Genetica (2014) 142:563–573
DOI 10.1007/s10709-014-9804-3

Genetic diversity of wild germplasm of "yerba mate" (*Ilex paraguariensis* St. Hil.) from Uruguay

Jimena Cascales · Mariana Bracco · Lidia Poggio · Alexandra Marina Gottlieb

Received: 8 August 2014 / Accepted: 22 November 2014 / Published online: 27 November 2014

Abstract The "yerba mate" tree, *Ilex paraguariensis* St. Hil., is a crop native to subtropical South America, marketed for the elaboration of the highly popular "mate" beverage. The Uruguayan germplasm occupies the southernmost area of the species distribution range and carries adaptations to environments that considerably differ from the current production area. We characterized the genetic variability of the germplasm from this unexplored area by jointly analyzing individuals from the diversification center (ABP, Argentina, Brazil and Paraguay) with 19 nuclear and 11 plastidic microsatellite markers. For the Uruguayan germplasm, we registered 55 alleles (18 % private), and 80 genotypes (44 % exclusive), whereas 63 alleles (28.6 % private) and 81 genotypes (42 % exclusive) were recorded for individuals from ABP. Only two plastidic haplotypes were detected. Distance-based and multilocus genotype analyses showed that individuals from ABP intermingle and that the Uruguayan germplasm is differentiated in three gene-pools. Significant positive correlations between genetic and geographic distances were detected. Our results concur in that ABP individuals harbor greater genetic variation than those from the tail of the distribution, as to the number of alleles (1.15-fold), H_e (1.19-fold), R_s (1.39-fold), and the between-group genetic distances (1.16-fold). Also the shape of the genetic landscape interpolation analysis suggests that the genetic variation decays southward towards the Uruguayan territory. We showed that Uruguayan germplasm hosts a combination of nuclear alleles not present in the central region, constituting a valuable breeding resource. Future conservation efforts should concentrate in collecting numerous individuals of "yerba mate" per site to gather the existent variation.

Keywords *Ilex paraguariensis* · Nuclear and plastidic SSR · Genetic diversity · Uruguayan accessions

Introduction

The tree crop *Ilex paraguariensis* St. Hil. (Aquifoliaceae Bartl.), commonly known as the "yerba mate" tree, is a dioecious perennial, native to subtropical regions of South America. Its natural area of distribution is restricted to north eastern Paraguay, the southern states of Brazil, the north eastern region of Argentina, and reaches Uruguay solely as patchy populations (Giberti 2001; Fig. 1). In the central area of the species distribution, these trees are usually found among specimens of *Araucaria angustifolia* (Bertol.) Kuntze. As a subtropical plant, it requires high temperatures (20–23 °C annual average) and high soil and environmental humidity; its rainfall requirements are about 1,500–1,800 ml/year, mostly from September to February. It thrives in permeable deep red sandy-clayey soils, rich in phosphoric acid, potassium and iron oxide, where water does not stagnate.

Electronic supplementary material The online version of this article (doi:10.1007/s10709-014-9804-3) contains supplementary material, which is available to authorized users.

J. Cascales · M. Bracco · L. Poggio · A. M. Gottlieb (✉)
Laboratorio de Citogenética y Evolución (LaCyE), Departamento de Ecología, Genética y Evolución, Instituto IEGEBA (CONICET-UBA), Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Intendente Güiraldes y Costanera Norte s/n, 4to piso, Pabellón II, Ciudad Universitaria, C1428EHA Ciudad Autónoma de Buenos Aires, Argentina
e-mail: gottlieb@ege.fcen.uba.ar

J. Cascales · M. Bracco · L. Poggio · A. M. Gottlieb
Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Avenida Rivadavia 1917, C1033AAJ Ciudad Autónoma de Buenos Aires, Argentina

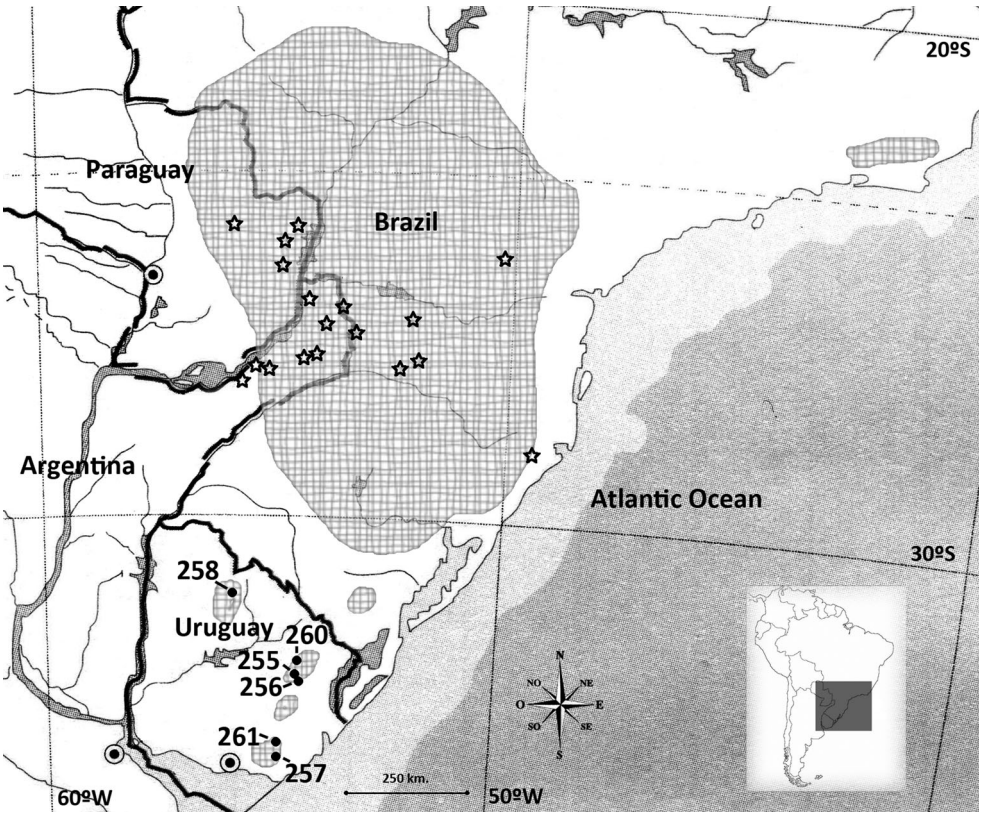

Fig. 1 Natural distribution area of *I. paraguariensis* (*in grey*), according to Giberti (2001) and Grela (2004), and collection sites of the trees analyzed in the present study. The *stars* denote materials from Argentina, Brazil and Paraguay; whereas the *dots* indicate materials from Uruguay. For details, see Table 1

Nowadays, natural populations of *I. paraguariensis* are found growing wild only in Brazil, Paraguay and Uruguay (Giberti 2011b). The Argentinean germplasm of "yerba mate" is currently maintained through institutional in situ conservation efforts, as natural populations are practically inexistent. Unlike other *Ilex* species, its economic value relies on the usage of their leaves and twigs to prepare a very popular infusion called "mate", which has a great socio-cultural significance in the region.

