0.50 to 0.62 (Crispim et al. 2019).

The genetic diversity parameter of Nei (1973) for all the *C. phaea* accessions analyzed here can be considered low (HT' = 0.10) (Table 4). The genetic diversity among the population is low based on the statistics GST' = 0.19 and Θ = 0.09. The fixation indexes within populations were 0.57 (H_s) and 0.13 (*f*). In general, the regional groups of accessions did not show evidence on inbreeding. In comparison, the analyses of two populations of *C. adamantium* and *C. pubescens* each from GO indicated a significant structured diversity, with more diversity found within (~75%) than between populations in both species (Miranda et al. 2016). No inbreeding was detected in any *C. adamantium* and *C. pubescens* population, but higher genetic diversity was registered among individuals of *C. pubescens* when compared with individuals of *C. adamantium* (Miranda et al. 2016). When seven populations of *C. adamantium* from MS, GO, and Paraguay were investigated for population genetic structure, the AMOVA analysis indicated that the structure showed significance but with low diversity among populations (F_{ST} = 0.06), and more diversity within populations (Crispim et al. 2018). Therefore, in all cases for the *Campomanesia* species, most of the genetic diversity occurred within populations, with little variation among populations. Similar pattern with high genetic diversity within populations has been described for other Myrtaceae species, such as in three populations of *E. uniflora* (Ferreira-Ramos et al. 2008), and *E. dysenterica* (Zucchi et al. 2002). Conversely, 13 populations of the Myrtaceae camu-camu (*Myrciaria dubia*), native to the Amazon, analyzed by seven microsatellite loci indicated a significant deficit of heterozygotes (He=0.218 to 0.680; and Ho=0.137 to 0.527), with a high genetic diversity among the populations, but also a high degree of inbreeding within the populations (Šmíd et al. 2017).

Considering that cambuci flowers are self-incompatible (Cordeiro et al. 2017; Tokairin et al. 2018), and therefore, dependent on cross-pollination for fruit set, it might have a high allele share among individuals within a given population. Flowering occurs during warm and humid months, and in natural stands of cambuci, the synchrony of flowering among individual trees can reach 50% of the individuals, which may contribute to allele dispersion within populations (Cordeiro et al. 2017). Bees with nocturnal or crepuscular habits, such as *Megalopta sodalis*, *Megommation insigne*, *Ptiloglossa latecalcarata*, *Zikanapis seabrai*, and *Apis mellifera* are responsible for pollination of cambuci flowers (Cordeiro et al. 2017). Seed dispersal in *Campomanesia* appears to be predominantly performed by animals, particularly small primates (Gressler et al. 2006). Seeds are dispersed during the favorable period for germination (Van Schaik et al. 1993; Morellato and Leitao-Filho 1996; Tabarelli and Peres 2002). The pattern of flowering and seed dispersal favor to have more diversity within populations than among.

Phylogeny of Campomanesia phaea by ITS sequence

The genus *Campomanesia* is part of the tribe Myrteae, which holds most of the New World Myrtaceae including all Brazilian species (Landrum 1986). *Campomanesia* is a well-defined genus within the Myrtaceae, with distinctive morphological features (Landrum 1986; Luber et al. 2020). However, the distinction among *Campomanesia* species is more elusive, with some species showing large variation, but others exhibit little infraspecific morphological diversity (Landrum 1986).

To place the Brazilian *C. phaea* in the context of the genus, and to try to define the potential relationship of the species with the other congeners, we sequenced the ITS region of the ribosomal gene. The ribosomal gene contains highly conserved sequences (18S, 5.8S and 26S), separated by two more variable transcribed regions, named ITS1 and ITS2), useful markers to elucidate the evolutive history at various taxonomic levels (Hsaio et al. 1994). We amplified, cloned and sequenced fragments from two accessions (#20 and #41), originally collected in Paraibuna, SP. The sequenced fragments had 721 bp, and upon analysis confirmed the identity against National Center for Biotechnology Information (NCBI) as *Campomanesia* (GenBank id# MT433815 and MT433816). The sequenced products from both accessions of *C. phaea* (#20 and #41) were analyzed with other ITS sequences available at NCBI: *Campomanesia* sp. (AM234078.1), *C. laurifolia* (MK313875.1), *C. guazumifolia* 1 (MK313874.1), *C. guazumifolia* 2 (MG708054.1), *C. guazumifolia* 3 (AM234076.1), *C. ilhoensis* (MH445990.1), *C. xanthocarpa* (KF421011.1), *C. xanthocarpa* 1 (MG708055.1), *C. xanthocarpa* 2 (MG708055.1), *C. xanthocarpa* 3 (KF421010.1), *C. guaviroba* (MG707974.1), *C. hirsuta* (MG707973.1), *C. velutina* (MF954026.1), *C. adamantium* (MF954025.1), *C. pubescens* (AM234077.1), and the sequence from *Psidium guajava* (AY487283.1) was used as an outgroup.

