An Acad Bras Cienc (2020) 92(1): e20190113 DOI 10.1590/0001-3765202020190113
Anais da Academia Brasileira de Ciências | *Annals of the Brazilian Academy of Sciences*
Printed ISSN 0001-3765 I Online ISSN 1678-2690
www.scielo.br/aabc | www.fb.com/aabcjournal

BIOLOGICAL SCIENCES

Unraveling the genetic complexity of a cultivated breeding population of "yerba mate" (*Ilex paraguariensis* St. Hil.)

DANIELA I. PAIVA, JIMENA CASCALES, MARÍA EVA N. ROSETTI, RAFAEL A. SCHERER, MARÍA E. GAUCHAT & ALEXANDRA M. GOTTLIEB

Abstract: We report the first comprehensive multilocus molecular characterization of cultivated *Ilex paraguariensis* plants belonging to a breeding program. Using nuclear and homologous chloroplastidic molecular markers, we have genotyped 158 plants from four plantation sites. Analyses of the nuclear data (187 variable dominant loci) allowed detecting high diversity (0.569), the occurrence of four distinct genetic clusters, and a low but significant differentiation among sites. Additionally, 20 chloroplastidic alleles were identified applying five microsatellite polymorphic markers, and a high chloroplastidic diversity was recognized (0.505); two haplogroups were distinguished amongst the 63 haplotypes detected. Our results from both nuclear and plastidic markers indicate that most genetic variation reside within plantations sites (≥ 95%), and that these plantations were established on highly variable materials (either as seeds or plantlets) derived from, at least, 63 maternal lineages. Moreover, our study suggests that the genetic structure within each plantation site was most likely shaped by past admixture favored by farmers´ practices during the establishment of each plantation. Also, subsequent constraints in gene flow and/or a low level of shared polymorphism among plantations could have contributed to current structure.

Key words: chloroplastidic microsatellites, genetic diversity, genotyping, ISSR markers.

INTRODUCTION

Ilex paraguariensis St. Hil. (Aquifoliaceae Bartl.), commonly known as the "yerba mate" tree, is a subtropical dioecious evergreen species restricted to north eastern Paraguay, southern Brazil, north eastern Argentina, and Uruguay (Giberti 2008, 2011). The popular infusion "mate" is prepared from the aerial parts and is highly appreciated in several southern South American countries for its distinctive flavor and invigorating and nutraceutical properties; qualities that are attributed to the high concentration of secondary metabolites (Anesini et al. 2012, Berté et al. 2011, Heck & de Mejía 2007). Variation in consumer´s preferences at regional and international markets had fostered producer´s renewed interest in developing new varieties (Marx et al. 2003, Scherer et al. 2006). Notably, current breeding programs had relied purely on phenotypic traits, and although it is well established that genetic characterizations produce more stable and valuable information to implement modern conservation and improvement programs (Picca et al. 2004), records on the genetic variation of Argentinean "yerba mate" germplasms are restricted to a small number of individuals (Gottlieb et al. 2005, 2011, Gottlieb & Poggio 2010, 2014) due to a near elimination of natural forests. In contrast, the knowledge on Brazilian genetic

resources is centered on the survey of their natural populations (Gauer & Cavalli-Molina 2000, Neuman-Wendt et al. 2007, 2009, Pereira et al. 2013). In the present study, we conducted a comprehensive molecular characterization of a selected population of *I. paraguariensis* belonging to an already established Argentine breeding program (Marx et al. 2003, Rakocevic et al. 2012, Scherer 2001, Scherer et al. 2006). This breeding population was established without any information on the genetic background of the seeds and plantlets used to erect it. We aimed at unveiling useful genetic variation to promote ongoing breeding efforts. To best of our knowledge, this is the first study that had characterized cultivated "yerba mate" plants by simultaneously examining nuclear and chloroplastic genomes. Joint examination of nuclear and chloroplastic data deliver a more complete panorama of the genetic diversity involved; nuclear data are biparentally inherited whereas in most angiosperms chloroplast genomes are maternally transmitted, lack recombination and a lower mutation rate (Borsch & Quandt 2009, Rendell & Ennos 2003). For the nuclear genotyping we implemented the technique Inter-Simple Sequence Repeat (ISSR; Zietkiewicz et al. 1994) which offer high genomic coverage, simultaneous survey of polymorphic loci, and optimal data throughput by means of affordable protocols (Pradeep Reddy et al. 2002). This technique is particularly useful for non-model crop species, revealing variation at various levels (Al-Turki & Basahi 2015, Gaiero et al. 2011, Kiani et al. 2012, Kumar et al. 2016, Seyedimoradi & Talebi 2014). To assess variation of chloroplastidic genome, we capitalize a panel of microsatellite markers recently designed for *Ilex paraguariensis* (J Cascales, unpublished data) from massive sequencing plastomic data (Cascales et al. 2017).

MATERIALS AND METHODS

Field sampling

We collected young and healthy leaves from 158 plants growing at four commercial plantations within the Argentine productive area (Table I, Fig. 1). Leaves were taken at 1-2 meters from the base and were preserved in silica-gel until use. The plants surveyed here integrate the base population of on-going improvement schemes carried-out by local "yerba mate" producers and had been selected on their quantitative and qualitative traits, such as crown size, type of primary and secondary branches, number of leaves per shoot, and health status, by Scherer (2001). These plants have been grown mostly in alfisole soils at lowlands (182-240 m.a.s.l.) and at tree densities between 1250 and 2222 plants per hectare. The four productive plantations were built using individuals with different origins. Owners of "Establecimiento Azula" (A) selected plants from older plantations and from remnants of wild forests, these plants were then domesticated. Plants held at "Establecimiento Urfer" (U) and "Chacra Villa Monte" (M) were originally provided from nearby plantations and nurseries held at the central region of Misiones. In contrast, plants held at "Establecimiento Stvas" (L) derive from a selection process performed by the owners´ last three generations, though with undocumented criteria.

DNA extraction

For DNA isolation, ca. 0.10 grams of dry leaf blade were ground in liquid nitrogen, processed with the DNeasy Plant Quick Extraction kit (QIAgen Inc., Duesseldorf, Germany) following manufacturer's instructions, and then stored at -20 °C.

Table I. Details on the *Ilex paraguariensis* materials used for molecular characterization.

Plantation Site (code)	Geographical origin†	GPS coordinates	N	Plantation age (years)
	Province of Misiones			
Establecimiento Azula (A)	General Belgrano, Comandante Andresito, Comandante Andresito	25º 42' S, 53º 59' W	45	29 – 31
Chacra Villa Monte (M)	Iguazú, Puerto Esperanza, Puerto Esperanza	26º 00' S, 54º 32' W	34	25
Establecimiento Urfer (U)	Libertador General San Martín, Puerto Rico, Colonia Oro Verde	26º 51' S, 55º 07' W	38	30 – 80
	Province of Corrientes			
Establecimiento Stvas (L)	Ituzaingó, Colonia Liebig, Colonia Liebig	27º 55' S, 55º 51' W	41	no data

†, the corresponding Department, Locality, and Municipality are indicated. Establecimiento and Chacra refer to rural establishments. N, number of plants sampled from each site.

Figure 1. Map of Northeastern Argentina, with detail of Misiones Province and northeastern Corrientes Province, showing location of the "yerba mate" plantation sites surveyed in this report.