Argentina generates 62 % of the world production of "yerba mate", followed by Brazil (34 %) and Paraguay (4 %). These countries, together with Uruguay—a non-producer—, are the biggest consumers. The commercial exploitation of "yerba mate" in Argentina is restricted to the Provinces of Misiones and Corrientes (200,000 ha), due to the climate and soil requirements of the crop. There, the industrial processing of "yerba mate" represents a U$S 500 million annual income and involves approximately 30,000 families, from planting and harvesting to the point of sale (INYM, Instituto Nacional de la Yerba Mate 2014). The Brazilian production of "yerba mate" is limited to the states Rio Grande do Sul, Paraná, Santa Catarina and Mato Grosso do Sul (71,300 ha; IBRAMATE, Instituto Brasileiro da Erva-Mate 2014). In recent years overseas commerce has increased due to the assimilation of the habit of drinking "mate" in Middle East countries.

Several medicinal and stimulating properties are attributed to the high concentrations of methylxanthines (such as caffeine), flavonoids, triterpene saponins, polyphenols, minerals and vitamins, detected in the aqueous "mate" infusions (Filip et al. 2009; Anesini et al. 2012). Consequently, phytochemical studies abound in the literature (see Heck and De Mejia 2007, for a review). In contrast, the scarcity of genetic knowledge on the crop hampers its improvement and artificial selection with modern approaches.

A close relationship of *Ilex paraguariensis* trees sampled from Argentina, Brazil and Paraguay has been shown with AFLP marker analysis (Gottlieb et al. 2011). The natural populations of "yerba mate" from Brazil showed a narrow genetic base (Gauer and Cavalli-Molina 2000; Neumann-Wendt 2005; Pereira et al. 2013). Notably, nothing is known about the germplasm that thrives in the Uruguayan territory, which represents the southernmost area of the distribution range of the species (Giberti 2011a, b; Grela 2004). This country encompasses part of the Paranaense phytogeographic Province and of a transitional region with the Chaco phytogeographic Province (Grela 2004). The relevance of the Uruguayan "yerba mate" germplasm relies on its adaptation to local environments that considerably differ from the current productive area in terms of soil composition, temperature and precipitation

Table 1 Accessions of "yerba mate" surveyed in this study

Accession	Number of individuals	Georeference coordinates	Geographic origin
CA1/74	1	26°56′S 54°24′W	Selected progeny. Misiones, Argentina
CA8/74	1	26°56′S 54°24′W	Selected progeny. Misiones, Argentina
CA11/75	1	26°56′S 54°24′W	Selected progeny. Misiones, Argentina
CA51/75	1	26°56′S 54°24′W	Selected progeny. Misiones, Argentina
1	2	26°56′S 54°24′W	San Vicente, Guaraní, Misiones, Argentina[b]
34	2	26°53′S 54°15′W	San Vicente, Guaraní, Misiones, Argentina
39	1	26°34′S 54°01′W	San Pedro, San Pedro, Misiones, Argentina
47	1	26°25′S 53°38′W	B. de Irigoyen, Gral. M. Belgrano, Misiones, Argentina
50	3	27°15′S 55°32′W[a]	Paraje Colonia, Alberdi, San Ignacio, Misiones, Argentina
51	2	27°19′S 55°03′W[a]	Campo Viera, Oberá, Misiones, Argentina
195	1	27°27′S 55°44′W[a]	Candelaria, Candelaria, Misiones, Argentina
220	1	26°11′S 54°24′W	Puerto Esperanza, Iguazú, Misiones, Argentina
27	1	26°52′S 52°25′W	Xanxeré, Santa Catarina, Brazil[c]
28	2	27°06′S 52°37′W	Chapecó, Santa Catarina, Brazil
30	1	26°17′S 53°38′W	Dionísio Cerqueira, Santa Catarina, Brazil
100	1	25°03′S 51°31′W[a]	Turvo, Paraná, Brazil
92	1	26°14′S 52°21′W	Clevelandia, Paraná, Brazil
217	1	25°06′S 50°56′W[a]	Ivaí, Paraná, Brazil
224	1	24°50′S 54°40′W	San Alberto, Alto Paraná, Paraguay[c]
254	1	25°30′S 54°39′W[a]	Caaguazú, Caaguazú, Juan E. O'Leary, Ciudad Del Este, Paraguay
261	10	33°55′S 54°42′W	Lavalleja, Uruguay[e]
257	10	34°15′S 54°40′W	Maldonado, Uruguay
258	10	31°38′S 56°02′W	Tacuarembó, Uruguay
255	10	33°09′S 54°26′W	Treinta y Tres, Uruguay
256	10	33°11′S 54°25′W	Treinta y Tres, Uruguay
260	8	32°56′S 54°28′W	Treinta y Tres, Uruguay

[a] Coordinates inferred using Google Earth version 7.1.2.2041 (2013)

[b] Municipality, Department, Province, Country

[c] Municipality, State, Country

[d] Municipality, Department, Country

[e] Department, Country

regimes. Therefore, it constitutes a source of genetic variability with potentiality for genetic improvement programs. The assessment of the genetic variability is the first step for any plant breeding program. A comprehensive study of this genetic breeding resource will allow detecting and transferring agronomic features that may permit the expansion of the cultivation area further towards the South.

In this context, and aiming at contributing to the knowledge of the genetic variability harbored in *Ilex paraguariensis* trees growing wild in the tail of the species distribution area, we characterized Uruguayan samples with nuclear and plastidic microsatellite markers. We seek to answer the following questions: (1) does the Uruguayan germplasm exhibit a lower genetic diversity than that from the central area? (2) Does the Uruguayan germplasm constitute a single gene pool?

Materials and methods

We studied a set of 58 individuals from Uruguay, and 26 from Argentina, Brazil and Paraguay (hereafter, ABP), including selected samples (Table 1; Fig. 1). This material was provided by the "Banco de Germoplasma de Yerba Mate", at the "Estación Experimental Agropecuaria INTA Cerro Azul" (EEA-INTA-CA; Misiones, Argentina). With the exception of the selected progenies, all other materials derive from a pool of seeds gathered at each location, from several mother trees. Young leaves, preserved in silica-gel, were used for the isolation of total genomic DNA with the DNeasy Plant Quick Extraction kit (QIAgen Inc., Duesseldorf, Germany), following manufacturer's instructions.

Based on the number of alleles reported by Torimaru et al. (2004, 2007) and by Sosa et al. (2009) for other *Ilex*

species, we selected the ten most variable nuclear SSR (simple sequence repeat) markers, namely, *Ile01-47*, *Ile03-38*, *Ile03-53*, *Ile03-65*, *Ile03-86b*, *Ile04-02*, *Ile04-06*, *Ile04-17*, *Ile04-18* and *Ile05-83*. We also assayed the nine most informative nuclear SSR from those developed by Pereira et al. (2013) for *I. paraguariensis* (i.e., *Ipg-01*, *Ipg-06*, *Ipg-07*, *Ipg-10*, *Ipg-19*, *Ipg-23*, *Ipg-41*, *Ipg-46* and *Ipg-49*). In addition, we assayed 11 heterologous plastid SSR markers (*ccmp2*, *ccmp3*, *ccmp4*, *ccmp6*, *ccmp7*, *ccmp10*, *ZMCP3764*, *ZMCP7430*, *ZMCP17192*, *ZMCP18704* and *ZMCP20824*; Weising and Gardner 1999; Provan et al. 1999). These encompass the large single-copy region of the chloroplast, with the exception of locus *ccmp10* that extends over one inverted repeat.

