ISSN 1479-2621
Plant Genetic Resources: Characterization and Utilization; 1–11
doi:10.1017/S1479262114000756

Quantitative and qualitative genomic characterization of cultivated *Ilex* L. species

Alexandra Marina Gottlieb[1,2]* and Lidia Poggio[1,2]

[1]*Laboratorio de Citogenética y Evolución (LaCyE), Departamento de Ecología, Genética y Evolución, Instituto IEGEBA (UBA-CONICET), 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 and* [2]*Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Avenida Rivadavia 1917, C1033AAJ Ciudad Autónoma de Buenos Aires, Argentina*

Received 18 March 2014; Accepted 31 May 2014

Abstract

The development of modern approaches to the genetic improvement of the tree crops *Ilex paraguariensis* ('yerba mate') and *Ilex dumosa* ('yerba señorita') is halted by the scarcity of basic genetic information. In this study, we characterized the implementation of low-cost methodologies such as representational difference analysis (RDA), single-strand conformation polymorphisms (SSCP), and reverse and direct dot-blot filter hybridization assays coupled with thorough bioinformatic characterization of sequence data for both species. Also, we estimated the genome size of each species using flow cytometry. This study contributes to the better understanding of the genetic differences between two cultivated species, by generating new quantitative and qualitative genome-level data. Using the RDA technique, we isolated a group of non-coding repetitive sequences, tentatively considered as *Ilex*-specific, which were 1.21- to 39.62-fold more abundant in the genome of *I. paraguariensis*. Another group of repetitive DNA sequences involved retrotransposons, which appeared 1.41- to 35.77-fold more abundantly in the genome of *I. dumosa*. The genomic DNA of each species showed different performances in filter hybridizations: while *I. paraguariensis* showed a high intraspecific affinity, *I. dumosa* exhibited a higher affinity for the genome of the former species (i.e. interspecific). These differences could be attributed to the occurrence of homologous but slightly divergent repetitive DNA sequences, highly amplified in the genome of *I. paraguariensis* but not in the genome of *I. dumosa*. Additionally, our hybridization outcomes suggest that the genomes of both species have less than 80% similarity. Moreover, for the first time, we report herein a genome size estimate of 1670 Mbp for *I. paraguariensis* and that of 1848 Mbp for *I. dumosa*.

Keywords: flow cytometry; genome filter hybridization; representational difference analysis; reverse dot-blot hybridization

Introduction

The perennial trees *Ilex paraguariensis* St. Hill. and *Ilex dumosa* Reissek (*Aquifoliaceae* Bartl.) are native to tropical and subtropical South America, which are frequently found in southern Brazil, south-eastern Paraguay and north-eastern Argentina; only *I. paraguariensis* reaches Uruguay as patchy populations (Giberti, 1989, 2001, 2011). These species are dioecious trees, insect-pollinated, and have the same diploid chromosome number ($2n = 40$; Barral *et al.*, 1995; Greizerstein *et al.*, 2004).

* Corresponding author. E-mail: gottlieb@ege.fcen.uba.ar

The aerial parts of *I. paraguariensis*, the 'yerba mate' tree, are commercialized in the southern countries of South America to prepare a popular beverage called 'mate', highly appreciated for its peculiar flavour and stimulating properties. Recently, countries in the Middle East have incorporated the custom of drinking 'mate', with Syria being the largest consumer. Its choleretic, hypocholesteraemic, antioxidant and hepatoprotective properties (Berté *et al.*, 2011; Bracesco *et al.*, 2011) are attributed to the high concentration of xanthines, flavonoids, saponins, amino acids, minerals and vitamins (Heck and de Mejia, 2007). *I. dumosa*, known as 'yerba señorita', has comparatively mild effects due to a significant lower content of caffeoyl derivatives and flavonoids (32 and 10 times, respectively), and methylxanthine content varying from little to none (Filip *et al.*, 1998, 2001). For a long time, *I. dumosa* has been used as an adulterant (Giberti, 1989); however, nowadays, it is legally marketed as a low-caffeine alternative.