ISSR genotyping

All PCR amplification runs were performed using ca. 3.5 ng of genomic DNA, 100 ng of ISSR primer, 1X PCR buffer (Thermo Fisher Scientific, California, USA), 1.5 mM $MgCl_2$, 200 µM dNTP and 0.5 unit of Taq DNA polymerase (Thermo Fisher Scientific), in a final volume of 25 µL. The PCR program was as follows: an initial heating at 94 °C for 3 min, 35 cycles of denaturing at 94 ºC for 30 s, annealing at 46 °C for 59 s, extension at 72 °C for 90 s, and a final extension of 5 min at 72 °C. PCR products were checked by electrophoresis in 1.5 % (w/v) agarose gels in 1X TAE buffer (0.04 M Tris, 0.114 % v/v glacial acetic acid, 1 mM EDTA pH 8.0), and then stained with ethidium bromide, and visualized and photographed under UV light. A preliminary screening was performed using ten 3´-anchored, one 5'-anchored and two unanchored ISSR primers (Supplementary Material - Table SI) on a subsample of 12 individuals (three from each plantation site). The primers showing the most promising banding patterns (i.e., with more than five clear bands, between 300-1500 bp in size, as visualized in agarose gels) were selected for further use, namely YM4, YM5, YM6, YM7 and

YM11. Then, ISSR amplicons were separated through 6 % (w/v) high resolution denaturing (8 M urea) polyacrilamide gel electrophoresis (PAGE), carried out in 1X TBE buffer (0.09 M Tris, 0.09 M boric acid, 2 mM EDTA, pH 8.0) at 60 W. Bands were visualized by silver nitrate staining using the Silver Sequence Staining Reagents (Promega, USA) kit, according to manufacturer's instructions. Images were digitalized from dried gels.

Allele sizes were estimated from gel images by comparison against the molecular weight standards 100 bp DNA Ladder (Solis Biodyne) and ΦX174 DNA/HaeIII Markers (Promega) and were homologized among gels by using internal standard markers in every electrophoretic run. Each band was considered as a dominant allele for a given locus and used to construct a binary data matrix of presence (1) or absence (0). From raw data matrices, we removed monomorphic (≥ 0.95) and low frequency (≤ 0.33) alleles in order to diminish methodological errors generated during gel comparisons. Thus, six data matrices were constructed: one per ISSR primer plus a multiloci matrix involving solely individuals with positive information for the five primers.

Chloroplastid microsatellite characterization

A preliminary screening was performed using 14 microsatellite haploid loci (cpSSR; Table SI) over a subsample of 16 individuals (four from each plantation site). The eight loci that yielded putative polymorphic patterns when screened on agarose gels were selected for studying the whole sample. Amplification conditions were as above but the annealing temperature was set to 60 °C and the final volume scaled to 15 µL. PAGE and silver staining conditions were carried out as previously stated. Allele sizes were estimated from digitalized gel images by comparison against molecular weight standards 100 bp DNA Ladder and 30–330 AFLP DNA Ladder (Thermo Fisher Scientific), and against internal standards of known sizes. As chloroplasts are haploid organelles, each individual yielded a single band per locus. The recorded information was used to construct a data matrix by combining allele sizes (in base pairs) for each individual plant and locus.

Data Analysis

Genetic diversity of biparentally inherited data was estimated through Shannon index in POPGENE 1.32 (Yeh & Boyle 1997, 1999). To compare the performance of the ISSR markers in the evaluation of the genetic profiles, polymorphism information content (PIC) and resolving power (RP) were calculated manually according to Roldan-Ruiz et al. (2000) and to Prevost & Wilkinson (1999), respectively. Partitioning of genetic variation within and among plantation sites was further evaluated with the analysis of molecular variance (AMOVA; Excoffier et al. 1992) using GenAlEx 6.5 (Peakall & Smouse 2012); statistical significance of each variance component was assessed with 1000 permutations of the data and were corrected for multiple comparisons using Bonferroni procedures (Sokal and Rohlf 1995) with the same program. In order to identify genetic clusters without a priori delimitation, we conducted a Discriminant Analysis of Principal Components (DAPC; Jombart et al. 2010) using the adegenet 2.1.0 package (Jombart & Collins 2015) implemented in R 3.4.2 (R Core Team 2017). DAPC has been shown to efficiently separate clusters by maximizing the separation between groups while minimizing variation within groups (Jombart et al. 2010). Particularly, clusters were determined using the function find.clusters, with n.iter=100000 and n.start=10, evaluating a range from 1 to 40, by the K-means algorithm (Legendre & Legendre 1998). As the optimum value was determined as K=4, we used three discriminant

functions for all analyses. A loading plot -which indicates the most contributing alleles- was obtained with the function loadingplot and default settings (threshold= 0.01). A Kruskal-Wallis independence test was performed between the deduced genetic clusters and the samples grouped according to their origin (i.e., A, M, U, and L) as implemented in stats 3.4.2 package of R (R Core Team 2017).

Relationships among individual plants were evaluated with the combined ISSR matrix by performing DAPC analyses with adegenet. Also, genetic distances were obtained by applying the complementary value of the Dice similarity index in FAMD 1.31 (Schlüter & Harris 2006) and the distance matrix was used to construct a Neighbor-Joining (NJ; Saitou & Nei 1987) unrooted dendrogram, and to perform bootstrap analyses (1000 pseudoreplicates) with the same program. Resulting NJ was visualized with FigTree v1.4.3 (Rambaut 2016).

Chloroplastidic diversity across individuals was examined through the number of alleles (Na), effective number of alleles (Ne), and gene diversity index (h; Nei 1978) using GenAlEx. Also, the number of private alleles (those present in only one predefined group and absent in the others) and the number of rare alleles (i.e., those in frequency < 0.05) were recorded. Size information of cpSSR alleles was used to manually identify plastidic haplotypes; each haplotype resulted from the combination of a single allele per haploid locus. Haplotype diversity (Nei 1987) was calculated with ARLEQUIN 3.0 (Excoffier et al. 2005). Relationships among individual plants and among haplotypes were also investigated. The corresponding distance matrices were generated with the program Populations 1.2.32 (Langella 1999) by applying the shared allele distance index (= 1 - proportion of shared alleles). This index does not assume any mutational pattern and has a reduced variance (Goldstein et al. 1994). The distance matrices were used to build NJ unrooted dendrograms and to perform bootstrap analyses (1000 pseudoreplicates) with the same program. A DAPC was also performed using the by individual's matrix as explained above. The hierarchical genetic structure for plastidic markers was investigated with an AMOVA among pre-defined sub-populations (i.e., plantations sites and ISSR-defined genetic clusters), as indicated previously. Also, to evaluate the impact of size homoplasy on these cpSSR loci we conducted a multilocus linkage disequilibrium analysis based on a likelihood-ratio test in GENEPOP v.4.7 (Rousset 2008). Size homoplasy -an apparent similarity that masks evolutionary differences- can arise from compound or interrupted microsatellite motifs, or from loci showing variation in the flanking regions, resulting in fragments of equal size that are not identical by descent (Angioi et al. 2009, Provan et al. 2001, Wheeler et al. 2014). Contingency tables are created for all pairs of loci then, a G test or a probability test is computed for each table using the Markov chain algorithm of Raymond & Rousset (1995). Because all microsatellite loci are on the same DNA molecule, we expect to see linkage disequilibrium between microsatellite regions if there is little homoplasy.