All PCR amplifications were performed using ca. 30 ng of genomic DNA template, 60 ng of each primer, 10 % of 10X PCR buffer (Invitrogen Life Technologies, California, USA), 2 mM $MgCl_2$, 0.2 mM of each dNTP, 0.5 unit of Taq DNA polymerase (Invitrogen Life Technologies), in a final volume of 25 µl. Cycling consisted of an initial 4 min denaturation at 94 °C, followed by 30 cycles of 1 min at 94 °C, 1 min at 50–60 °C (annealing changing according to primer Tm), 1 min at 72 °C, and a final extension at 72 °C for 7 min, on a GenePro TC-E apparatus (Bioer Technology Ltd., Tokyo, Japan). PCR products were checked by electrophoresis in 1.5 % (w/v) agarose gels in 1× TAE buffer (0.04 M Tris, 0.114 % v/v glacial acetic acid, 1 mM EDTA pH 8.0). Gels were stained with ethidium bromide, visualized and photographed under UV light.

Aliquots of SSR amplifications were mixed with 10 µl of 98 % (v/v) formamide buffer, heat-denatured at 95 °C for 8 min and loaded onto 6 % (w/v) high resolution denaturing polyacrilamide gel (8 M urea). Electrophoresis was carried out in 1× TBE buffer (0.09 M Tris, 0.09 M boric acid, 2 mM EDTA, pH 8.0) at 60 W for about 2 h, in a Model S2 apparatus (Gibco BRL Sequencing System, Life Technologies, Maryland, USA). Bands were visualized by silver nitrate staining according to Bassam et al. (1991). Allele sizes were estimated from digitalized gel images, by comparison against the molecular weight standard *30–330 AFLP DNA Ladder* (Invitrogen Life Technologies), and by using internal standard markers in every run. To verify the allele sizes estimated, a set of nuclear alleles were randomly chosen for sequencing. These were isolated from polyacrylamide gels following the procedure of Stumm et al. (1997), PCR re-amplified as described above, and electrophoresed in 1.5 % (w/v) agarose gels for further purification with QIAquick Gel Extraction Kit (QIAgen Inc.). Sequences were obtained at the "Unidad de Genómica del Instituto de Biotecnología" (Centro Nacional de Investigación Agropecuaria INTA, Castelar, Argentina) using amplification primers and an ABI3130xl sequencer (Applied Biosystems). Chromatograms were edited with BioEdit 7.1.3.0 (Hall 1999). The sequences were deposited in GenBank (KP014023-KP014047).

In order to compare the genetic diversity harbored by the germplasms from the center (ABP) and tail (Uruguay) of the species distribution area, the following indices and parameters were calculated for each nuclear SSR locus using PowerMarker 3.25 software (Liu and Muse 2005): the major allele frequency (MAF), the number of genotypes, the observed heterozygosis (H_o), the gene diversity (H_e) and the polymorphism information content (PIC). Allelic richness (R_s) index (El Mousadik and Petit 1996) was calculated using the software Fstat 2.9.3.2 (Goudet 1995, 2001).

Differences in H_e and R_s among germplasms were tested for significance by a Kruskal–Wallis test using STATISTICA (StatSoft, Inc. 2001). The presence of group-specific alleles (hereafter referred to as private alleles, i.e., alleles present in only one group, in frequency $\geq$0.05, and absent in the others) and of rare alleles (i.e., those alleles in frequency <0.05) was examined. The discriminating power of the SSR loci was calculated with GenAlEx 6.501 (Peakall and Smouse 2012) through the estimate of the overall probability of identity (PI), according to Waits et al. (2001). The allele frequencies of plastid SSR were calculated by direct counting from individual records. Size information was used to detect the haplotypes; each haplotype is considered as one allele at a single haploid locus.

Nuclear SSR dataset was used to generate a genetic distance matrix among individuals, using the shared allele distance ($S_{AD} = 1 -$ proportion of shared alleles, Chakraborty and Jin 1993) in PowerMarker. This index does not assume any mutational pattern and has a reduced variance (Goldstein et al. 1995). The distance matrix was imported in SplitsTrees 4 (Huson and Bryant 2006) to build a network using the Neighbor-Net algorithm (NN, Bryant and Moulton 2004). The following settings were used to generate the NN: edge fitting as ordinary least squares, equal angle as the chosen splits transformation, least squares to modify weights and four maximum dimensions as the filtering option. The split graph generated yields a visual representation of conflicting signals in the data by presenting them as a series of parallel edges. The least squares fit (LSfit) between the pair-wise distances in the graph and distances in the matrix were computed with the SplitsTrees program. A Principal Coordinate Analysis (PCoA) was performed using the S_{AD} matrix in GeneAlEx, with default settings. In addition, the genetic relationships between the samples were inferred employing the model-based Bayesian approach implemented in STRUCTURE 2.3.4 (Pritchard et al. 2000). The number of genetic clusters (i.e., the number of reconstructed ideal populations, or *K*) evaluated

Table 2 Variation parameters of nuclear SSR loci

Locus[a]	MAF[b]	PIC[b]	Allelic range[b] (in bp)	Observations		Alleles[c]		Genotypes[d]		H_o^e		H_e^e		R_s^e	
				Uru	ABP	Uru	ABP	Uru	ABP	Uru	ABP	Uru	ABP	Uru	ABP
Ile01-47	0.619	0.450	186–196	58	22	3	5 (2)	4 (1)	6 (3)	0.345	***0.636***	0.485	0.517	2.22	4.37
Ile03-38	0.456	0.690	268–300	58	21	6 (2)	5 (1)	11 (8)	7 (4)	0.241	0.143	0.637	0.702	4.82	4.48
Ile03-53	0.838	0.271	144–150	55	19	4 (1)	5 (2)	4 (1)	6 (3)	***0.018***	0.158	0.154	0.565	2.46	4.36
Ile03-65	0.531	0.529	199–215	58	22	3	7 (4)	6 (4)	8 (6)	0.379	0.636	0.531	0.607	2.64	5.71
Ile03-86b	0.911	0.158	210–214	58	21	3	3	4 (1)	3	0.138	0.190	0.161	***0.177***	2.38	2.72
Ile04-06	0.639	0.477	189–209	58	21	3	6 (3)	4	8 (4)	0.224	0.619	0.293	0.660	2.53	5.09
Ile04-17	0.875	0.199	201–207	58	22	2	3 (1)	2	3 (1)	0.155	0.500	***0.143***	0.385	***1.91***	2.59
Ile04-18	0.494	0.444	235–241	58	22	3	4 (1)	5 (1)	5 (1)	***0.448***	0.545	0.523	0.458	2.64	3.68
Ipg-01	0.430	0.688	285–320	50	14	7	7	8 (2)	8 (2)	0.040	***0.071***	0.695	***0.798***	***5.08***	***6.99***
Ipg-06	0.546	0.555	140–154	58	18	4	5 (1)	5 (3)	7 (5)	0.310	0.222	0.474	0.451	3.01	4.43
Ipg-07	0.344	0.736	159–179	53	8	7 (5)	3 (1)	10 (9)	3 (2)	0.377	0.500	***0.742***	0.531	–	–
Ipg-10	0.450	0.592	360–380	57	13	4 (2)	2	7 (5)	3 (1)	0.404	0.308	0.594	0.497	3.79	***2***
Ipg-19	0.676	0.431	224–228	58	16	3	3	4	6 (2)	0.103	0.188	0.335	0.648	2.40	3
Ipg-46	0.377	0.610	172–208	50	15	3	5 (2)	6	8	0.300	0.333	0.659	0.689	3	4.85
Mean	**0.585**	**0.488**	–	**56.21**	**18.14**	**3.93**	**4.5**	**5.71**	**5.8**	**0.249**	**0.361**	**0.459**	**0.549**	**2.99**	**4.17**