Notably, the commercial production of 'yerba mate' is maintained by intensive clonal propagation of a small number of genotypes, restricting its genetic background and its response to pests and diseases. *I. dumosa* constitutes a natural reservoir of agronomically interesting genes, and is a suitable potential donor for introgression into the commercial lines of 'yerba mate', as it shows increased resistance to drought conditions and to the main pest (*Gyropsylla spegazziana*, *Psyllidae*) of 'yerba mate' (Prat Kricun, 2009).

Although the species have great socio-economic importance, the scarcity of basic genetic data halts the implementation of modern approaches to genetic improvement. Formal molecular microevolutionary studies concur in highlighting that the natural populations of 'yerba mate' still available in the wild, have a narrow genetic base (Gauer and Cavalli-Molina, 2000; Neumann-Wendt, 2005; Cascales, 2013). So far, only one study has characterized *I. dumosa* using molecular markers (Gottlieb *et al.*, 2011). In this context, the generation of genetic information about *I. paraguariensis* and *I. dumosa* will aid ongoing traditional improvement activities by enhancing the development of more modern programmes. To contribute to their knowledge, we initiated an exploration for specific distinctive features through a comparative characterization, using relatively low-cost methodological molecular approaches. We implemented a subtractive hybridization technique, such as the representational difference analysis (RDA; Lisitsyn *et al.*, 1993), coupled with single-strand conformation polymorphism (SSCP; Orita *et al.*, 1989) analysis of amplicons, bioinformatic characterization of nucleotide sequence data, and a series of reverse and direct dot-blot filter hybridization assays. Additionally, we estimated the genome sizes of the two species using flow cytometry. Therefore, we herein report for the first time the quantitative and qualitative differences between the genomes of *I. paraguariensis* and *I. dumosa*. Our approach could encourage research in other non-model plants, for which cutting-edge technologies (namely Next Generation Sequencing (NGS)) are still economically and methodologically unfeasible.

Materials and methods

Plant materials

Leaf samples of *I. paraguariensis* and *I. dumosa* were provided by the 'Banco de Germoplasma de Yerba Mate y Té' at the 'Estación Experimental INTA Cerro Azul (EEINTA-CA)', Misiones, Argentina. A total of two allopatric accessions were selected: *I. paraguariensis* EEINTA-CA no. CA1-74 (Misiones, Argentina), which is catalogued as the best leaf producer (Belingheri and Prat Kricum, 2000), and *I. dumosa* var. *dumosa* EEINTA-CA no. 20 (Santa Catarina, Brazil), which represents the wild phenotype (G. C. Giberti, pers. commun.). For DNA extraction and subsequent RDA assays, four individual plants per accession were used.

To prepare genomic probes for dot-blot assays, we used leaf samples of 'yerba mate' from accessions EEINTA-CA no. CA1-74 and CA4-74 and from six materials (namely SI-16♂, SI-19♀, SI-49♂, SI-67♀, G-18♀ and Y-383♂; Misiones, Argentina), kindly provided by 'Establecimiento Las Marías' (Gdor. Virasoro, Corrientes, Argentina). For *I. dumosa*, we used the leaf samples from the accession EEINTA-CA no. 20 and Argentinean accessions EEINTA-CA no. 7, 49, 48 and 222, and BACP no. 606, 607, 608, 609, 610 and 611 (Herbario del Centro de Estudios Farmacológicos y Botánicos – CEFYBO, held at "Museo de Farmacobotánica de la Facultad de Farmacia y Bioquímica", Universidad de Buenos Aires, Argentina).

For flow cytometry analysis, we obtained fresh leaf samples from seven plantlets ('yerba mate': SI-19, SI-49, SI-67, YM1 and YM sn; 'yerba señorita': CRV7/4, CRV48/10, YS1 and YS sn), kindly donated by Dr P. A. Sansberro ('Instituto de Botánica del Nordeste IBONE', Universidad Nacional del Nordeste, Corrientes, Argentina).

DNA extraction

Young leaves preserved in silica gel were used for the extraction of total genomic DNA with the DNeasy Plant kit (Qiagen Inc., Valencia, California, USA), according to the manufacturer's instructions. Quality control and quantification were carried out by agarose gel (0.8%, w/v) electrophoresis and by comparison with a DNA molecular

size standard (Lambda EcoRI/HindIII; Promega Corp., Madison, Wisconsin, USA). Gels were stained with ethidium bromide and photographed under UV light.