The evolutionary history was inferred using the UPGMA method (Schlee et al. 1975). The optimal tree is shown (Figure 2). The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura et al. 2004) and are in the units of the number of base substitutions per site. The proportion of sites where at least one unambiguous base is present in at least 1 sequence for each

descendent clade is shown next to each internal node in the tree. This analysis involved 18 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There was a total of 826 positions in the final dataset. Evolutionary analyses were conducted in MEGA X (Kumar et al. 2018).

The clustering analysis formed two groups with strong branch support (89%), plus the outgroup species *P. guajava* (Figure 2). The first group contained two subgroups, supported by robust bootstrap (95%). The first subgroup contained both *C. phaea* accessions, clustered with robust branch support (88%) together with the samples of *C. hirsuta* and *C. laurifolia*, both species with a wide distribution in the Atlantic Forest, similar to *C. phaea* (Landrum 1986). The second subgroup (branch support of 86%) included *C. ilhoensis* and *C. guazumifolia*, species with current distribution in the Brazilian dry area ('caatinga'). This subgroup appears to contain species from the coast region of Brazil. The second subgroup showed a robust branch support (82%), and it contained an unknown *Campomanesia* species, *C. xanthocarpa*, *C. adamantium*, *C. velutina,* and *C. pubescens*, all Brazilian species that occur in the Brazilian savanna, denominated 'cerrado'. The results suggest that *C. phaea* probably evolved in the Atlantic Forest. The clear distinction between the species was not resolved by the ITS phylogeny since there was an overlap of grouping between *C. phaea* and *C. hirsuta*.

Choice of accessions for core collection

Ex-situ germplasm collection represents an important source for plant species conservation and breeding (Odong et al. 2013). For species with recalcitrant seeds, germplasm collection must be kept as living individuals, which can represent a large use of limited investment, since collections can reach large planting areas. Seeds from some Myrtaceae species appear to be recalcitrant, not tolerating desiccation (Maluf and Pisciottano-Ereio 2005). Regarding *C. phaea* seeds, there is limited information about long-term seed storage. In one study, *C. phaea* non-desiccated seeds maintained 100% germination rate for 180 days when stored at 8°C in plastic bags (Maluf and Pisciottano-Ereio 2005). Seed germination viability of non-desiccated seeds was observed up to 240 days at 8°C, while naturally seed drying reduced seedling? vigor. Thus, the seeds of *C. phaea* can be considered, at least, as partially orthodox, because they can be stored under low temperatures for about 240 days (Maluf and Pisciottano-Ereio 2005). However, for long-term conservation of a heterozygous out-crossing genotypes, an active germplasm collection must be preserved to keep the genetic identity of selected cultivars.

The maintenance and management of large germplasm collections are expensive and inefficient due to possible genotype redundancies and/or duplications, and because of the difficulty to perform a detailed evaluation of all the conserved individuals (Grenier et al. 2000). To improve the efficiency for maintenance of collections, and to establish a representative source of the genetic diversity, the concept of core collections was established (Frankel 1984). A core collection represents the maximum diversity from all the studied populations with the least number of individuals and redundancy (Egbadzor et al. 2014). To define the core collection, we used the sub-tree with maximum length implemented in the

Darwin package (Bernard et al. 2018). This method searches for a subgroup of genotypes, minimizing the redundancy among them and limiting the loss of diversity (Campoy et al. 2016).

The genotypes were chosen based on the genetic distance reflected in a diversity tree. The genotypes with higher genetic distances correspond to those that have higher quantities of uncommon characters, i.e., they are genetically distinct (Billot et al. 2013). The most distinct genotypes were identified using the removed edge value, the maximum length edge and the sphericity index. Campoy et al. (2016) used the reference value of 0.008 as the limit parameter for removed edge value to select individuals with minimal redundancy. Here, we were able to select 18 accessions to form the core collections (Table 5). This core collection represents 12% of the analyzed individuals, comprised of five individuals from Mogi das Cruzes (27, 42, 56, 75, and 77), four individuals from Juquitiba (101, 106, 109, and 112), four individuals from Paraibuna (8, 12, 14, and 18), three from Salesópolis (141, 145, and 146), and two from Ribeirão Pires (119 and 120).

Declarations

Funding: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq): MSc scholarship to the first author; Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Grant No. 2014/126063: financial support to this research.

Conflicts of interest/Competing interests: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this manuscript.

Availability of data and material (data transparency): The datasets of the current study are presented in this article or are available upon a reasonable request.