[a] Loci denoted with *Ile* are heterologous for "yerba mate", whereas *Ipg* denote species-specific loci

[b] Calculations based on N = 80. *MAF* major allele frequency, *PIC* polymorphism information content

[c] The number of private alleles is indicated between parenthesis

[d] The number of genotypes exclusively detected in each group is indicated between parenthesis

[e] The maximum and minimum values in each group are indicated in *bold* and *italic* fonts. R_s calculations were based on N = 13. Uru, stands for Uruguayan accessions; ABP, refers to accessions from Argentina, Brazil and Paraguay

ranged from 1 to 10, running ten replicates per *K* value, and 10^6 iterations of each Markov chain of Montecarlo (MCMC) after a burn-in period of 10^5. The analysis was performed considering the admixture model, unlinked loci, correlated allele frequencies (Falush et al. 2003), and assuming that the samples have an unknown origin. The optimum *K* value was determined by examining the posterior probabilities of the data (Ln P(D)) with respect to the number of clusters, then the run with the highest Ln P(D) was considered. Furthermore, STRUCTURE results were evaluated using an ad-hoc criterion based on the second order rate of change of the log probability of the data between successive *K* values (ΔK) (Evanno et al. 2005) with STRUCTURE HARVESTER 0.6.93 (Earl and vonHoldt 2012). An individual was assigned to a given cluster based on a probability of membership (*Q*) higher than an arbitrary cutoff value of 80 % ($Q > 80$ %). The partitioning of genetic variation was evaluated by the analysis of molecular variance (AMOVA; Excoffier et al. 1992; Excoffier 2007) using GeneAlEx. This was done to further study the effect of the geographic origin over the genetic variation taking into account five regions: (a) Argentina; (b) Brazil; (c) northern Uruguay (Tacuarembó); (d) mid-eastern Uruguay (Treinta y Tres); and (e) southern Uruguay (Lavalleja and Maldonado). Paraguay was not considered for this analysis. The significance of each component of the variance was computed in GeneAlEx, based on 999 permutations.

The relationship between genetic and geographic distances was evaluated with the Mantel test using the Alleles In Space 1.0 software (AIS; Miller 2005). Individuals CA1/74, CA8/74, CA11/75 and CA51/75 were removed from this analysis, as they are artificial progenies. When no Global Positioning Systems (GPS) coordinates were available, these were inferred using Google Earth version 7.1.2.2041 (2013). Geographical distances were obtained by converting the longitude/latitude coordinates to Universal Transverse Mercator (UTM) coordinates, and then calculating the Euclidean distances (in km) using the PBSmapping package for R version 3.0.2 (R Core Team 2013). For this analysis, geographical distances were log-transformed. Next, AIS was used to generate a "genetic landscape shape" through an interpolation procedure, employing geographical and residual genetic distances to construct a connectivity network of sampled areas, based on Delaunay triangulation. This allows the visualization of the genetic diversity distribution over a tri-dimensional surface, where the x- and y-axes represent the geographical space and the z-axis reflects genetic distances. A uniform grid size of 100 × 100, and a distance weighting value

(a) of 1 were used, after testing different values for these parameters (50 × 50; a = 0.5–2).

Results

We surveyed 19 nuclear microsatellite loci of which 14 rendered clear and legible banding patterns (Table 2). Eight heterologous primer-pairs, out of the ten assayed, were successfully transferred to "yerba mate" amplifying in most individuals (range: 74–80). Of the nine species-specific primers tested, six effectively amplified in 61–76 individuals. Specimens showing ≥50 % of missing data were excluded from subsequent analyses. All the primer combinations generated polymorphic markers; the range of MAF values is 0.344–0.911 (average 0.585) and the range of PIC values is 0.158–0.736 (average 0.488). The discrimination power for this set of 14 loci is 3.03×10^{-09}, a value within the accepted range for codominant markers (Waits et al. 2001). We found 73 alleles and 116 genotypes in 80 individuals. Parameters of nuclear SSR variability were computed for the Uruguayan individuals on the one hand, and for the Argentinean, Brazilian and Paraguayan specimens on the other (Table 2). For the Uruguayan germplasm, we registered 55 alleles (range: 2–7 per locus), and 80 genotypes (range: 2–11 per locus); for Argentinean, Brazilian and Paraguayan individuals, we detected 63 alleles (range: 2–7 per locus) and 81 genotypes (range: 3–8 per locus). Among the genotypes recorded in each group, 44 % were exclusively found in the first group, whereas 42 % were found solely in the second group. Even though between both sets the differences in H_e are non-significant, the differences in R_s are significant (R_s: $H_{(\text{d.f.}=1;\ N=26)} = 5.09$, $P < 0.024$). In total, 28 private alleles were detected, of which the Uruguayan gene pool has five in frequencies >0.1 (range: 0.123–0.553) and the other group has four (frequency range: 0.105–0.310). No rare alleles were detected throughout the data set.

Among the 11 plastidic microsatellite loci assayed (heterologous), only locus *ZMCP7430* showed two alleles (frequencies: 0.583 and 0.416); the remaining loci are monomorphic. Thus, only two haplotypes were detected