RESULTS

ISSR amplification was effectively accomplished in > 93 % of the individuals (over a total of 158); the number of bands recorded per primer varied from 49 to 99 (mean 68.6; Table II). A total of 343 bands were registered, of which 119 variable bands were in a frequency below the established threshold (i.e., < 33 %) and 37 were fixed (i.e., constant) across individuals, and were thus excluded. Therefore, the edited matrices

Table II. ISSR characterization of "yerba mate" cultivated plants.

Matrix name							Mean nº loci / individual †					AMOVA	DAPC		
							Plantation sites				Diversity (H') range ‡	Among sites (%)	NC	IA	CA
	N	n_i	n_f	Md (%)	PIC	RP	A	M	U	L					
YM4	147	60	31	1.51	0.31	19.65	15 a	19 b	22 b	21 b	0.463 (M) – 0.555 (L)	11	9	56	3
YM5	150	99	35	4.99	0.26	23.57	21 a	20 a	21 a	20 a	0.479 (M) –0.513 (L)	3	7	68	4
YM6	149	72	50	2.52	0.35	35.69	29 a	29 a	25 b	28 a	0.539 (U) – 0.597 (A)	4	4	99	10
YM7	147	63	40	3.96	0.31	21.22	26 a	27 ab	30 b	30 b	0.391 (L) – 0.496 (A)	5	4	97	4
YM11	148	49	31	3.38	0.36	19.35	22 a	18 b	18 b	21 a	0.506 (U) – 0.624 (A)	6	4	89	5
Combined	128	---	187	3.11	---	---	115 a	114 a	115 a	122 a	0.552 (U) – 0.582 (A)	#	4	100	26

N, number of individuals included in each matrix; n_i, number of loci of the raw matrix; n_f, number of loci of the edited matrix. Md, missing data; PIC, polymorphism information content calculated on raw matrices; RP, resolving power calculated on edited matrices. †For each matrix, identical letters indicate non-significant differences among sites in the mean number of loci per individual (P > 0.05). ‡ The corresponding site, indicated by the single letter code (A, M, U, L; see Table I), is denoted between parentheses. NC, number of clusters discriminated; IA, percentage of individuals showing a cluster assignment >90%; CA, number of most contributing alleles to the discriminant functions. #, please refer to Table IVa.

showed 31-50 variable bands and 3.27 % missing data, on average (Table II). PIC and RP values were jointly maximized for primer YM6, which yielded the highest number of informative alleles. Significant differences were detected among plantation sites in the mean number of bands (loci) per individual for matrices derived from primers YM4, YM6, YM7 and YM11. Plantation site A appeared as the most diverse with data from primers YM6, YM7 and YM11, whereas the lowest values varied according to the primer used. In all cases, the AMOVA indicated that most of the variation resides within sites, and that YM5 was the primer that more strongly evidenced this (97 %). Data derived from primers YM4 and YM5 allowed discrimination of more clusters but with fewer individuals robustly assigned to a cluster.

The combined matrix comprised 128 individuals and 187 variable ISSR loci (3% missing data; Table II). The mean number of bands per individual did not differ statistically among sites, and none private loci were found. The mean diversity was 0.569, being the site A the most diverse and the site U the least one. The mean genetic distances estimated across all individuals, within sites, and among sites, resulted in comparable values (0.300, 0.288 and 0.305, respectively; Table III). As before, most of the genetic variation resides within plantation sites (95%); a low but significant genetic differentiation was found among sites (Table IVa). DAPC discriminated four clusters to which individuals were assigned with full probability (Figs. 2-3; Fig. S1). Twenty-six loci contributed to cluster discrimination with a load > 0.01 (Table II); of those, only three showed a load > 0.05 (Fig. S2). The mean genetic distance within DAPC clusters was

Table III. Mean genetic distances (complementary of Dice index) calculated within and among plantation sites ("Sites"), indicated by the single letter code (A, M, U, L; see Table I), and mean values estimated within and among DAPC genetic clusters ("Clusters"). Values estimated within groups are in bold type; distances among groups are in plain type.

		Sites						Clusters			
		A	**M**	**U**	**L**			**1**	**2**	**3**	**4**
Sites	**A**	**0.303**				Clusters	1	**0.256**			
	M	0.320	**0.296**				**2**	0.277	**0.226**		
	U	0.322	0.306	**0.285**			**3**	0.324	0.305	**0.326**	
	L	0.297	0.292	0.291	**0.268**		**4**	0.294	0.291	0.343	**0.278**

0.271, reaching 0.306 among clusters (Table III), whereas the AMOVA indicates that ISSR clusters are significantly differentiated and that most variation resided within clusters (Table IVb) Clusters 2, 3 and 4 were composed by different proportions of individuals from the four sites (Fig. 2); notably, cluster 1 did not include representatives from site U. The independence between plantation sites and individual´s assignment to a genetic cluster was verified (Kruskal-Wallis test: X^2= 2.33, df= 3, p-value= 0.506). In the Neighbor-Joining, only a small number of individuals formed groups robustly supported by the data. Still, the DAPC clusters were roughly recovered in the unrooted dendrogram (Fig. 4).

Chloroplastidic characterization of 156 individuals with eight microsatellite loci allowed identification of 23 alleles (ca. 3% of missing data). Of those, 20 derived from five polymorphic loci (ca. 4% of missing data) which, on average, produced four alleles per loci (range= 2-5; frequency range= 0.007-0.809; Table V). A total of six rare alleles were registered (frequency range= 0.007-0.035) among which one -the 354 bp allele from locus *rps18-rpl20*- was private to plantation site L; the number of effective alleles ranged from 1.470 to 2.617. Excluding 28 individuals with missing data for at least one locus, the combination of cpSSR alleles resulted in 63 haplotypes. Fifty-four percent of the haplotypes were in a frequency of 0.008 whereas a single haplotype reached 10.2 % (Table VI). Thirty-eight private haplotypes were found distributed among plantations, with site A showing the highest number of private maternal lineages (namely, 12). Nineteen haplotypes were shared between two sites, further five haplotypes were shared among three sites, and only one was common to all sites (haplotype 30; Table VI). The minimum number of maternal lineages registered among plantation sites varied between 21 and 28. All plantation sites showed high haplotypic diversity values (range= 0.974-0.983). Most chloroplast variation was found within plantations sites (96%; Table IVc). Likewise, the chloroplastic variation among the four ISSR DAPC clusters indicated that most variation was held within those groups (97%; Table IVd). In both cases, a low but significant genetic differentiation had been obtained among sites (Table IVc-d). To determine whether size homoplasy is likely to occur in *Ilex paraguariensis*, we examined linkage disequilibrium between the five informative microsatellite loci studied. Two of the five cpSSR loci (*rps18-rpl20* / *psbI-trnS*) showed significant linkage disequilibrium (X^2=17.402, df= 8, p=0.02; Table SII).