Representational difference analysis

The RDA can be performed without prior sequence information, and implies a subtractive hybridization technique applied to a representation of the genome (Zoldos *et al.*, 2001). This technique relies on the occurrence of missing restriction sites and/or genome rearrangements, such as deletions, insertions, duplications and translocations (Chen *et al.*, 1998). We performed two reciprocal RDA assays. In the first assay, pooled genomic DNA from *I. paraguariensis* was employed as the tester (i.e. used in limited amounts) and that from *I. dumosa* was used as the driver to block the sequences shared by the two species, hereafter referred to as the RDA-Pa/Du assay. In the second assay, *I. dumosa* was used as the tester and *I. paraguariensis* as the driver, hereafter referred to as the RDA-Du/Pa assay. The RDA was performed, following the procedure proposed by Panaud *et al.* (2002), using approximately 400 ng of each pooled genomic DNA. The first oligonucleotide adapter set (R-set) was ligated to restriction fragments. Amplicons were obtained via polymerase chain reaction (PCR) as described previously by Panaud *et al.* (2002). The subtraction step was initiated by removing the R-adaptors from tester amplicons and then ligating the second set of adaptors (N-set). The N-ligated tester was mixed with the R-ligated driver in a 1:100 ratio, and ethanol-precipitated overnight (− 20°C). The following steps (resuspension, hybridization, PCR amplification and nuclease treatment) were performed as described previously by Panaud *et al.* (2002). A second PCR was run for 20 cycles. Prior to the second round of subtractive hybridization, the N-adaptors were removed from the tester and replaced by the J-adaptor set. In this round, tester-J/driver-R ratios of 1:100, 1:500 and 1:1000 were assayed. In the third round, the J-adaptors were removed from the tester and replaced by the N-adaptor set. In this cycle, tester-N/driver-R ratios of 1:100, 1:500 and 1:1000 were used. Provisional results were checked at each step by electrophoresis in 2% (w/v) agarose gels, as described previously. The tester-N/driver-R amplicons derived from the third round (1:1000 ratio) were used for cloning with the pGEM-T Easy Vector Cloning Kit (Promega Corporation, Madison, Wisconsin, USA), following the manufacturer's instructions. Isolated white colonies were only selected for purification of the plasmid vector plus the insert, which was achieved using the QIAprep Spin Miniprep kit (Qiagen Inc., Valencia, California, USA).

Nucleotide screening of clones

The nucleotide variation in isolated RDA clones was initially screened via SSCP analyses (Orita *et al.*, 1989), following the procedures described by Gottlieb and Poggio (2010). PCR amplicons were obtained from each purified clone by using universal primers T7 and SP6. Air-dried polyacrylamide gels were scanned and visually analysed. Clones were clustered according to their banding patterns.

Sequence characterization

From each SSCP cluster, at least one clone was selected for its characterization. Nucleotide sequences were obtained with universal primers T7 and SP6 at the 'Servicio Interno de Genotipificación y Secuenciación de ADN–INTA' (Castelar, Buenos Aires, Argentina), using an Applied Biosystems (Thermo Fischer Scientific, Waltham, Massachusetts, USA) automated 3130 XL DNA sequencer. Chromatograms were proofread and sequences manually edited in BioEdit (Hall, 1999). Boundaries of the plasmid vector and of the oligonucleotide adaptors were determined and eliminated. All sequences were compared with each other, and the G+C content was calculated. Blast searches were performed using the BlastN, TBlastX and BlastX algorithms of the NCBI (http://www.ncbi.nlm.nih.gov/Blast) and *E*-value threshold (probability of random match) $\leq 10^{-10}$. Additional searches were conducted against the Gypsy Database (Lloréns *et al.*, 2008, 2011) and the Conserved Domain Database (CDD) of the NCBI (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) using a threshold $\leq 10^{-6}$. The RepeatMasker program (version open-4.0.1; Smit *et al.*, 1996–2010) was used for the screening of interspersed repeats and low-complexity DNA sequences, assuming queries to be *Arabidopsis thaliana*. The sequences obtained herein were deposited in GenBank (HN280454–HN280499).