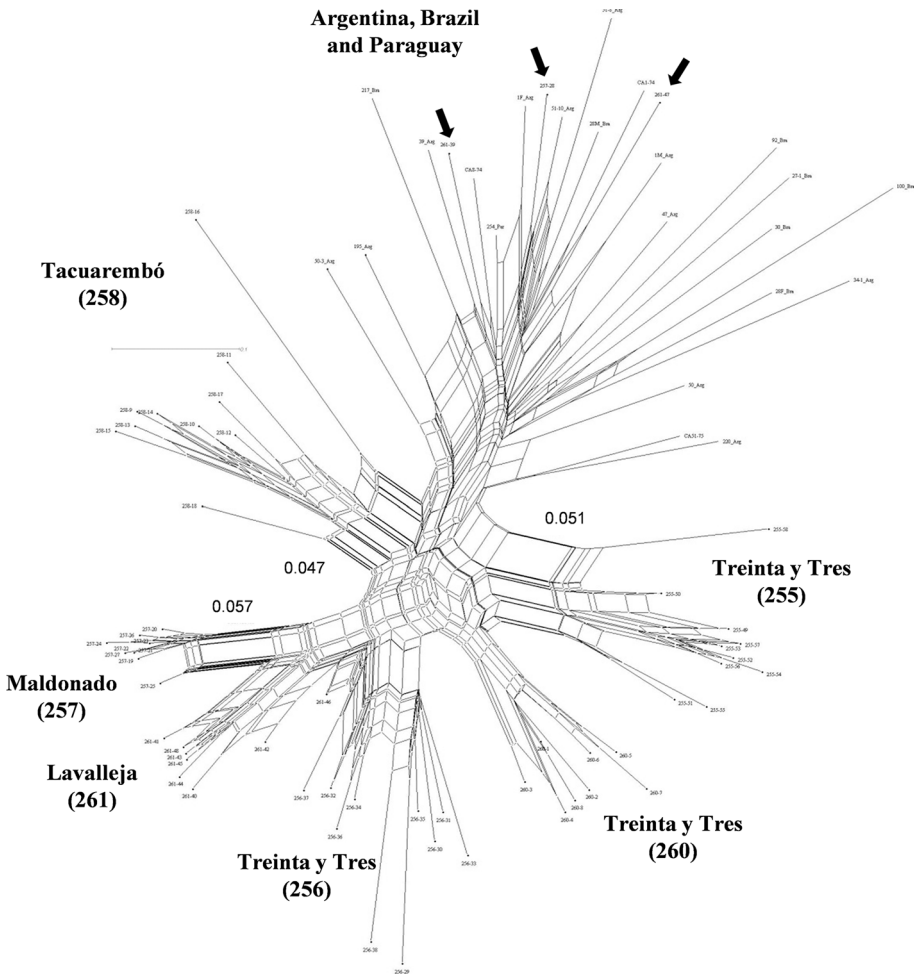

Fig. 2 Neighbor-Net split graph based on shared allele distances (S_{AD}) among "yerba mate" individuals. The *scale bar* is in genetic distance units. The terminal branches of Uruguayan individuals are marked with a *dot*. The weight of each edge, shown in *numbers*, is proportional to its length

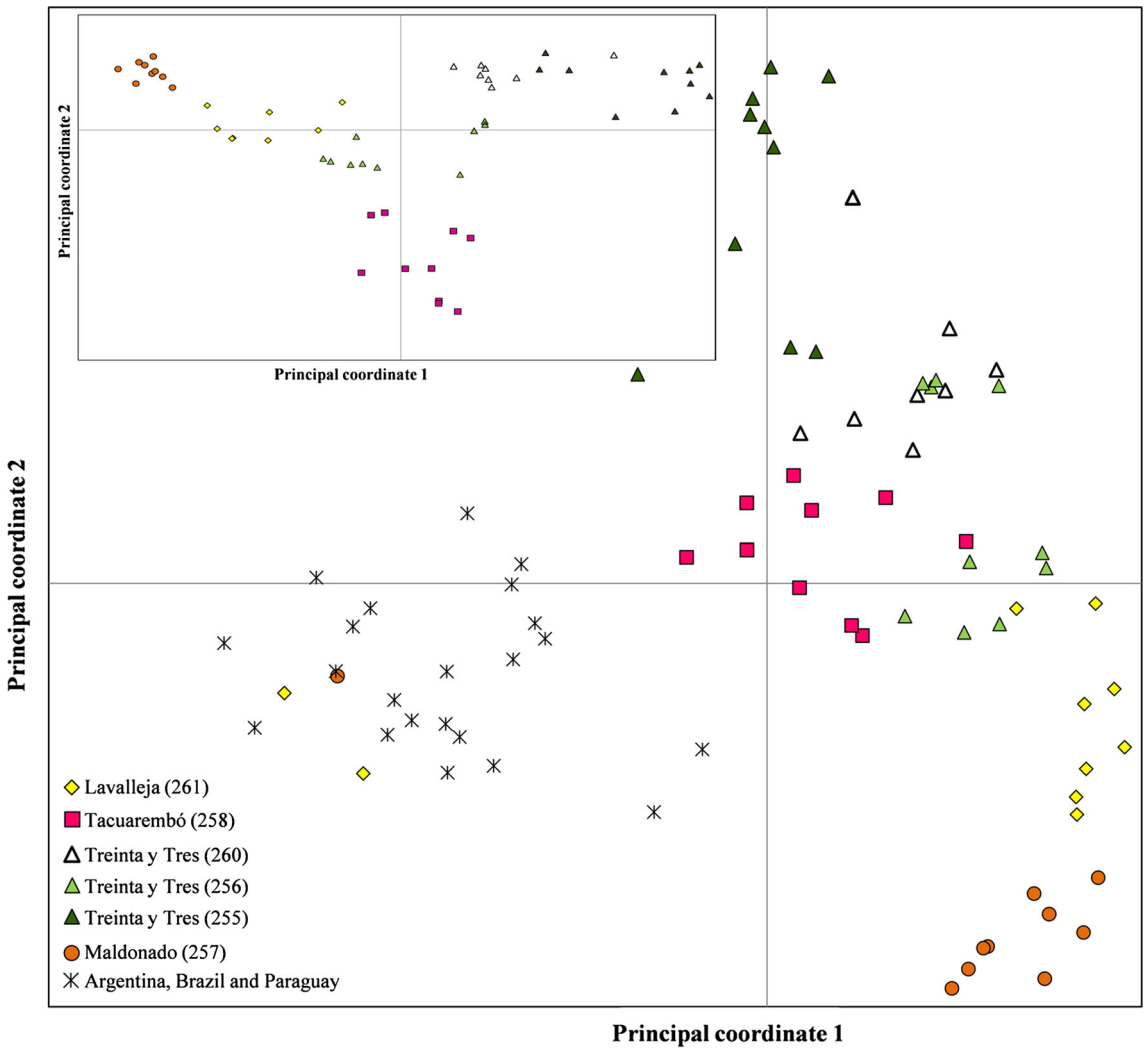

Fig. 3 Principal Coordinates Analysis (PCoA) of *I. paraguariensis* individuals. The first and second axes explain 18.01 and 11.77 % of the variation, respectively. *Inset*, PCoA performed on Uruguayan individuals; the first axis explains 21.84 % of the variation, whereas the second axis explains 14.30 %

across the individuals studied. The set of individuals from Argentina and Brazil show both haplotypes, whereas individuals from Paraguay exhibit the least frequent haplotype and those from Uruguay show the most frequent one.

The nuclear SSR dataset was used to generate a pairwise distance matrix (S_{AD}). Individuals from ABP and from Uruguay have an average $S_{AD} = 0.491$ and 0.424, respectively; whereas the between groups distance is 0.596. Within the latter group, the specimens from Lavalleja, Treinta y Tres, Tacuarembó and Maldonado show a $S_{AD} = 0.330$, 0.316, 0.287 and 0.186, respectively. Among these collection sites, individuals gathered from Lavalleja and Maldonado are the least distant (0.364), whereas those from Tacuarembó are maximally distant from those from Maldonado (0.506) (Table S1). The NN split graph based on the distance matrix (LSfit = 90.558) shows that individuals from ABP appear intermingled but apart from Uruguayan specimens, which cluster in six groups according to the geographic origin of the individuals. However, two specimens from Lavalleja and one from Maldonado clustered within ABP (arrows in Fig. 2). When these outliers are excluded, the distance values within Lavalleja and Maldonado drop to 0.130 and 0.07, respectively (Table S1). The heaviest splits in the NN are those that set apart individuals from Maldonado, Treinta y Tres (accession 255) and Tacuarembó (Fig. 2). The PCoA shows the separation between ABP and Uruguayan germplasms through the first coordinate; the second coordinate clearly discriminates Maldonado and Treinta y Tres (accession 255) from the rest (Fig. 3). The PCoA performed solely with Uruguayan samples, excluding the outliers, shows more clearly the grouping of the individuals according to their origin (Fig. 3, inset).