Code availability (software application or custom code): not applicable

Authors' contributions: Rafael Oliveira Moreira: Investigation, Data curation, Writing – Original Draft. Eduardo de Andrade Bressan: Investigation, Writing - Review & Editing. Horst Bremer Neto: Investigation, Data curation. Angelo Pedro Jacomino: Project administration, Funding acquisition. Antonio Figueira: Conceptualization, Writing - Review & Editing. Francisco de Assis Alves Mourão Filho: Conceptualization, Project administration, Writing - Review & Editing.

Acknowledgments

The authors thank Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for the MSc scholarship to the first author; Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Grant No. 2014/126063 for financial support to this research. The authors also thank Instituto H & H Fauser, Instituto Auá de Empreendedorismo Socioambiental, Instituto Ecofuturo/Parque das Neblinas, Parque Andreense, Parque Nascentes de Paranapiacaba, Instituto Florestal/ Secretaria de Infraestrutura e Meio Ambiente do Estado de São Paulo, for their technical support during this research.

References

1. Alves RM, Sebbenn AM, Artero AS, et al (2007) High levels of genetic divergence and inbreeding in populations of cupuassu (Theobroma grandiflorum). Tree Genet Genomes 3:289–298. https://doi.org/10.1007/s11295-006-0066-9
2. Anderson JA, Churchill GA, Autrique JE, et al (1993) Optimizing parental selection for genetic linkage maps. Genome 36:181–186. https://doi.org/10.1139/g93-024
3. Bernard A, Barreneche T, Lheureux F, Dirlewanger E (2018) Analysis of genetic diversity and structure in a worldwide walnut (Juglans regia L.) germplasm using SSR markers. PLoS One 13:e0208021. https://doi.org/10.1371/journal.pone.0208021
4. Bianchini FG, Balbi RV, Pio R, et al (2017) Parents choice and genetic divergence between cambuci fruit tree accessions. Crop Breed Appl Biotechnol 17:214–220. https://doi.org/10.1590/1984-70332017v17n3a33
5. Bianchini FG, Balbi RV, Pio R, et al (2016) Caracterização morfológica e química de frutos de cambucizeiro. Bragantia 75:10–18. https://doi.org/10.1590/1678-4499.096
6. Billot C, Ramu P, Bouchet S, et al (2013) Massive Sorghum Collection Genotyped with SSR Markers to Enhance Use of Global Genetic Resources. PLoS One 8:e59714. https://doi.org/10.1371/journal.pone.0059714
7. Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. 32:314–331
8. Bressan EA, Rossi ML, Gerald LT, Figueira A (2014) Extraction of high-quality DNA from ethanol-preserved tropical plant tissues. BMC Res Notes 7:. https://doi.org/10.1186/1756-0500-7-268
9. Campoy JA, Lerigoleur-Balsemin E, Christmann H, et al (2016) Genetic diversity, linkage disequilibrium, population structure and construction of a core collection of Prunus avium L. landraces and bred cultivars. BMC Plant Biol 16:49. https://doi.org/10.1186/s12870-016-0712-9
10. Cordeiro GD, Pinheiro M, Dötterl S, Alves-dos-Santos I (2017) Pollination of Campomanesia phaea (Myrtaceae) by night-active bees: a new nocturnal pollination system mediated by floral scent. Plant Biol 19:132–139. https://doi.org/10.1111/plb.12520
11. Creste S, Neto AT, Figueira A (2001) Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gels by silver staining. Plant Mol Biol Report 19:299–306. https://doi.org/10.1007/BF02772828
12. Crispim B, Bajay M, de Vasconcelos A, et al (2018) Relationship between Genetic Variability and Land Use and Land Cover in Populations of Campomanesia adamantium (Myrtaceae). Diversity 10:106. https://doi.org/10.3390/d10040106
13. Crispim B do A, Déo TG, Fernandes J dos S, et al (2019) Development and characterization of microsatellite markers in Campomanesia adamantium, a native plant of the Cerrado ecoregions of South America. Appl Plant Sci 7:. https://doi.org/10.1002/aps3.11287