The relationship among the 63 haplotypes showed an arrangement in two main haplogroups (Fig. 5) that are in great correspondence with the distribution of the two alleles from locus *accD-psaI* (379 and 381 bp); solely haplotypes 7, 35 and 44 which have the allele 379-bp, appeared nested within the haplogroup "381-bp". Still, no significant bootstrap values (ie, >50%) were retrieved. The relationships among individual´s haplotypes showed a similar backbone structure but, in this case, solely haplotype 44 appeared nested within haplogroup "379-bp", though without bootstrap support (Fig. S3). No evident correspondence had been detected either with plantation sites or ISSR clusters.

Table IV. Results from analyses of molecular variance (AMOVA) performed on combined ISSR (a-b) and cpSSR (c-d) matrices, considering (a-c) the plantation sites and deduced genetic clusters (b-d) as hierarchical groups.

	Source	Variation	df	Sum of squares	Variance components	Percentage of variation	Fixation index
a.	**ISSR combined**	Among plantation sites	3	315.839	2.137	5%	0.054 [a]
		Within plantation sites	124	4647.559	37.480	95%	
		Total	127	4963.398	39.617	100%	
b.	**ISSR combined**	Among genetic clusters	3	486.887	3.979	10%	0.099 [a]
		Within genetic clusters	124	4476.511	36.101	90%	
		Total	127	4963.398	40.080	100%	
c.	**cpSSR combined**	Among plantation sites	3	27.292	0.149	4%	0.043 [b]
		Within plantation sites	152	503.708	3.314	96%	
		Total	155	531.001	3.463	100%	
d.	**cpSSR combined**	Among genetic clusters	3	6.972	0.034	3%	0.027 [c]
		Within genetic clusters	124	154.560	1.246	97%	
		Total	127	161.531	1.280	100%	

[a] p= 0.001; [b] p= 0.002; [c] p= 0.009.

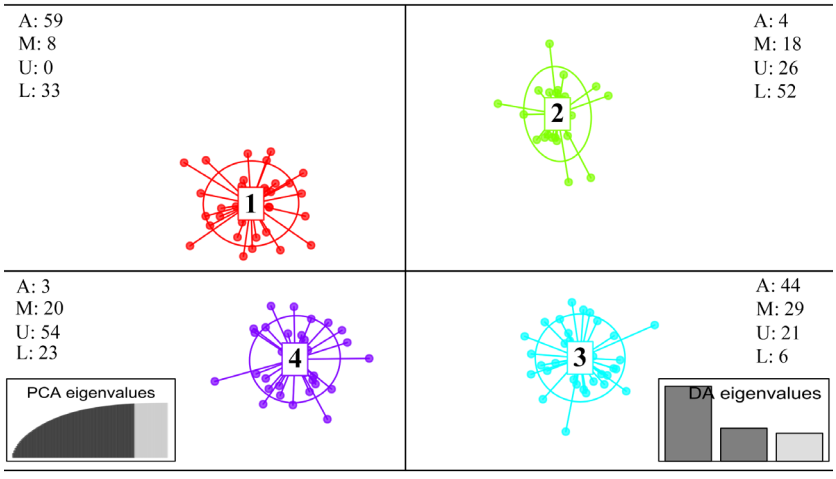

Figure 2. Genetic clustering. Scatterplot of the Discriminant Analysis of Principal Components (DAPC) depicting relative intra-group and amongst-group genetic distances. Dots represent individual samples and centroids are indicated with numbers (1-4). Inset indicates the number of principal components retained by DAPC and the cumulative variation among these (55.76%, 23.72% and 20.51%, respectively). The relative contribution of individuals from each site (A, M, U and L; see Table I) to the corresponding genetic cluster (1-4), is expressed in percentage.

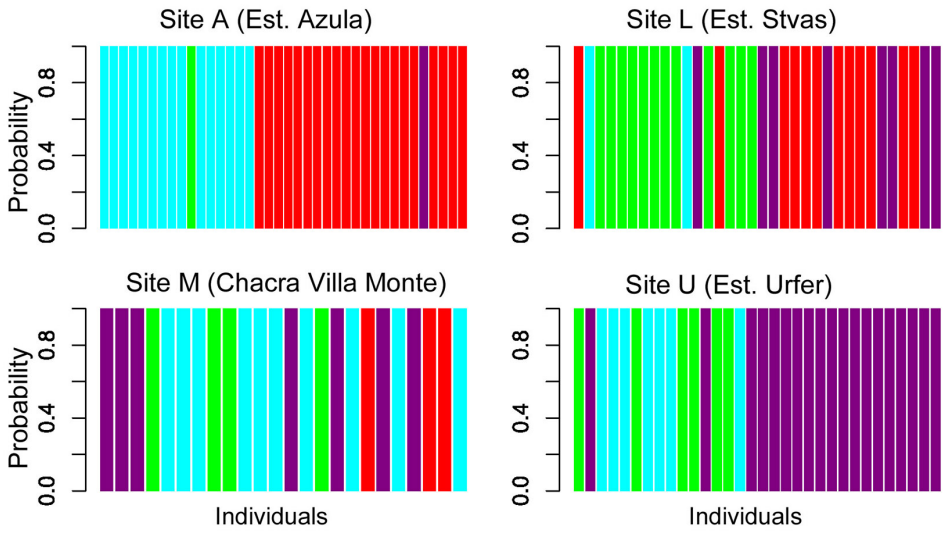

Figure 3. Estimated composition of each "yerba mate" plantation site surveyed based on the individuals´ genetic background. Each individual plant is represented by a thin vertical segment, which can be partitioned into K (K=4) colored segments that represent the individual´s estimated membership probability to the K cluster.

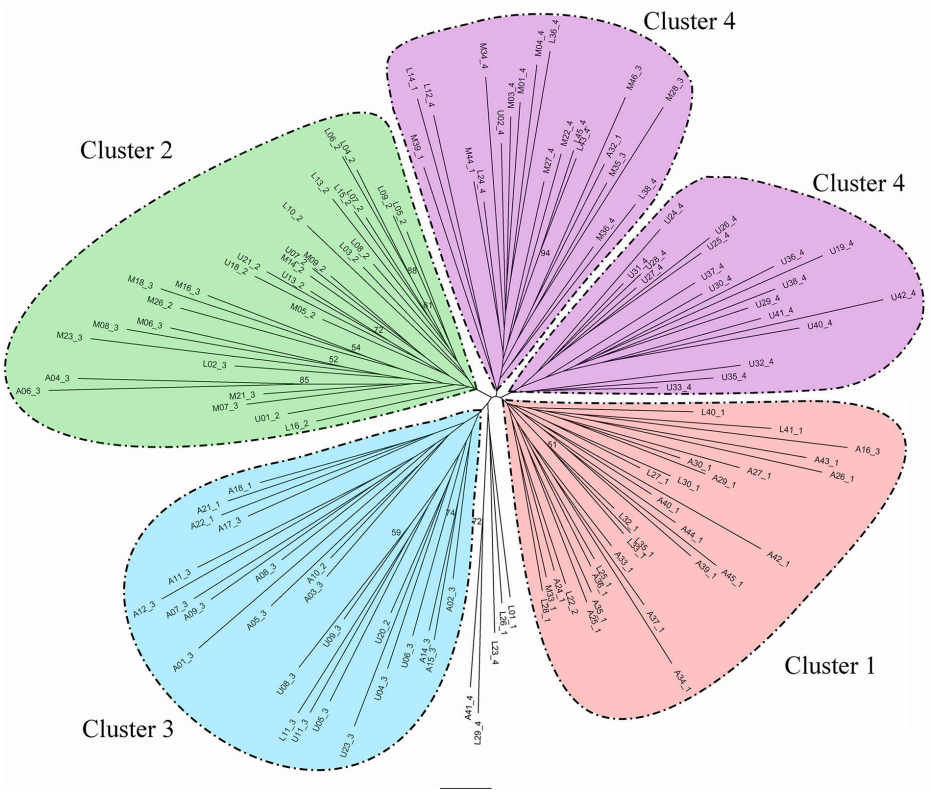

Figure 4. Neighbor-Joining unrooted dendrogram of "yerba mate" individual plants based on ISSR derived genetic distances (Dice distances). Individuals are indicated by their internal number followed by the number of the corresponding DAPC genetic cluster (1-4). Dotted lines circumscribe groups in rough correspondence with DAPC clusters. Bootstrap values ≥ 50 % are shown on branches.