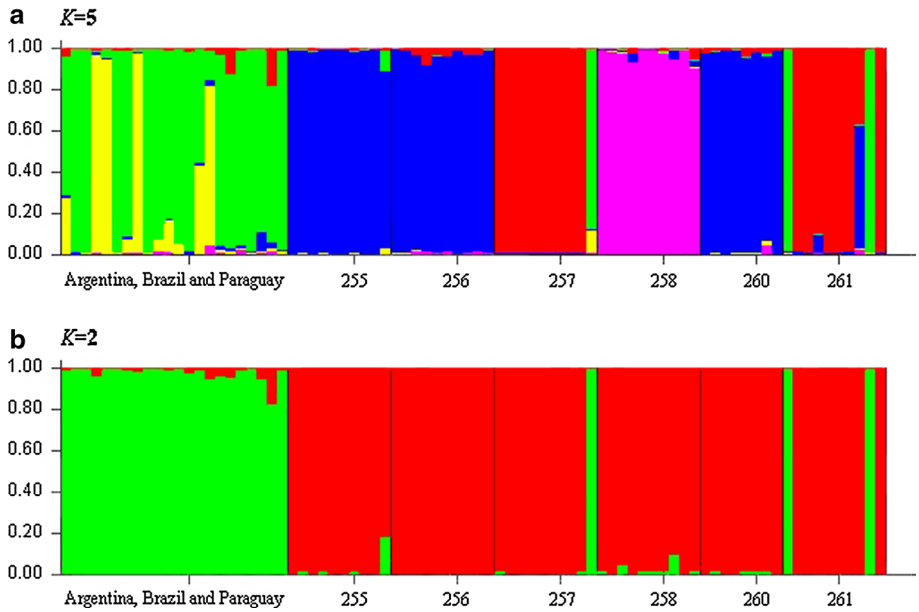

Fig. 4 Graphic representations of the ideal gene-pools detected for **a** K = 5 and **b** K = 2. Each individual is represented by a *vertical line*; *shades* indicate the contribution of each inferred gene pool (*K*) to the individuaĺs genome. *Thin black lines* delimitate the collection sites considered. The run with the highest Ln P(D) was chosen for graphical representation of each *K*

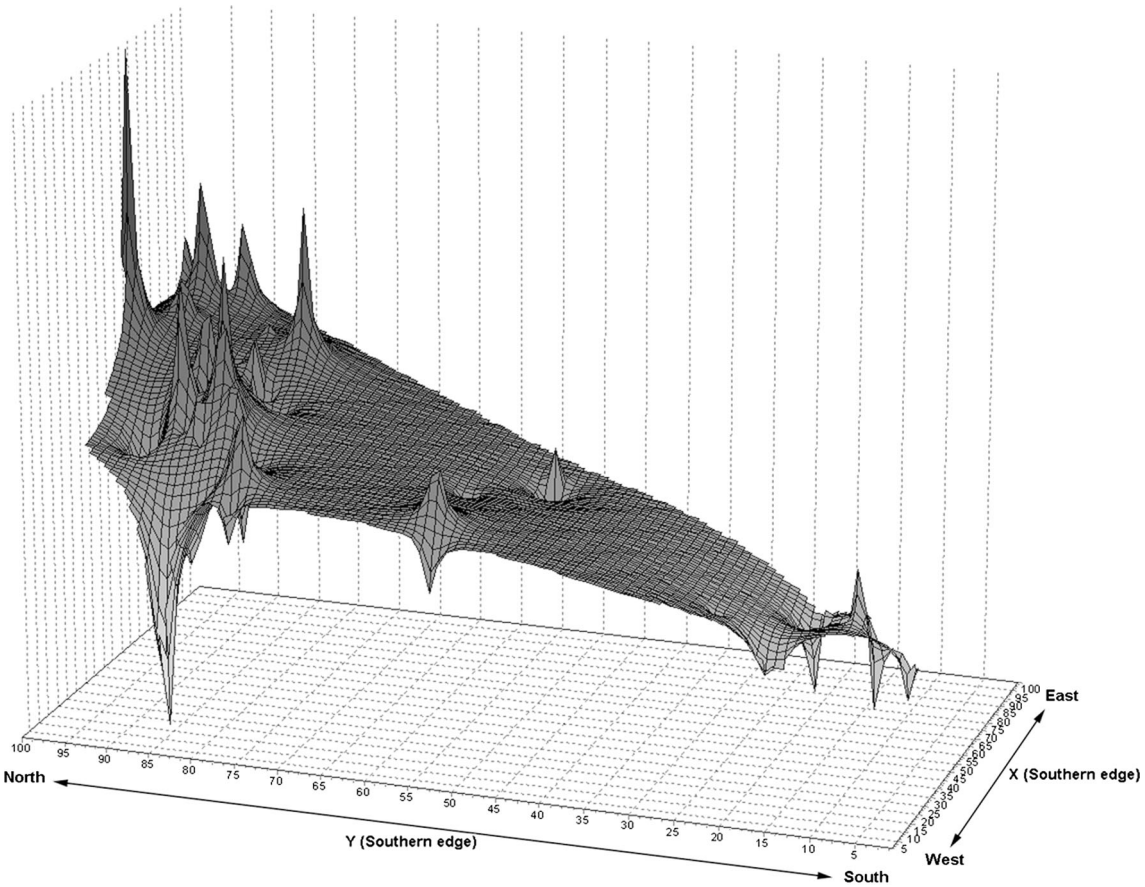

Fig. 5 Genetic Landscape Shape interpolation analysis using a 100 × 100 grid and a distance weighting parameter **a** of 1. The x- and y-axes represent the sampling area, and z-axis reflects the genetic distances

The analysis of individual multilocus genotypes with the Bayesian approach identified that the log-likelihood value for data conditional on *K*, Ln P(D), reached a plateau at $K = 5$. For this *K* value, individuals from Argentina distinguish from some Brazilians, whereas individuals from Uruguay are arranged into three gene-pools (each group

has an average $Q > 82$ %) (Fig. 4a). One group is composed by individuals from Lavalleja and Maldonado; another involves the individuals from Tacuarembó, and the last group is formed by the samples from Treinta y Tres. The above mentioned three outliers, appear again more related to Argentinean or Brazilian gene-pools than to those from Uruguay ($Q = 97$–98 %). The fragmentation scheme obtained here for Uruguayan samples agrees with the partitions delimited in the NN network (Fig. 2) and the PCoA (Fig. 3). The assessment of ΔK following the method of Evanno et al. (2005) identified the optimal value of this statistic at $K = 2$. For this K value, one partition is formed by all individuals from Uruguay (with $Q > 81$ %) and the other by samples from ABP plus the three outlier Uruguayan individuals ($Q > 99$ %) (Fig. 4b).

To examine the effect of collection sites on the partitioning of genetic variation and its distribution, an AMOVA was performed considering the five groups as the uppermost hierarchical levels (i.e., Argentina, Brazil and three regions within Uruguay). This analysis revealed that 14 % of the genetic variation is among the groups recognized (d.f. = 4; $\varphi = 0.142$; $P \leq 0.001$); 27 % is among collection sites within each group (d.f. = 3; $\varphi = 0.317$; $P \leq 0.001$) and 59 % is within collection sites (d.f. = 71; $\varphi = 0.414$; $P \leq 0.001$). Among the Uruguayan collection sites, no significant differences were found for H_e and R_s indices.

A significant positive correlation between genetic and geographic distances was detected for the entire set of samples ($r = 0.699$; $P < 0.001$) and for the Uruguayan data set (Mantel test, $r = 0.473$; $P < 0.001$). The genetic landscape shape interpolation analysis shows a directional pattern of the genetic diversity. The surface plot revealed that greater genetic distances are found in specimens from the northern region than in southern individuals (Fig. 5).