Dot-blot filter hybridization assays

Species-specificity assessment of the RDA-derived fragments and the estimation of their copy number within the genome of each species were performed through reverse dot-blot assays in triple- to quintuplicate, following the protocol described by Sabot *et al.* (2004). The digoxigenin labelling and detection kit (Roche, Pacheco, Buenos Aires, Argentina) was employed, according to the manufacturer's protocols, to generate non-radioactive species-specific genomic probes. We pooled equal amounts of total DNA from eight *I. paraguariensis* plants, on the one hand, and from

11 *I. dumosa* plants, on the other hand. A minimum of 3 μg from each bulked genomic DNA were labelled independently. The selected RDA-derived fragments were blotted onto a positively charged nylon filter membrane (Genescreen Plus, Perkin Elmer Inc., Boston, Massachusetts, USA), in aliquots (1–2 μl) of known amounts and fixed under UV light, according to the manufacturer's instructions. As controls for hybridization reactions and internal references for quantification of the signals, we blotted aliquots (1 μl) from serial dilutions (up to 156 pg/μl) of unlabelled bulked genomic DNA with respect to the probe used.

The molecular affinity between the genomes of the two species was assessed through direct dot-blot assays, in quadruplicates. Serially diluted DNA of each species (up to 110 pg/μl) was consecutively blotted, in 1 μl aliquots onto the same hybridization filter membrane. Each row functioned either as the control for hybridization reactions (intraspecific controls) and as the reference for quantification of the signals, or as the hybridization targets (i.e. for interspecific hybridizations). A control for the staining procedure, provided by the kit, was included in each assay. For each hybridization temperature assayed, two such membranes were prepared. At the established temperature, each membrane was hybridized with a single digoxigenin-labelled genomic probe. Optimal hybridization temperatures were calculated using the formula suggested by the kit which considers: fragment length, G+C content and average melting temperature subtracting 20°C or 25°C. The resulting optimal temperatures were 40 and 50°C. Post-hybridization washes comprised two 5 min washes in 2× saline–sodium citrate (SSC) and 0.1% sodium dodecyl sulphate (SDS) at room temperature, followed by two 15 min washes in 0.5× SSC and 0.1% SDS at the corresponding hybridization temperature. The experimental conditions used are regarded as stringent, allowing up to 18% mismatches between the probe and the targets. Staining was performed according to the manufacturer's protocol. Hybridization signals were quantified with Adobe Photoshop using the rapid selection tool and histogram function, and comparing the intensity value of each query dot with dot intensities of known amounts of the corresponding genomic DNA. Following Linares Miquel (1997), we assumed that differences in the intensity of hybridization signals were due to the differences in copy number. For estimation of copy number, we considered the quantity and the length of the blank DNA blotted, and the quantity of the reference DNA blotted and its genome size (Linares Miquel, 1997). For *I. paraguariensis* and *I. dumosa*, we used the genome sizes estimated here (in pg; see below) multiplied by 0.978×10^{-9} to obtain the DNA content in base pairs (bp) (Doležel *et al.*, 2003).

Flow cytometry

Nuclear suspensions were prepared by simultaneously chopping 1 cm^2 of fully expanded leaves (maintained either in darkness for 24 h or under natural conditions) or roots of *Ilex* spp., and 1 cm^2 of the internal standard leaf (i.e. *Zea mays* cultivar CE-777, 2C = 5.43 pg; Lysák and Doležel, 1998) using a razor blade and 2.5 ml of LB01 buffer (Doležel *et al.*, 1989) supplemented with 0.4% (w/v) of polyvinylpyrrolidone 10 and bovine serum albumin. The crude suspension was treated with 10 μl RNase (20 mg/ml), filtered through a nylon mesh (50 μm) and incubated for approximately 30 min at room temperature. Then, 500 μl propidium iodide (50 mg/ml) were added, and staining was allowed to develop for either 30 or 5 min in darkness. The measurements were made on a CyFlow Ploidy Analyser (Partec GmbH, Münster, Germany) at the facility of the 'Instituto de Floricultura – INTA' (Castelar, Buenos Aires, Argentina). Histograms were evaluated using the Flowing software 2.5.0 (P. Terho, Turku Centre for Biotechnology, University of Turku, Finland). For each sample, at least 5000 particles were analysed. Estimates of nuclear DNA content were considered only when the coefficients of variation were below 5%. The genome size was calculated by multiplying the DNA content of the standard (in pg) by the ratio of the mean fluorescent intensity of *Ilex* spp. to that of the standard. The seeds of the internal reference were kindly provided by Dr J. Doležel (Laboratory of Molecular Cytogenetics and Cytometry, Institute of Experimental Botany, Czech Republic). Estimation of genome sizes (in bp) was done as explained above.