Table V. Chloroplastidic characterization of *Ilex paraguariensis* individuals with microsatellite markers.

locus	***rpl14-rpl16***		***rps18-rpl20***		***psbI-trnS***		***psbZ-trnS***		***accD-psaI***		***infA-rps8***		***psbM-trnD***		***intron rps16***			
	A	f	A	f	A	f	A	f	A	f	A	f	A	f	A	f		
	448	*0.013*	**350**	*0.035*	**350**	*0.019*	**262**	*0.026*	**379**	0.373	**351**	1	**420**	1	**467**	1		
	450	0.121	**351**	0.183	**351**	0.635	**263**	0.151	**381**	0.627								
	451	0.809	**352**	0.585	**352**	0.224	**264**	0.546										
	452	0.051	**353**	0.190	**353**	0.122	**265**	0.243										
			354	*0.007*			**266**	*0.033*										
N	156		142		156		152		142		156		156		156		**Av.**	**SE**
Na	4		5		4		5		2		1		1		1		**--**	--
Ne	1.470		2.424		2.135		2.617		1.879		--		--		--		**2.105**	0.202
I	0.632		1.093		0.956		1.168		0.661		--		--		--		**0.902**	0.110
h	0.320		0.587		0.532		0.618		0.468		--		--		--		**0.505**	0.053

A, allele size in base pairs; f, allelic frequency, rare alleles (< 0.05) are in italics; N, number of individuals recorded; Na, number of different alleles; Ne, number of effective alleles; I, Shannon's Information Index; h, diversity. Av., average; SE, standard error.

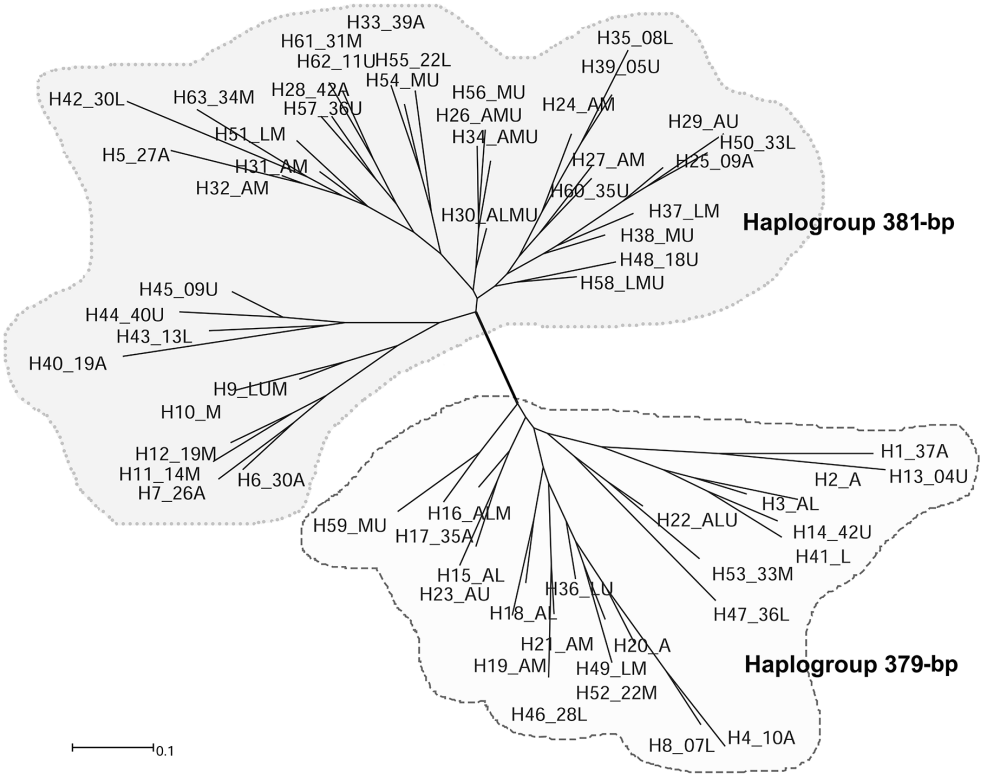

Figure 5. **Neighbor-Joining unrooted dendrogram of "yerba mate" haplotypes. Each terminal (or leaf) corresponds to a haplotype (H1 – H63; see Table VI). Additionally, haplotypes shared among plantation sites have an indication of the sites from where they were registered (i.e., A, M, U, L), private haplotypes were labeled either by the plantation site´s identification letter or by the individual´s identification alphanumeric code.**

Table VI. Frequencies of chloroplastidic haplotypes per plantation site and other statistics.

	Haplotype identification number	Plantation sites			
		A	M	U	L
Haplotypic frequencies per site	1/ 4/ 5/ 6/ 7/ 17/ 25/ 28/ 33/ 40	0.008			
	8/ 35/ 42/ 43/ 46/ 47/ 50/ 55				0.008
	10/ 11/ 12/ 52/ 53/ 61/ 63		0.008		
	13/ 14/ 39/ 44/ 45/ 48/ 57/ 60/ 62			0.008	
	2	0.016			
	3	0.008			0.008
	9		0.008	0.008	0.008
	15	0.008			0.008
	16	0.008	0.008		0.016
	18	0.008			0.008
	19	0.008	0.008		
	20	0.023			
	21	0.008	0.008		
	22	0.008		0.023	0.008
	23	0.008		0.008	
	24	0.008	0.008		
	26	0.023	0.016	0.008	
	27	0.008	0.008		
	29	0.008		0.008	
	30	0.023	0.031	0.023	0.023
	31	0.023	0.016		
	32	0.023	0.008		
	34	0.008	0.008	0.023	
	36			0.016	0.008
	37		0.008		0.008
	38		0.023	0.008	
	41				0.016
	49		0.008		0.008
	51		0.008		0.008
	54		0.008	0.008	
	56		0.008	0.008	
	58		0.008	0.016	0.008
	59		0.008	0.016	
Number of individuals		39	33	31	25
Private haplotypes		12	7	9	10
Maternal lineages		28	25	22	21
Haplotypic diversity		0.978	0.979	0.974	0.983
± standard deviation		0.011	0.015	0.015	0.017

DISCUSSION

The ISSR technique, as performed here, probed efficient in assessing genetic diversity and variation of cultivated "yerba mate" plants; the values obtained for PIC and RP are in the range of those documented for other non-model crops (Grativol et al. 2011), and suggest that YM6 is the most descriptive primer. Therefore, future genotyping attempts on wild or cultivated *Ilex paraguariensis* germplasms could benefit from using this marker, for instance, when a rapid pilot screening is necessary. The stringent conditions applied here for matrix construction and data processing, aided in minimizing banding pattern recording errors and ensured reliable results. In addition, the genomic information of each individual was maximized through banding patterns combination.