14. de Araújo FF, Neri-Numa IA, de Paulo Farias D, et al (2019) Wild Brazilian species of Eugenia genera (Myrtaceae) as an innovation hotspot for food and pharmacological purposes. Food Res. Int. 121:57–72
15. Drewes H. and Sánchez L. (2017). The varying surface kinematics in Latin America: VEMOS 2009, 2015, and 2017. SIRGAS Symposium 2017. Mendoza, Argentina. Available at: www.sirgas.org Drewes H., Kuglitsch F., Ádám J., Rózsa S. (2016). The Geodesist's Handbook 2016. Journal of Geodesy. Vol. 90. Issue 10.
16. Egbadzor KF, Ofori K, Yeboah M, et al (2014) Diversity in 113 cowpea [Vigna unguiculata (L) Walp] accessions assessed with 458 SNP markers. J Korean Phys Soc 3:1–15. https://doi.org/10.1186/2193-1801-3-541
17. Fagundes BS, da Silva LF, Giacomin RM, et al (2016) Transferability of Microsatellite Markers Among Myrtaceae Species and Their use to Obtain Population Genetics Data to Help the Conservation of the Brazilian Atlantic Forest. Trop Conserv Sci 9:408–422. https://doi.org/10.1177/194008291600900122
18. Ferreira-Ramos R, Laborda PR, de Oliveira Santos M, et al (2008) Genetic analysis of forest species Eugenia uniflora L. through of newly developed SSR markers. Conserv Genet 9:1281–1285. https://doi.org/10.1007/s10592-007-9458-0
19. Frankel OH (1984) Genetic perspectives of germplasm conservation. In: Genetic manipulation: impact on man and society. pp 161–170
20. Goudet J (1995) FSTAT, a program to estimate and test gene diversities and fixation indices. J Hered 86:
21. Grattapaglia D, Mamani EMC, Silva-Junior OB, Faria DA (2015) A novel genome-wide microsatellite resource for species of Eucalyptus with linkage-to-physical correspondence on the reference genome sequence. Mol Ecol Resour 15:437–448. https://doi.org/10.1111/1755-0998.12317
22. Grenier C, Bramel-Cox PJ, Noirot M, et al (2000) Assessment of genetic diversity in three subjects constituted from the ICRISAT sorghum collection using random vs non-random sampling procedures: A. Using morpho-agronomical and passport data. Theor Appl Genet 101:190–196. https://doi.org/10.1007/s001220051468
23. Gressler E, Pizo M a., Morellato LPC (2006) Polinização e dispersão de sementes em Myrtaceae do Brasil. Rev Bras Botânica 29:509–530. https://doi.org/10.1590/S0100-84042006000400002
24. Hsaio C, Chatterton NJ, Asay KH, Jensen KB (1994) Phylogenetic relationships of 10 grass species: An assessment of phylogenetic utility of the internal transcribed spacer region in nuclear ribosomal DNA in monocots. Genome 37:112–120. https://doi.org/10.1139/g94-014
25. Kawasaki ML, Landrum LR (1997) A rare and potentially economic fruit of Brazil: Cambuci, Campomanesia phaea (Myrtaceae). Econ. Bot. 51:403–405
26. Kelkar YD, Tyekucheva S, Chiaromonte F, Makova KD (2008) The genome-wide determinants of human and chimpanzee microsatellite evolution. Genome Res 18:30–38. https://doi.org/10.1101/gr.7113408

27. Kijas JM, Fowler JC, Garbett CA, Thomas MR (1994) Enrichment of microsatellites from the citrus genome using biotinylated oligonucleotide sequences bound to streptavidin-coated magnetic particles. Biotechniques 16:656–60, 662
28. Kumar S, Stecher G, Li M, et al (2018) MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35:1547–1549. https://doi.org/10.1093/molbev/msy096
29. Landrum LR (1986) Campomanesia, Pimenta, Blepharocalyx, Legrandia, Acca, Myrrhinium, and Luma (Myrtaceae). In: Flora Neotropica Monogr. pp 1–178
30. Lewis, Zaykin D (2000) Genetic Data Analysis: Computer Program for the Analysis of Allelic Data, Version 1.0. (d15)
31. Lorenzi H (1992) Árvores brasileiras: manual de identificação e cultivo de plantas arbóreas nativas do Brasil, 2a ed. Instituto Plantarum de Estudos da Flora, Nova Odessa SP Brazil
32. Luber J, Christ JA, Ferreira MF da S, Carrijo TT (2020) Species delimitation within Campomanesia (Myrtaceae) using morphology and amplification profiles of microsatellite markers. Rev Bras Bot 43:131–137. https://doi.org/10.1007/s40415-020-00583-x
33. Lucas EJ, Bünger MO (2015) Myrtaceae in the Atlantic forest: their role as a 'model' group. Biodivers Conserv 24:2165–2180. https://doi.org/10.1007/s10531-015-0992-7
34. Maluf AM, Pisciottano-Ereio WA (2005) Desiccation and storage of cambuci seeds. Pesqui Agropecu Bras 40:707–714. https://doi.org/10.1590/s0100-204x2005000700012
35. Martins WS, Lucas DCS, Neves KFS, Bertioli DJ (2009) WebSat - A Web Software for MicroSatellite marker development. Bioinformation 3:282–283. https://doi.org/10.6026/97320630003282
36. Miranda EAGC, Boaventura-Novaes CRD, Braga RS, et al (2016) Validation of EST-derived microsatellite markers for two cerrado-endemic Campomanesia (Myrtaceae) species. Genet Mol Res 15:1–6. https://doi.org/10.4238/gmr.15017658
37. Morellato PC, leitao-Filho HF (1996) Reproductive Phenology of Climbers in a Southeastern Brazilian Forest. Biotropica 28:180. https://doi.org/10.2307/2389073
38. Nogueira AM, Ferreira A, da Silva Ferreira MF (2016) Transferability of Microsatellites from Psidium guajava to Eugenia, Myrciaria, Campomanesia, and Syzygium Species (Myrtaceae). Plant Mol Biol Report 34:249–256. https://doi.org/10.1007/s11105-015-0921-7
39. Odong TL, Jansen J, van Eeuwijk FA, van Hintum TJL (2013) Quality of core collections for effective utilisation of genetic resources review, discussion and interpretation. Theor Appl Genet 126:289–305. https://doi.org/10.1007/s00122-012-1971-y
40. QGIS Development Team, 2021. QGIS Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.osgeo.org
41. Risterucci AM, Duval MF, Rohde W, Billotte N (2005) Isolation and characterization of microsatellite loci from Psidium guajava L. Mol Ecol Notes 5:745–748. https://doi.org/10.1111/j.1471-8286.2005.01050.x

42. Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132:365–386. https://doi.org/10.1385/1-59259-192-2:365
43. Sanches Azevedo MC, Silva RRE, Jacomino AP, Genovese MI (2017) Physicochemical variability of cambuci fruit (Campomanesia phaea) from the same orchard, from different locations and at different ripening stages. J Sci Food Agric 97:526–535. https://doi.org/10.1002/jsfa.7756
44. Schlee D, Sneath PHA, Sokal RR, Freeman WH (1975) Numerical Taxonomy. The Principles and Practice of Numerical Classification. Syst Zool 24:263. https://doi.org/10.2307/2412767
45. Šmíd J, Kalousová M, Mandák B, et al (2017) Morphological and genetic diversity of camucamu [Myrciaria dubia (Kunth) McVaugh] in the Peruvian Amazon. PLoS One 12:1–15. https://doi.org/10.1371/journal.pone.0179886
46. Tabarelli M, Peres CA (2002) Abiotic and vertebrate seed dispersal in the Brazilian Atlantic forest: Implications for forest regeneration. Biol Conserv 106:165–176. https://doi.org/10.1016/S0006-3207(01)00243-9
47. Tamura K, Nei M, Kumar S (2004) Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc Natl Acad Sci U S A 101:11030–11035. https://doi.org/10.1073/pnas.0404206101
48. Tokairin T de O, Silva APG da, Spricigo PC, et al (2018) Cambuci: a native fruit from the Brazilian Atlantic forest showed nutraceutical characteristics. Rev Bras Frutic 40:. https://doi.org/10.1590/0100-29452018666
49. Vallilo MI, Garbelotti ML, Oliveira EDE, et al (2005) Características físicas e químicas dos futos do cambucizeiro (Campomanesia phaea). Rev Bras Frutic 27:241–244. https://doi.org/http://dx.doi.org/10.1590/S0100-29452005000200014
50. van Schaik CP, Terborgh JW, Wright SJ (1993) The Phenology of Tropical Forests: Adaptive Significance and Consequences for Primary Consumers*. Annu Rev Ecol Syst 24:353–377. https://doi.org/10.1146/annurev.es.24.110193.002033
51. Zalapa JE, Cuevas H, Zhu H, et al (2012) Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. Am. J. Bot. 99:193–208
52. Zucchi MI, Brondani RP V., Pinheiro JB, et al (2002) Transferability of microsatellite markers from Eucalyptus spp. to Eugenia dysenterica (Myrtaceae family). Mol Ecol Notes 2:512–513. https://doi.org/10.1046/j.1471-8286.2002.00297.x

Tables

Table 1. Characterization of 14 microsatellite *loci* for cambuci (*Campomanesia phaea*).