Results

The RDA clones isolated from both assays (208 in total) were grouped according to their SSCP banding patterns in 51 clusters (data not shown). One to ten representatives of each cluster (67 in total) were sequenced and characterized with bioinformatic tools and similarity searches; in total, 7318 bp were generated for *I. paraguariensis* and 7420 bp for *I. dumosa* (4124 and 5879 bp, respectively, when redundancy is discounted). Based on the characterization, nucleotide data were arranged in 46 sequence types and, additionally, in three major groups (see Tables S1 and S2, available online). The first group involved sequences showing either spurious affinities or no similarity with any sequence deposited in the databases surveyed (group A, $N = 91$; Fig. 1(a)). The second group showed significant blast hits with plant entries at either the nucleotide or protein level (group B; $N = 67$). Eleven sequence types were recognized as conserved domains of retrotransposons

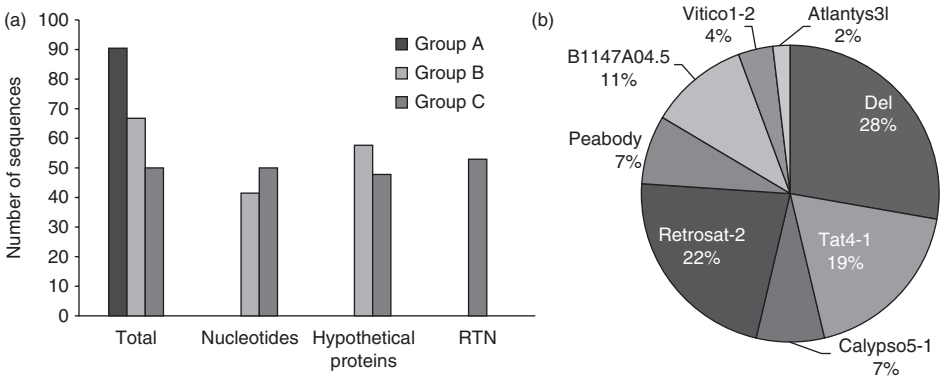

Fig. 1. (a) Schematic representation of the results from bioinformatic characterization and similarity searches. The description of groups A–C is detailed in the Results section. 'Total' refers to the total number of sequences per group. 'Nucleotides' refer to the number of sequences showing significant blast hits with entries from the databases surveyed at the nucleotide level. 'Hypothetical proteins' refer to the number of sequences showing significant blast hits to protein targets. 'RTN' refers to the number of sequences showing significant blast hits to retrotransposons. (b) Pie chart showing the retroelement composition in the representational difference analysis-derived sequences.

(namely aspartic peptidase, chromodomain, integrase, polyprotein precursor, RNaseH and retrotranscriptase), which belonged to seven Ty3/Gypsy and Copia lineages (Fig. 1(b)). Moreover, the RepeatMasker program applied on group B detected an integrase domain of the Atlantys3I long terminal repeat (LTR)/Gypsy retrotransposon from *Arabidopsis* (sequence type 2.9; 63% identity; 72% query coverage). In addition, the CDD search detected conserved domains of metabolic enzymes (i.e. calcium-binding motif, tyrosine kinase catalytic domain and nucleoside diphosphate kinase active site). For sequence types 1.7, 2.13, 2.14 and 2.15, no additional information was obtained, and thus remained unidentified. The third group of sequences matched plant ribosomal, mitochondrial or plastid DNA (group C; $N = 50$).