Surprisingly, our results indicate that the four Argentine "yerba mate" plantation sites maintain a genetic diversity (0.569) 3.5 times higher than that estimated for natural populations from Brazil (0.163; Gauer & Cavalli-Molina 2000). It must be considered, though, that Gauer & Cavalli-Molina (2000) recorded their data with other dominant markers (RAPDs) and using low resolution agarose gels. Then, the genetic heterogeneity uncovered here suggests that the productive plantations under study were established on highly variable materials (either as seeds or plantlets), a fact that is in correspondence with their multiple provenances, such as old plantations, local and institutional nurseries, and remnants of wild forests (Scherer 2001). It is also likely that germplasm exchange between "yerba mate" producers have favored the maintenance of high diversity; still, the life history traits exhibited by this species -like obligatory open pollination- could also have contributed. Nonetheless, the diversity level detected could be somehow overestimated due to the stringent polymorphism threshold employed (33%). The set of cultivated plants studied are also more diverse than other commercial crops screened with ISSR markers, such as *Jatropha curcas* L (0.434; Grativol et al. 2011), *Camelia sinensis* (L) Kuntze var. *assamica* (JW Mast.) Kitam. (0.418; Ji et al. 2011), and *Curcuma longa* L (0.377; Singh et al. 2012), but similar to the diversity shown by natural populations of *Coffea arabica* L (0.55; Aga et al. 2005) and sweet cherry cultivars (0.546; Ganopoulos et al. 2011). The level of heterogeneity revealed for the breeding population is similar to that estimated for 34 "yerba mate" individuals gathered from the species´ diversification center (mean Dice genetic distance= 0.298; range= 0.090-0.689; Gottlieb et al. 2011). Yet, Brazilian natural populations showed even more heterogeneity (mean distance= 0.413; Gauer & Cavalli-Molina 2000). Our results could be attributable to the fact that the breeding population studied is grounded on four gene pools that were most probably shaped by past exchange and transfer of materials among local producers and settlers.

In line with previous reports on *I. paraguariensis,* we found that most of the genetic variance (> 95%) is accumulated within each location surveyed. For instance, Gauer & Cavalli-Molina (2000), Neuman-Wendt et al. (2007, 2009) and Cascales et al. (2014), detected the greatest variation within populations or within plant´s provenance (85%, 87%, 88% and 59%, respectively); this structure was attributed to the species´ characteristics (i.e., obligatory outcrossing, perennial, and long-lived). Likewise, while studying the caffeine and theobromine contents of the same set of plants that have been genotyped here, Marx et al. (2003) obtained a variance accumulated within plantation sites (> 67%).

The microsatellite chloroplastic markers employed in the present study were assayed,

for the first time, over a large sampling. No previous genotyping efforts applying homologous cpSSR markers are available, in the literature, for "yerba mate". Of the 14 cpSSRs tested, our results validate five markers as they showed polymorphism in the analyzed samples. Because the plastid genome is a non-recombining molecule, all these chloroplastidic markers are physically linked and uniparentally inherited. Yet, we observed significant linkage disequilibrium solely between cpSSRs *rps18-rpl20* and *psbI-trnS,* suggesting that their alleles are unlikely to be derived by homoplasy (Angioi et al. 2009). In contrast, the occurrence of size homoplasy cannot be completely ruled-out for the alleles from cpSSRs *rpl14-rpl16, psbZ-trnS* and *accD-psaI*. Size homoplasy can lead to underestimates of population subdivision and genetic divergence between populations. However, as stated by Angioni et al. (2009) the low mutation rate of cpSSRs (in relation to nuclear microsatellites) makes size homoplasy a negligible problem when comparing intraspecific gene pools or closely related species. Therefore, our intraspecific survey allowed the recognition of high chloroplastidic diversity (0.505; Table V) and that the breeding population comprise individuals derived from, at least, 63 maternal lineages or haplotypes. This plastidic variation, though, resulted insufficient to distinguish robust groups and to converge into an optimal number of DAPC clusters (not shown). However, the haplotypes detected appear related in two groups. The low chloroplastidic differentiation found and the level of shared haplotypes among plantations could have been caused by anthropic manipulation of seeds and seedlings. Then, the low, yet significant, levels of differentiation detected with both nuclear and plastidic markers suggest the contribution of constraints to gene flow, and/or a low level of shared polymorphism among plantations. The fact that nuclear fixation index is higher than chloroplastidic F_{ST} may suggest some restriction in dispersion through pollen.

The phenotypic selection study performed by Scherer (2001) reported that "yerba mate" plants from site L were more uniform than those from the other sites. This was explained by Scherer (2001) through the selection processes carried-out at the "Establecimiento Stvas" across three owners´ generations. The uniformity mentioned above was not observed in the present genotyping analyses, as plants from site L did show nuclear and chloroplastidic characteristics like those of plants from the other sites. Thus, our outcomes reinforce the notion that, for these plants, the relationship between phenotypic and genotypic variation is not straightforward. Additionally, present results confirm that the selection cycles started from a sufficiently variable base population (Scherer 2001).

As a side outcome, it was detected two pairs of individuals showing the most contrasting fingerprinting patterns. These plants, namely M23 - U08 and U42 - L11, belong to different plantation sites, genetic groups, and maternal lineages, and involve compatible specimens (i.e., pistillate/ staminate). More importantly, these individuals were never crossed. Their utilization as target genotypes could bring renewed opportunities for development of novel varieties.

On one hand, our results highlight that the breeding population of "yerba mate" is sufficiently variable to ensure success to ongoing and forthcoming breeding programs. On the other hand, this genetically diverse population constitutes an interesting allelic pool with potentiality to cope, for instance, with the emergence of new or more virulent pathogenic diseases, or to satisfy market requirements. Considering the size of the base

population from which these plants were taken (ca. 900.000), it is clear that more genotypes are still to be uncovered.

Acknowledgments

The authors are grateful to Ing. Agr. A Nahirñak from Pindó S.A. (Misiones, Argentina) for helping during sample collections, and to Drs. GR Cueto and JC Vilardi for statistical assistance with the R environment. Microsatellite loci used in this study were designed by J Cascales as part of her Phd. Thesis (FCEN-UBA, 2018). Grants from the Universidad de Buenos Aires (UBACyT 20020100100859BA and 20020170100614BA), the Agencia Nacional de Investigaciones Científicas y Técnicas (PICT 2015-2525) and the Proyecto convocatoria especial UNAM (SCTyP I128) are gratefully acknowledged.

REFERENCES

AGA E, BEKELE E & BRYNGELSSON T. 2005. Inter-simple sequence repeat (ISSR) variation in forest coffee trees (*Coffea arabica* L.) populations from Ethiopia. Genetica 124: 213-221.