Discussion

It is known that genetic resources provide the basic material for plant breeding programs and that a prerequisite for this is to understand the amount and distribution of the genetic variation contained within them. Our outcomes indicate that the scattered populations of *Ilex paraguariensis* thriving in the Uruguayan territory represent a valuable genetic reservoir that deserves further exploration and a comprehensive conservational action.

Herein, we have jointly characterized Uruguayan wild "yerba mate" trees and representatives from the species diversification center, including selected progenitors, with co-dominant molecular markers. For the whole region surveyed we found an average $H_e = 0.504$. Previous studies concerned solely the Brazilian germplasm; a low diversity was found in southern populations using RAPD markers ($H_e = 0.163$, 148 individuals; Gauer and Cavalli-Molina 2000). Through isoenzymatic analysis a higher variability was detected for southern populations in comparison to those from mid-western and southeastern Brazil (average $H_e = 0.221$; Neumann-Wendt 2005). This structuring was associated to inbreeding and/or genetic drift arising from habitat fragmentation caused by deforestation, and/or natural selection. These studies found most of the genetic variation within the populations (>85 %), and this was attributed to the life history characteristics (i.e., obligatory outcrossing, perennial, long-lived species). More recently, the survey of two southeastern Brazilian populations reported high diversity using species-specific SSR markers ($H_e = 0.60$, 48 individuals; Pereira et al. 2013). This ample range of values may be explained both by the methodological approaches and by the intrinsic variation of each population surveyed.

The present results indicate that individuals of *Ilex paraguariensis* from the diversification center, despite their scattered and less intensive sampling, harbor greater genetic variation than those from the tail of the distribution (i.e., Uruguay). This is supported by the higher average estimates obtained for the group thriving in the central area as to the number of alleles (1.15-fold), H_e (1.19-fold), R_s (1.39-fold), and the between-group genetic distances (1.16-fold). Also the shape of the genetic landscape interpolation analysis suggests that the genetic variation decays southward, that is, towards the Uruguayan territory. The distance-based analyses (NN and PCoA) clearly discriminate ABP individuals from the Uruguayans, and within the latter group they further distinguish Tacuarembó, Maldonado and Treinta y Tres (255) as the most divergent individuals. Moreover, the Bayesian analyses also suggest that the germplasm from Uruguay is genetically differentiated in more than one gene-pool. We showed that this germplasm hosts a combination of nuclear alleles not present in the central region and thus it constitutes a valuable breeding resource. Consequently, this territory deserves the implementation of intensive germplasm conservation efforts to surpass the ongoing genetic erosion. These efforts should concentrate in collecting numerous individuals per site, rather than a few from different locations, as indicated by our genetic variation partition analysis.

Two distinct Floras have been proposed for the Uruguayan territory (Grela 2004), one westward and another eastward. At the geomorphological level, eastern Uruguay roughly continues the southern Brazilian regions "Depressão Central Gaúcha" and "Planalto Sul-Riograndense", explaining the presence of subtropical species, like *I. paraguariensis*, in such extreme (Grela 2004; Speranza et al. 2007). The Uruguayan Eastern Flora is subdivided in two areas, one to the north-west of the Rivera and

Tacuarembó Departments, and the other from the Cerro Largo Department, southwards to Maldonado. Genetic and biogeographical studies report a complete lack of seed flow between these two areas (Speranza et al. 2007; Turchetto et al. 2014). The data gathered here for *I. paraguariensis* trees from Uruguay broadly correspond to the above subdivisions. Precisely, the Uruguayan germplasm mostly associate in three distinct gene-pools according to the geographical origin. The northern germplasm (Tacuarembó) is situated in the first area mentioned in the precedents studies, whilst the mid-eastern (Treinta y Tres) and southern germplasms (Lavalleja and Maldonado) are located in the second region. The present work contributes to the knowledge of the genetic diversity of plant species from Uruguay, that so far have been focused on species of *Bromus, Paspalum, Stipa, Turnera* and on the *Petunia axillaris* complex (Rivas 2001; Speranza 2005; Speranza et al. 2007; Vidal et al. 2011; Turchetto et al. 2014).

According to Grela (2004), southwards, climatic conditions become increasingly adverse for subtropical species thriving in Uruguay. Particularly, the "yerba mate" tree requires high temperatures and precipitation regimes for seed germination and seedlings survival. The major climatic changes that occurred in the region during the Quaternary Period mostly affecting precipitation regimes, determined current species distribution patterns (Speranza et al. 2007). During the dry-cold phases, the subtropical elements may have survived in refugial areas like valleys, hilly and lowland areas (Speranza et al. 2007) which could explain the patchy distribution pattern of *I. paraguariensis* within Uruguay. This pattern could also be resultant of secondary southwards dispersions with scattered founder events, mediated by frugivorous birds. Herein, we have detected three individuals from southern Uruguay that are consistently associated with the germplasm from the species diversification center. This was detected with the distance-based clustering analyses and the STRUCTURE analysis; the latter also allowed the recognition of admixed individuals. All these may have been transported southwards by birds (for instance, Turdidae) in recent times. Alternatively, this could be accounted for by a sampling error or by an involuntary mixture of seeds during the handling.

Our survey on the cytoplasmic variation allowed the detection of two plastidic haplotypes. So far, there is a single study assessing chloroplastidic SSR variability in *Ilex* species. Rendell and Ennos (2003) found a low variability (8 haplotypes) for the chloroplast genome of the European *I. aquifolium* (six SSR loci surveyed in 16 populations). The lower SSR variation found in the chloroplast genome in respect to the nuclear genome was also observed by other authors (for example, Kaundun and Matsumoto 2002); this could be explained by the conserved nature of the plastid genome (i.e., lower mutation rate than nuclear genomes; Bock 2007). Alternatively, this could also be accounted for by the usage of heterologous SSR plastidic markers that failed to evidence the plastidic variation in *I. paraguariensis*. We expect that the sequencing of the "yerba mate" plastome, currently undertaken in our laboratory, would allow the design of species-specific plastidic markers, and subsequently, the detection of cytoplasmic variants for *I. paraguariensis* (Cascales et al. unpublished).

Acknowledgments The authors are indebted to SD Prat Kricun, LD Belingheri, R Bubilo and RM Mayol from EEA INTA Cerro Azul (Misiones, Argentina) who kindly provided specimens from the living collections. We are also grateful to JC Vilardi for helping us with the use of R software, and to MEN Rosetti that helped improving an earlier version of the manuscript. Grants from the Consejo Nacional de Investigaciones Científicas y Técnicas (PIP 12-200801-00342), the Universidad de Buenos Aires (UBA 20020100100859) and the Agencia Nacional de Investigaciones Científicas y Técnicas (PICT 2010-1665 and PICT 2011-2420) are gratefully acknowledged.

Conflict of interest The authors declare no conflict of interest.