The dot-blot assays performed with isolated RDA-derived fragments showed differences in hybridization signals, at both temperatures (40 and 50°C) and with both genomic probes (Fig. 2(a)). The estimation of copy number for the unidentified sequences (group A) indicated that *I. paraguariensis* had 1.20×10^5 copies on average (range 3.95×10^4–3.22×10^5) and that *I. dumosa* had 6.45×10^4 copies (range 1.74×10^3–2.66×10^5). Most fragments appeared more abundant in the genome of *I. paraguariensis* (range 1.21- to 39.62-fold; Table 1). For retrotransposon sequences, the genomes of both species showed a similar average content; *I. paraguariensis* had 1.14×10^5 copies (range 1.85×10^3–2.88×10^5), whereas *I. dumosa* contained 1.08×10^5 copies (range 2.03×10^3–2.63×10^5). However, most fragments appeared more abundant in the genome of *I. dumosa* (range 1.41- to 35.77-fold; Table 1).

The dot-blot assays for testing interspecific genomic affinity showed that, at a more stringent condition (50°C), only intraspecific hybridization signals were detected. At the lower stringency (40°C), both intra- and interspecific signals were evident, but *I. paraguariensis* was always detected more efficiently, even with genomic probes from *I. dumosa*. The quantification of signal intensities indicated that *I. dumosa* required, on average, 4.16–22 times higher amounts of DNA to produce dot intensities equivalent to those yielded by *I. paraguariensis* (Fig. 2(b)).

The estimation of genome size by means of flow cytometry suggested that, under the standard conditions (30 min nuclei staining), the cytosolic compounds of *I. paraguariensis* severely interfered with the dying process, affecting even the internal standard and impeding any measurement (data not shown). Depending on the samples surveyed from *I. dumosa*, we detected two opposite performances; the cytosolic crude extracts cause either a severe interference of nuclei staining or only a mild effect. When the nuclei staining procedure was developed for 5 min, it yielded clear peaks for all the samples, tissues (leaves or roots) and growth conditions (light and darkness) (Fig. 3). Thus, for *I. paraguariensis*, the estimated genome size value was $2C = 1.71 \pm 0.08$ pg (=1670 Mbp) and for *I. dumosa* the value was $2C = 1.89 \pm 0.07$ pg (=1848 Mbp).

Discussion

The trees *I. paraguariensis* ('yerba mate') and *I. dumosa* ('yerba señorita') possess peculiar phytochemical properties valued for human consumption, for which they have been largely studied (for a review, see Heck and de Mejia, 2007). However, information about the basic genetic aspects of these species is scarce. At present, only the natural populations of *I. paraguariensis* have been formally investigated under a microevolutionary framework (Gauer and Cavalli-Molina, 2000; Neumann-Wendt, 2005; Cascales, 2013; Pereira *et al.*, 2013).