AL-TURKI TA & BASAHI MA. 2015. Assessment of ISSR based molecular genetic diversity of Hassawi rice in Saudi Arabia. Saudi J Biol Sci 22: 591-599.

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: 299-304.

ANGIOI SA, DESIDERIO F, RAU D, BITOCCHI E, ATTENE G & PAPA R. 2009. Development and use of chloroplast microsatellites in *Phaseolus* spp. and other legumes. Plant Biol 11: 598-612.

BERTÉ KAS, BEUX MR, SPADA PKWDS, SALVADOR M & HOFFMANN-RIBANI R. 2011. Chemical composition and antioxidant activity of yerba-mate (*Ilex paraguariensis* A. St.-Hil., Aquifoliaceae) extract as obtained by spray drying. J Agr Food Chem 59: 5523-5527.

BORSCH T & QUANDT D. 2009. Mutational dynamics and phylogenetic utility of noncoding chloroplast DNA. Plant Syst Evol 282: 169-199.

CASCALES J, BRACCO M, GARBEROGLIO MJ, POGGIO L & GOTTLIEB AM. 2017. Integral Phylogenomic Approach over *Ilex* L. species from Southern South America. Life 7: 47.

CASCALES J, BRACCO M, POGGIO L & GOTTLIEB AM. 2014. Genetic diversity of wild germplasm of "yerba mate" (*Ilex paraguariensis* St. Hil.) from Uruguay. Genetica 142: 563-573.

EXCOFFIER L, LAVAL G & SCHNEIDER S. 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinform Online 1: 47-50.

EXCOFFIER L, SMOUSE PE & QUATTRO JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131: 479-491.

GAIERO P, MAZZELLA C, AGOSTINI G, BERTOLAZZI S & ROSSATO M. 2011. Genetic diversity among endangered Uruguayan populations of *Butia* Becc. species based on ISSR. Plant Syst Evol 292: 105-116.

GANOPOULOS IV, KAZANTZIS K, CHATZICHARISIS I, KARAYIANNIS I & TSAFTARIS AS. 2011. Genetic diversity, structure and fruit trait associations in Greek sweet cherry cultivars using microsatellite based (SSR/ISSR) and morpho-physiological markers. Euphytica 181: 237-251.

GAUER L & CAVALLI-MOLINA S. 2000. Genetic variation in natural populations of maté (*Ilex paraguariensis* A. St.-Hil., Aquifoliaceae) using RAPD markers. Heredity 84: 647-656.

GIBERTI GC. 2008. Aquifoliaceae. In: Zuloaga F et al. (Eds), Catálogo de las Plantas vasculares del Cono Sur (Argentina, Sur de Brasil, Chile, Paraguay y Uruguay), Vol. 2. Dicotyledoneae: Acanthaceae–Fabaceae (Abarema – Schizolobium), Missouri: Monographs in Systematic Botany, Missouri Botanical Garden, p. 1143-1146.

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

GOLDSTEIN DB, RUIZ LINARES A, CAVALLI-SFORZA LL & FELDMAN MW. 1994. An Evaluation of Genetic Distances for Use With Microsatellite Loci. Genetics 139: 463-471.

GOTTLIEB AM & POGGIO L. 2010. Genomic screening in dioecious "yerba mate" tree (*Ilex paraguariensis* A. St. Hill., Aquifoliaceae) through representational difference analysis. Genetica 138: 567-578.

GOTTLIEB AM & POGGIO L. 2014. Quantitative and qualitative genomic characterization of cultivated *Ilex* L. species. Plant Genet Resour-C 13: 142-152.

GOTTLIEB AM, GIBERTI GC & POGGIO L. 2005. Molecular analyses of the genus *Ilex* (Aquifoliaceae) in southern South America, evidence from AFLP and its sequence data. Am J Bot 92: 352-369.

GOTTLIEB AM, GIBERTI GC & POGGIO L. 2011. Evaluación del germoplasma de *Ilex paraguariensis* e *Ilex dumosa* (Aquifoliaceae). Bol Soc Arg Bot 46: 113-123.

GRATIVOL C, DA FONSECA LIRA-MEDEIROS C, HEMERLY AS & FERREIRA PCG. 2011. High efficiency and reliability of inter-simple sequence repeats (ISSR) markers for evaluation of genetic diversity in Brazilian cultivated *Jatropha curcas* L. accessions. Mol Biol Rep 38: 4245-4256.

HECK CI & DE MEJÍA EG. 2007. Yerba mate tea (*Ilex paraguariensis*): A comprehensive review on chemistry, health implications and technological considerations. J Food Sci 72: R138-151.

JI PZ, LI H, GAO LZ, ZHANG J, CHENG ZQ & HUANG XQ. 2011. ISSR diversity and genetic differentiation of ancient tea (*Camellia sinensis* var. *assamica*) plantations from China: Implications for precious tea germplasm conservation. Pak J Bot 43: 281-291.

JOMBART T & COLLINS C. 2015. A tutorial for Discriminant Analysis of Principal Components (DAPC) using adegenet 2.0.0. http://adegenet.r-forge.r-project.org/files/tutorial-dapc.pdf.

JOMBART T, DEVILLARD S & BALLOUX F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Gen 11: 94.

KIANI M, MEMARIANI F & ZARGHAMI H. 2012. Molecular analysis of species of *Tulipa* L. from Iran based on ISSR markers. Plant Syst Evol 298: 1515-1522.

KUMAR A, MISHRA P, BASKARAN K, SHUKLA AK, SHASANY AK & SUNDARESAN V. 2016. Higher efficiency of ISSR markers over plastid *psbA-trnH* region in resolving taxonomical status of genus *Ocimum* L. Ecol Evol 6: 7671-7682.

LANGELLA O. 1999. Populations 1.2.32. Available via http://bioinformatics.org/~tryphon/populations/.

LEGENDRE P & LEGENDRE L. 1998. Numerical ecology. 2nd ed., Amsterdam: Elsevier, 853 p.

MARX F, JANSSENS MJJ, URFER P & SCHERER RA. 2003. Caffeine and theobromine composition of mate (*Ilex paraguariensis*) leaves in five plantations of Misiones, Argentina. Plant Food Hum Nutr 58: 1-8.

NEI M. 1978 Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89: 583-590.

NEI M. 1987. Molecular Evolutionary Genetics. Columbia University Press, New York, 512 p.

NEUMAN-WENDT S, DE SOUSA VA, QUOIRIN M, SEBBENN AM, MAZZA MC & STURION JA. 2007. Caracterização genética de procedências e progênies de *Ilex paraguariensis* St. Hil. utilizando marcadores RAPD. Sci For 73: 47-53.

NEUMAN-WENDT S, DE SOUSA VA, SEBBENN AM, STURION JA, LUÍZ F, SANTOS ECS & QUOIRIN M. 2009. Inheritance and linkage relationships of allozyme variants of *Ilex paraguariensis* St. Hil. Braz Arch Biol Tech 52: 1443-1451.

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 Pla Sci 1: 1200245.

PICCA A, HELGUERA M, SALOMÓN N & CARRERA A. 2004. Marcadores Moleculares. In: Echenique V et al. (Eds), Biotecnología y Mejoramiento vegetal, Ediciones INTA, Argentina, 61-68.

PRADEEP REDDY M, SARLA N & SIDDIQ EA. 2002. Inter simple sequence repeat (ISSR) polymorphism and its application in plant breeding. Euphytica 128: 9-17.