Locus	Sequence (5’ – 3’)	Motif	Tm (°C)	Size (bp)	NCBI sequence id
Cam.ph01	ATGATGATAGGGTGCCAAAGAG ACGTCGATTTGTAGGGGATTTA	$(A)_{13}$	57°C	271	MT010565
Cam.ph02	GCGTCTCTGGGATTTATATGGA AAGGGAAACCGGAAAGTATTGT	$(AG)_7$	57°C	343	MT010566
Cam.ph03	ACAGATGTCTCGATGGAATCAC CGCCTTTCTTTTATTGGAGAAG	$(CT)_{13}$	59°C	251	MT010567
Cam.ph04	GTCGGAGGTTTTAAGGTTGAGA CTCTCTCTCTCTCTTGCGATCC	$(AGG)_6$	55°C	209	MT010568
Cam.ph05	GAGCTGTTGCTCAGTTCCG CTTTGACGAGCTTTCTGGTTCT	$(TC)_9$	57°C	288	MT010569
Cam.ph06	CTATCATCGGAGCGAGCTTTAC GGACTAACTCTGCTTTTCCCCT	$(AG)_{13}$	55°C	256	MT010570
Cam.ph08	TTGAGTCCAGGGATTGTCTATG ATCTGAACAAGATTCGACTTGC	$(T)_{12}$	55°C	127	MT010571
Cam.ph09	GAGGACCAGCAATACAGAGAGG TAGAAAACCAGGGAGTTACCGA	$(CT)_{12}$	59°C	372	MT010572
Cam.ph10	GACAGAAGCCCCATATTTTCAT AAATCTAGTCCCAACCAAGCAA	$(T)_{10}$	55°C	302	MT010573
Cam.ph11	TGCTTGGTTGGGACTAGATTTT CATAATTCACGAGTGGTCAGGA	$(A)_8$	55°C	162	MT010573
Cam.ph12	AACTCGGGAGCTTTTGTGATAC AAGAGGACGAGATGTAGGCAG	$(T)_{13}CGTTTG(A)_{11}$	55°C	310	MT010574
Cam.ph13	ATATGGGGTGCGTTTATTCATC ATGAATAAACGCACCCCATATC	$(CT)_9$	55°C	213	MT010575
Cam.ph15	TAATGGGTATTGCTTTTGAGGG ACTCTACTTCTCACCACGGCAT	$(T)_{11}$	55°C	342	MT010576
Cam.ph16	AGAGGGGTACTTTTCGGTTTTC GCTTCTTGGACAGCATAGGAAT	$(GA)_9$	55°C	258	MT010577

Table 2. Characterization of the loci and parameters of genetic diversity, number of alleles per locus (*A*), observed heterozygosity (H_o), expected heterozygosity (H_e), and Polymorphism Information Content (PIC) in 145 accessions of cambuci (*Campomanesia phaea*).

Locus	Allele sizes (bp)	*A*	H_o	H_e	PIC
Cam.ph03	166 - 191	6	0.50	0.66	0.517
Cam.ph04	197 - 218	4	1.00	0.60	0.490
Cam.ph05	218 - 225	2	0.35	0.46	0.341
Cam.ph06	236 - 256	6	0.52	0.76	0.633
Cam.ph09	295 - 326	4	0.69	0.65	0.564
Cam.ph13	266 - 278	4	0.27	0.69	0.454
Average	**-**	**4.33**	**0.55**	**0.64**	**0.499**

Table 3. Parameters of genetic diversity, including the sampled group of accessions, number of evaluated accesses (***n***), percentages of polymorphic loci (***P***), average number of alleles per locus (***A***), observed heterozygosity (H_o), expected heterozygosity (H_e), and fixation index (***f***), based on microsatellites evaluated in five cambuci (*Campomanesia phaea*) regional group of accessions.

Regional group of accessions	*n*	*P (%)*	*A*	H_o	H_e	*f*
Paraibuna	52	0.60	4.16	0.53	0.60	0.11
Mogi das Cruzes	47	0.61	4.33	0.58	0.61	0.05
Juquitiba	14	0.53	3.66	0.42	0.53	0.20
Ribeirão Pires	22	0.57	3.66	0.65	0.57	-0.13
Salesópolis	10	0.55	3.33	0.54	0.55	0.02
Mean	29	0.57	3.83	0.54	0.57	0.05

Table 4. Diversity within populations (H_s), total diversity (H_T), proportion of the genetic diversity between populations (G_{ST}), total fixation index (F), divergence between populations (Θ), fixation index (f), estimated based on sic microsatellite loci in five regional group of accessions of cambuci (*Campomanesia phaea*).

Locus	H_s	$H_{T'}$	$G_{ST'}$	F	Θ	f
Cam.Ph03	0.58	0.08	0.14	0.27	0.10	0.25
Cam.Ph04	0.58	0.05	0.08	-0.64	0.05	-0.66
Cam.Ph05	0.45	0.06	0.12	0.26	0.06	0.23
Cam.Ph06	0.69	0.08	0.15	0.32	0.08	0.31
Cam.Ph09	0.64	0.10	0.19	0.03	0.10	-0.05
Cam.Ph13	0.50	0.24	0.50	0.62	0.17	0.60
Average	**0.57**	**0.10**	**0.19**	**0.16**	**0.09**	**0.13**

Table 5. Selected individuals of cambuci (*Campomanesia phaea*) to comprise a central population, according to parameters of genetic diversity and minimum redundancy, calculated by DARWIN program.