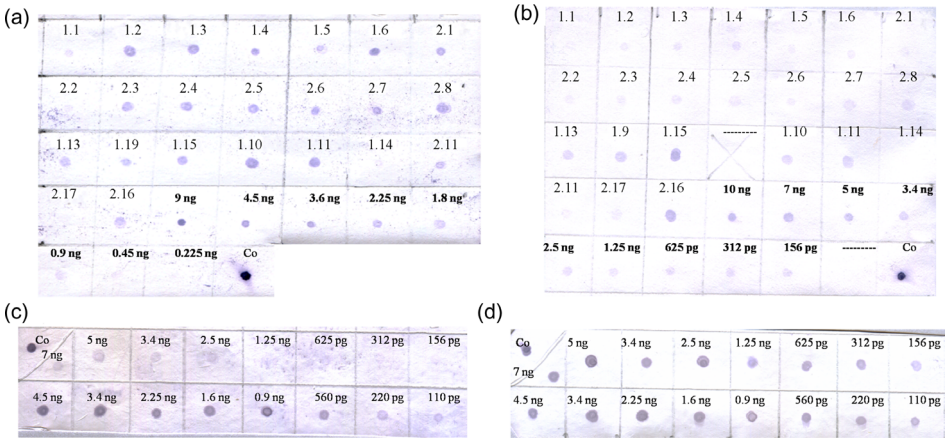

Fig. 2. (a) Examples of filter hybridizations of representational difference analysis-derived fragments by means of reverse dot-blot assays. The membrane showing dot blots from fragments of groups A and B, hybridized with a digoxigenin (DIG)-labelled genomic probe from *Ilex paraguariensis*. The amount (in ng) of unlabelled genomic DNA from *I. paraguariensis* is indicated in bold font. The dot blots are named as in Table 1. A staining control (Co) dot is included. (b) The membrane showing dot blots hybridized with a DIG-labelled genomic probe from *Ilex dumosa*. The amount (in ng) of unlabelled *I. dumosa* genomic DNA is indicated in bold font. (c) Examples of filter hybridizations of genomic DNA by means of direct dot-blot assays. Hybridization with a DIG-labelled genomic probe from *I. paraguariensis* is shown. The dot blots of the upper row correspond to the Co and to the serial dilutions of unlabelled total DNA from *I. dumosa*; the dot blots of the lower row correspond to the serial dilutions of unlabelled total DNA from *I. paraguariensis* (i.e. intraspecific control). Here, 3.4 ng of genomic DNA from *I. dumosa* have an intensity equal to 0.11 ng of genomic DNA from *I. paraguariensis*. (d) Hybridization with a DIG-labelled genomic probe from *I. dumosa* is shown. Here, the dot blots of the upper row correspond to the Co and to the intraspecific control (i.e. serial dilutions of unlabelled total DNA from *I. dumosa*); the dot blots of the lower row correspond to the serial dilutions of unlabelled total DNA from *I. paraguariensis*. Here, 7 ng of genomic DNA from *I. dumosa* are equivalent to 4.5 ng of genomic DNA from *I. paraguariensis*.

Those studies agree in showing evidence for low population divergence, explained by the occurrence of gene flow mediated by birds and/or by a recent fragmentation of habitats. Although such population data are lacking for *I. dumosa*, the phylogenetic and genetic relationships for these species have been assessed (Gottlieb *et al*., 2005, 2011). Still, the paucity of molecular information can be evidenced by a simple GenBank search. Currently, in GenBank, there are 77 entries for *I. paraguariensis*, most of which are microsatellite sequences (41), followed by chloroplast sequences (22) and by nuclear entries (14) corresponding to nine different regions. For *I. dumosa*, the data available are even fewer; there are only nine nuclear entries for three regions and four plastid sequences. In this study, we have characterized 67 genomic regions with bioinformatic tools, thus increasing the availability of non-redundant sequence data by 18% for *I. paraguariensis* and by 114% for *I. dumosa*.

Moreover, the information provided here demonstrates that the two species differ quantitatively and qualitatively in highly repetitive DNA, and in their genome sizes. The repeated recovery of some genomic regions indicates that the SSCP, while helped detecting sequence redundancies, overestimated the actual nucleotide variation of some fragments; at the same time, it suggests that isolated regions mostly involve repetitive DNA. A quantitative difference detected between the species relies on the group of non-coding sequences (group A), which appear on average 8.5 times more abundantly in the genome of *I. paraguariensis* than in the genome of *I. dumosa*. The high copy numbers estimated for this group clearly support their classification as highly repeated DNA. It is known that hybridization signals revealed with total DNA probes, such as those used here, are significant only if blotted sequences are highly repeated (Sabot *et al*., 2004). The fact that this group shows no significant similarity with any other sequence data deposited in public repositories inclines us to follow the criteria applied by Nagarajan *et al.* (2008) and to consider these as genomic regions specific for the genus *Ilex*. These repetitive sequences will be used as molecular probes in fluorescence *in situ* hybridization assays, to examine their distribution pattern over the chromosomal complement. Also, they have the potentiality for development of random and/or region-specific markers, with specificity for these particular species and with potentiality for future exploration of breeding populations.