PREVOST A & WILKINSON MJ. 1999. A new system of comparing PCR primers applied to ISSR fingerprinting of potato cultivars. Theor Appl Gen 98: 107-112.

PROVAN J, POWELL W & HOLLINGSWORTH PM. 2001. Chloroplast microsatellites: new tools for studies in plant ecology and evolution. Trends Ecol Evol 16: 142-147.

R CORE TEAM. 2017. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/.

RAKOCEVIC M, JANSSENS MJJ & SCHERER RA. 2012. Light responses and gender issues in the domestication process of yerba-mate, a subtropical evergreen. In: Bezerra AD & Ferreira TS (Eds), Evergreens*:* Types, Ecology and Conservation, New York: Nova Science Publishers, p. 63-96.

RAMBAUT A. 2016. FigTree v1.4.3 software. Institute of Evolutionary Biology, University of Edinburgh. Retrieved from http://tree.bio.ed.ac.uk/software/figtree/.

RAYMOND M & ROUSSET F. 1995. An exact test for population differentiation. Evolution 49: 1280-1283.

RENDELL S & ENNOS RA. 2003. Chloroplast DNA diversity of the dioecious European tree *Ilex aquifolium* L. (English holly). Mol Ecol 12: 2681-2688.

ROLDAN-RUIZ I, DENDAUW J, VAN BOCKSTAELE E, DEPICKER A & DE LOOSE M. 2000. AFLP markers reveal high polymorphic rates in ryegrass (*Lolium* spp.). Mol Breed 6: 125-134.

ROUSSET F. 2008 GenePop'007: a complete re-implementation of the GenePop software for Windows and Linux. Mol Ecol Res 8: 103-106.

SAITOU N & NEI M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4: 406-425.

SCHERER RA, JANSSENS MJJ, MARX F, URFER P & SCHNEIDER E. 2006. Saponin content and quality-related traits of mass-selected yerba maté (*Ilex paraguariensis* A. St.-Hil.) trees. J Herbs, Spices Med Plants 12: 73-85.

SCHERER RA. 2001. Mass and nursery selection of yerba mate (*Ilex paraguariensis*) for quality improvement in Argentina.Curvillier Verlag Göttingen, Germany. 203 p.

SCHLÜTER PM & HARRIS SA. 2006. Analysis of multilocus fingerprinting data sets containing missing data. Mol Ecol Notes 6: 569-572.

SEYEDIMORADI H & TALEBI R. 2014. Detecting DNA polymorphism and genetic diversity in Lentil (*Lens culinaris* Medik.) germplasm: comparison of ISSR and DAMD marker. Physiol Mol Biol Pla 20: 495-500.

SINGH S, PANDA MK & NAYAK S. 2012. Evaluation of genetic diversity in turmeric (*Curcuma longa* L.) using RAPD and ISSR markers. Ind Crop Prod 37: 284-291.

SOKAL RR & ROHLF FJ. 1995. Biometry: the principles and practice of statistics in biological research, 3rd ed., New York, WH Freeman and Co, 887 p.

WHEELER GL, DORMAN HE, BUCHANAN A, CHALLAGUNDLA L & WALLACE LE. 2014. A review of the prevalence, utility, and caveats of using chloroplasts simple sequence repeats for studies of plant biology. Appl Plant Sci 2: 1400059.

YEH FC & BOYLE T. 1999. PopGene Version 131: Microsoft Window-based freeware for population genetic analysis. Retrieved from https://sites.ualberta.ca/~fyeh/popgene_download.html.

YEH FC & BOYLE TJB. 1997. Population genetic analysis of co-dominant and dominant markers and quantitative traits. Belg J Bot 129: 157.

ZIETKIEWICZ E, RAFALSKI A & LABUDA D. 1994. Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. Genomics 20: 176-183.

SUPPLEMENTARY MATERIAL

Table SI. ISSR primer names with their respective nucleotide sequences, and chloroplast microsatellite (cpSSR) loci assayed with their respective forward (F) and reverse (R) sequences and repeat motives. In bold type, the loci selected for further study.

Table SII. Results from linkage disequilibrium analysis between pairs of cpSSR informative loci, in the overall sample (Fisher's method). The degree of freedom was eight for all comparisons.

Figure S1. Group assignment for each "yerba mate" individual plant derived from DAPC analysis. Red blocks indicate full assignment probability (p=1.0). Crosses represent an individual. Sites are indicated by their identification code (see Table I); inferred clusters are indicated in the lower row with numbers (1-4).

Figure S2. Loading plot showing allele contribution to discriminant analysis. The grey line marks the threshold default value. Numbers above each line indicate the PCR primer used.

Figure S3. Midpoint-rooted Neighbor-Joining dendrogram of "yerba mate" individual haplotypes. Each terminal (or leaf) corresponds to an individual plant named with its corresponding haplotype number (H1 – H63; see Table VI) followed by the alphanumeric code of each plant and its ISSR genetic cluster number. The two haplogroups are indicated.

How to cite

PAIVA DI, CASCALES J, ROSETTI MEN, SCHERER RA, GAUCHAT ME & GOTTLIEB AM. 2020. Unraveling the genetic complexity of a cultivated breeding population of "yerba mate" (*Ilex paraguariensis* St. Hil.). An Acad Bras Cienc 92: e20190113. DOI 10.1590/0001-3765202020190113.

Manuscript received on January 31, 2019;
accepted for publication on May 26, 2019

DANIELA IVANA PAIVA[1,3]
https://orcid.org/0000-0002-5564-479X

JIMENA CASCALES[2,3]
https://orcid.org/0000-0003-0854-6385

MARÍA EVA NATALIA ROSETTI[2,3]
https://orcid.org/0000-0001-8849-9181

RAFAEL ALEJANDRO SCHERER[4]
https://orcid.org/0000-0002-2913-0437

MARÍA ELENA GAUCHAT[1]
https://orcid.org/0000-0002-1141-5653

ALEXANDRA MARINA GOTTLIEB[2,3]
https://orcid.org/0000-0002-7620-8276

[1]Estación Experimental Agropecuaria Montecarlo, Instituto Nacional de Tecnología Agropecuaria (EEA-INTA Montecarlo), Av. El Libertador, 2472, 3384, Misiones, Argentina

[2]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

[3]Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Godoy Cruz, 2290, C1425FQB, Ciudad Autónoma de Buenos Aires, Argentina

[4]Pindo S.A., Juan Domingo Perón 303, 3378, Puerto Esperanza, Misiones, Argentina

Correspondence to: **Alexandra Marina Gottlieb**
E-mail: gottlieb@ege.fcen.uba.ar

Author contributions

DIP: collected samples, performed the gathering and statistical analysis of the data, prepared the first draft of the manuscript. JC: designed the microsatellite loci used, assisted in wet lab procedures. MENR: helped DIP on linkage disequilibrium analysis and interpretation of results. RAS facilitated access to plant material and databases and contributed with some financial support. MEG designed the field collection and helped collecting samples; also discussed results and contributed to the manuscript. AMG designed the study, performed some illustrations, wrote the manuscript and provided financial support. MEG and AMG were the advisors of DIP. In addition, all authors edited and critically revised the manuscript, and approved the final version.