References

Anesini C, Turner S, Cogoi L, Filip R (2012) Study of the participation of caffeine and polyphenols on the overall antioxidant activity of mate (*Ilex paraguariensis*). LWT-Food Sci Technol 45(2):299e–304e

Bassam BJ, Caetano-Anollés G, Greshoff PM (1991) Fast and sensitive silver staining of DNA in polyacrylamide gels. Anal Biochem 196:81–84

Bock R (2007) Structure, function and inheritance of plastid genomes. In: Bock R (ed) Cell and molecular biology of plastids. Springer-Verlag, Berlin Heidelberg, pp 29–64

Bryant D, Moulton V (2004) Neighbor-net: an agglomerative method for the construction of phylogenetic networks. Mol Biol Evol 21:255–265

Chakraborty R, Jin L (1993) A unified approach to study hypervariable polymorphisms: statistical considerations of determining relatedness and population distances. In: Pena SDJ, Chakraborty R, Epplen JT, Jeffreys AJ (eds) DNA fingerprinting: state of the science. Birkhauser Verlag, Basel, Switzerland, pp 153–175

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–361

El Mousadik A, Petit R (1996) High level of genetic differentiation for allelic richness among populations of the argan tree (*Argania spinosa* (L.) Skeels) endemic to Morocco. Theor Appl Genet 92:832–839

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. Mol Ecol 14:2611–2620

Excoffier L (2007) Analysis of population subdivision. In: Balding D, Bishop M, Cannings C (eds) Handbook of statistical genetics, 3rd edn. Wiley, New York, pp 980–1020

Excoffier L, Smouse P, Quattro J (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479–491

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587

Filip R, Ferraro GE, Bandoni AL, Bracesco N, Nunes E, Gugliucci A, Dellacassa E (2009) Mate (*Ilex paraguariensis*). In: Imperato F (ed) Recent advances in phytochemistry. Research Signpost, Kerala, India, pp 113–131

Gauer L, Cavalli-Molina S (2000) Genetic variation in natural populations of maté (*Ilex paraguariensis* A. St. Hil., Aquifoliaceae) using RAPD markers. Heredity 84:647–654

Giberti GC (2001) Diferentes aspectos del género *Ilex* (Aquifoliaceae). Corología, arquitectura floral, posición sistemática. Doctoral Thesis, Universidad de Buenos Aires, Argentina, 154 pp

Giberti GC (2011a) La "yerba mate" (*Ilex paraguariensis*, Aquifoliaceae) en tempranos escritos rioplatenses de Bonpland y su real distribución geográfica en Sudamérica austral. Bonplandia 20:203–212

Giberti GC (2011b) South American *Ilex* raw materials for regional infusions: advanced arboreal crop systems versus wild forest extractive production. In: Filip R (ed) Multidisciplinary approaches on food science and nutrition for the XXI century. Kerala, India, pp 17–27

Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW (1995) An evaluation of genetic distances for use with microsatellite loci. Genetics 139:463–471

Google Earth version 7.1.2.2041 (2013) http://www.google.com/earth. Accessed 12 Mar 2014

Gottlieb AM, Giberti GC, Poggio L (2011) Evaluación del germoplasma de *Ilex paraguariensis* e *Ilex dumosa* (Aquifoliaceae). Bol Soc Argent Bot 46:113–123

Goudet J (1995) FSTAT (version 1.2): a computer program to calculate F-statistics. J Hered 86:485–486

Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). http://www2.unil.ch/popgen. Université de Lausanne, Lausanne

Grela IA (2004) Geografía florística de las especies arbóreas de Uruguay: propuesta para la delimitación de dendrofloras. Msc thesis, Universidad de la República, Montevideo, Uruguay, 97 pp

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Series 41:95–98

Heck CI, Gonzalez De Mejia E (2007) Yerba mate tea (*Ilex paraguariensis*): a comprehensive review on chemistry, health implications, and technological considerations. J Food Sci 72:R138–R151

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23:254–267

Instituto Brasileiro da Erba-Mate (IBRAMATE) (2014) Mateando. Informativo do IBRAMATE 1, Ilópolis, Brazil, 15 pp

Instituto Nacional de la Yerba Mate (INIM) (2014) http://www.inym.org.ar/

Kaundun SS, Matsumoto S (2002) Heterologous nuclear and chloroplast microsatellite amplification and variation in tea, *Camellia sinensis*. Genome 45:1041–1048

Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128–2129

Miller MP (2005) Alleles In Space (AIS): computer software for the joint analysis of interindividual spatial and genetic information. J Hered 96:722–724

Neumann-Wendt S (2005) Genética de populações em *Ilex paraguariensis* St. Hil. Doctoral Thesis, Universidade Federal do Paraná, Brazil, 165 pp

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. Bioinformatics 28:2537–2539

Pereira MF, Ciampi AY, Inglis PW, Souza VA, Azevedo VCR (2013) Shotgun sequencing for microsatellite identification in *Ilex paraguariensis* (Aquifoliaceae). Appl Plant Sci. doi:10.3732/apps.1200245

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Provan J, Lawrence P, Young G, Wright F, Bird R, Paglia GP, Cattonaro F, Morgante M, Powell W (1999) Analysis of the genus *Zea* (Poaceae) using polymorphic chloroplast simple sequence repeats. Plant Syst Evol 218:245–256

R Core Team (2013). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Rendell S, Ennos RA (2003) Chloroplast DNA diversity of the dioecious European tree *Ilex aquifolium* L. (English holly). Mol Ecol 12:2681–2688

Rivas M (2001) Sistema reproductivo y estructura genética de poblaciones de *Bromus auleticus* Trinius ex-Nees (Poaceae): Estudio mediante isoenzimas. Agrociencia 5:32–40

Sosa PA, González Pérez MA, González-González EA, Rivero-Santana E, Naranjo-Cigala A, Oostermeijer G, Pestano-Broto JJ, Van Hengstum T (2009) Biología de la conservación de endemismos vegetales de los parques nacionales canarios: caracterización genética y demográfica. Proyectos de investigación en parques nacionales: 2006–2009

Speranza PR (2005) Evolutionary Patterns in the Dilatata Group (*Paspalum*, Poaceae): a polyploid/agamic complex. Doctoral Thesis; Florida: University of Florida, 119 pp

Speranza PR, Seijo JG, Grela A, Solís Neffa VG (2007) Chloroplast DNA variation in the *Turnera sidoides* L. complex (Turneraceae): biogeographical implications. J Biogeogr 34:427–436

StatSoft, Inc. (2001). STATISTICA (data analysis software system), version 6. www.statsoft.com

Stumm G, Vedder H, Schlegel J (1997) A simple method for isolation of PCR fragments from silver stained polyacrylamide gels by fine needle scratching. Tech Tips Online 2:116–117

Torimaru T, Tani N, Tsumura Y, Hiraoka K, Tomaru N (2004) Development and polymorphism of simple sequence repeat DNA markers for the evergreen shrub *Ilex leucoclada* M. Mol Ecol Notes 4:531–533

Torimaru T, Tani N, Tsumura Y, Nishimura N, Tomaru N (2007) Effects of kin-structured seed dispersal on the genetic structure of the clonal dioecious shrub *Ilex leucolada*. Evolution 61:1289–1300

Turchetto C, Fagundes NJR, Segatto ALA, Kuhlemeier C, Solís Neffa VG, Speranza PR, Bonatto SL, Freitas LB (2014) Diversification in the South American Pampas: the genetic and morphological variation of the widespread *Petunia axillaris* complex (Solanaceae). Mol Ecol 23:374–389

Vidal R, González A, Gutiérrez L, Umaña R, Speranza PR (2011) Distribución de la diversidad genética y sistema reproductivo de *Stipa neesiana* Trin. et Rupr. Agrociencia Uruguay 15:1–12

Waits LP, Luikart G, Taberlet P (2001) Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. Mol Ecol 10:249–256

Weising K, Gardner RC (1999) A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. Genome 42:9–19