Individual	Accession Group	External edges	Current tree length
8	Paraibuna	9.14805	15.99264
12	Paraibuna	9.32629	15.99264
14	Paraíbuna	9.11246	15.99264
18	Paraibuna	9.36332	15.99264
27	Mogi das Cruzes	9.21778	15.99264
42	Mogi das Cruzes	9.50135	15.99264
56	Mogi das Cruzes	9.42409	15.99264
75	Mogi das Cruzes	9.30516	15.99264
77	Mogi das Cruzes	9.30516	15.99264
101	Juquitiba	9.86387	15.99264
106	Juquitiba	9.58960	15.99264
109	Juquitiba	9.92272	15.99264
112	Juquitiba	9.81931	15.99264
119	Ribeirão Pires	9.65264	15.99264
120	Ribeirão Pires	9.68294	15.99264
141	Salesópolis	10.03181	15.99264
145	Salesópolis	10.03181	15.99264
146	Salesópolis	10.15388	15.99264

Figures

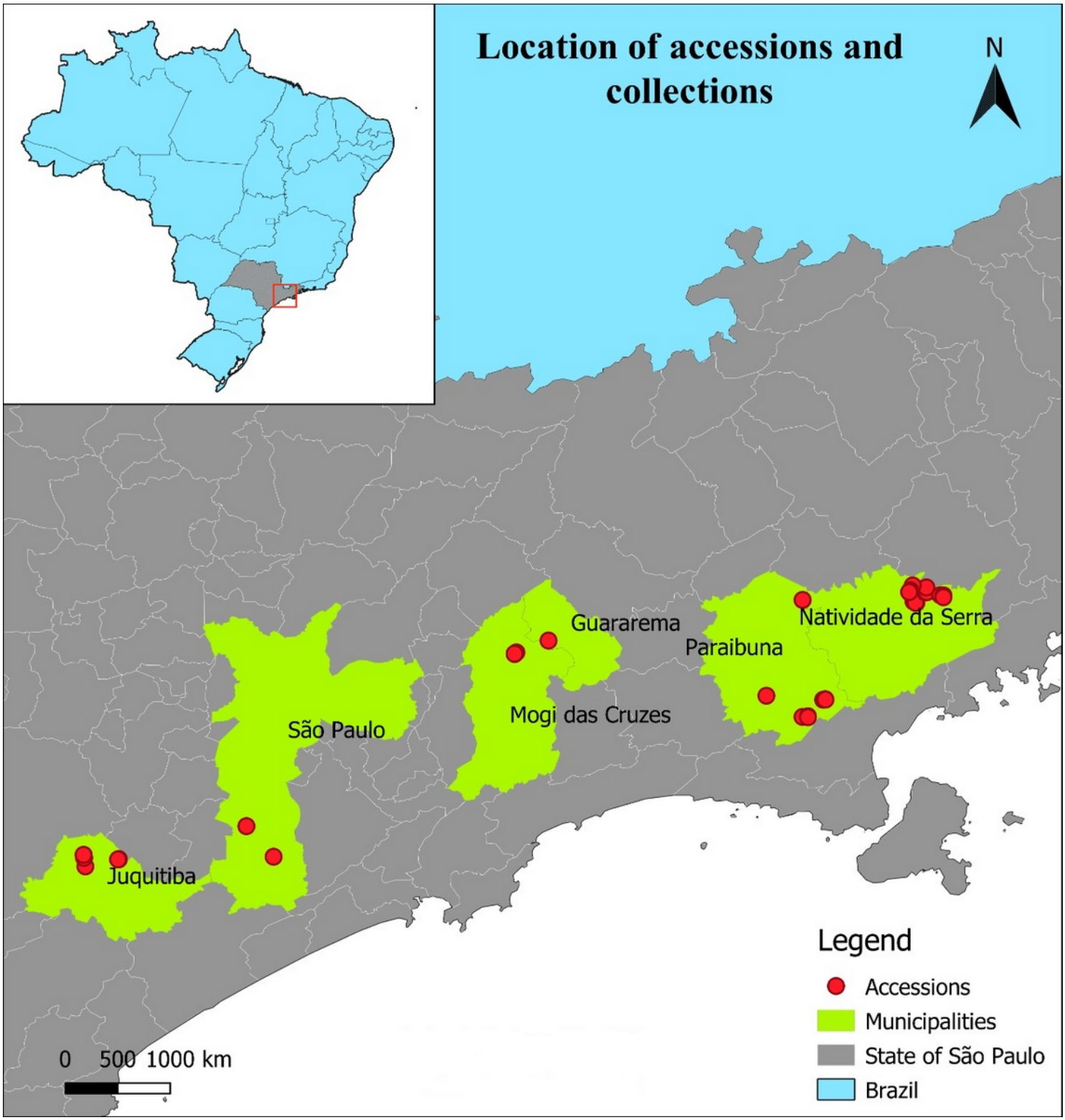

Figure 1

Geographic location of collected accessions. The map is based on the Coordinate System SIRGAS (Drewes H. and Sánchez L. (2017)), and the image was generated using Qgis project (2021).

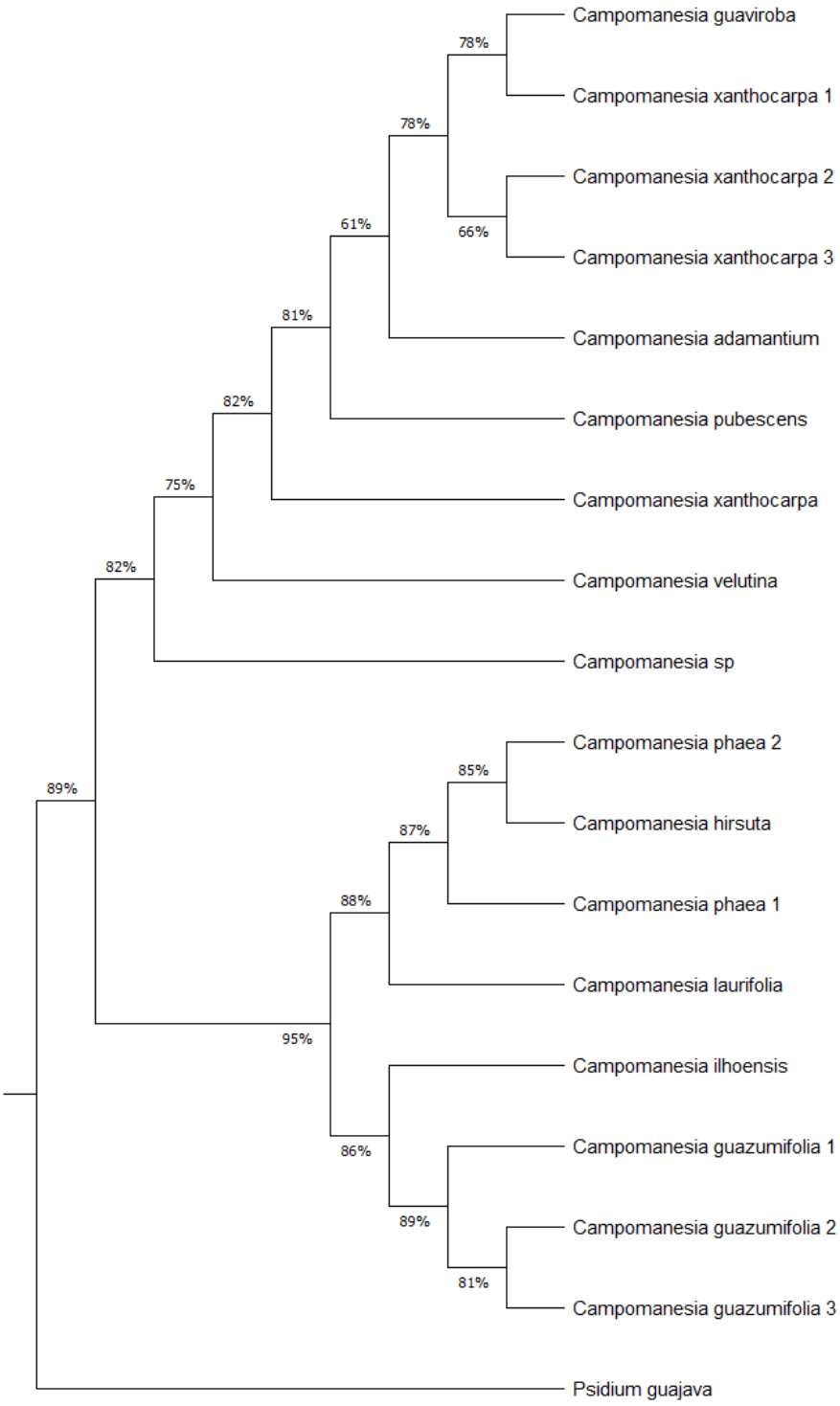

Figure 2

Phylogenetic analysis of 11 species da família de Campomanesia, including Psidium guajava as an outgroup.