Our outcomes also indicate that genomes of both species show a similar average copy number of the mobile elements (group B) isolated herein. We are aware that our estimates may be biased since we cannot distinguish full-length elements (i.e. autonomous) from

Table 1. Estimation of copy number for selected sequence types of groups A and B

Sequence type[a]	Copy number in *Ilex paraguariensis* genome[b]	Copy number in *Ilex dumosa* genome[b]	Ratio[c]
Group A			
Pa/Du-RDA			
1.1	*3.95 × 10^4*	*1.74 × 10^3*	22.7
1.2	7.50 × 10^4	2.20 × 10^4	3.41
1.3	5.05 × 10^4	3.78 × 10^4	1.34
1.4	6.05 × 10^4	6.49 × 10^4	0.93 (1.07)
1.5	6.59 × 10^4	4.06 × 10^4	1.62
1.6	7.66 × 10^4	1.93 × 10^3	39.62
1.8	1.27 × 10^4	6.10 × 10^4	0.21 (4.81)
Du/Pa-RDA			
2.1	5.96 × 10^4	3.22 × 10^4	1.85
2.2	2.29 × 10^5	3.00 × 10^4	7.65
2.3	**3.22 × 10^5**	**2.66 × 10^5**	1.21
2.4	2.33 × 10^5	1.89 × 10^5	1.23
2.5	1.70 × 10^5	2.73 × 10^4	6.22
2.6	1.41 × 10^5	9.93 × 10^3	14.24
2.7	9.22 × 10^4	8.08 × 10^4	1.14
2.8	7.14 × 10^4	9.81 × 10^4	0.73 (1.37)
2.10	9.11 × 10^4	4.05 × 10^4	2.25
2.13	1.80 × 10^5	6.60 × 10^4	2.73
2.14	2.00 × 10^5	9.11 × 10^4	2.20
Group B			
Pa/Du-RDA			
1.9	9.26 × 10^4	1.66 × 10^5	0.56 (1.8)
1.10	7.89 × 10^4	1.11 × 10^5	0.71 (1.41)
1.11	9.12 × 10^4	**2.63 × 10^5**	0.35 (2.88)
1.13	*1.85 × 10^3*	6.62 × 10^4	0.03 (35.77)
1.14	2.83 × 10^5	1.56 × 10^5	1.81
1.15	3.37 × 10^4	*2.03 × 10^3*	16.61
1.16	1.07 × 10^5	1.13 × 10^5	0.95 (1.06)
Du/Pa-RDA			
2.11	6.45 × 10^4	5.47 × 10^3	11.78
2.16	**2.88 × 10^5**	1.32 × 10^5	2.18
2.17	4.06 × 10^4	1.27 × 10^5	0.32 (3.14)

RDA, representational difference analysis.
[a] For GenBank accession numbers, see Tables S1 and S2 (available online). [b] Average of triple to quintuple replicates. The numbers in bold and italic fonts represent the maximum and minimum values per group of sequences, respectively. [c] Ratio of the average copy number estimate of *I. paraguariensis* to *I. dumosa*; the values given in parentheses indicate the inverse ratio.

truncated retrotransposons (i.e. non-autonomous, possibly nested, defective or inactive), though similar estimates have been reported for other plants. The genome of cocoa tree, *Theobroma cacao*, was estimated to contain approximately 6.75 × 10^4 copies of transposable elements, although this estimate was regarded as a rough underestimation (Argout *et al*., 2011). The genome of *Populus trichocarpa* has over 5000 copies of retrotransposons with a prevalence of Gypsy-like elements (Tuskan *et al*., 2006); recently, 1479 full-length LTR retrotransposons were identified in poplar (Cossu *et al*., 2012). In *Eucalyptus*, 24–226 copies of transcriptionally active Copia-like LTR retrotransposons have been reported (Marcon *et al*., 2011). A preliminary analysis of repetitive DNA from pilot bacterial artificial chromosome (BAC)-end sequences indicates that over 98% of the genome of the tea crop, *Camellia sinensis*, could be repetitive (Lin *et al*., 2011). We detected conserved domains associated with the replication and integration of mobile elements from the Ty3/Gypsy Del lineage, Chromoviridae branch (i.e. Del, Retrosat-2 and Peabody), the Athila/Tat lineage (i.e. Calypso5-1, Tat4-1 and B1147A04.5) and the Ty1/Copia Retrofit clade (i.e. Vitico1-2) according to the current classification of retrotransposons (Lloréns *et al*., 2008, 2011). We found that the *Ilex paraguariensis* retrotransposon element (IPRE) retroelement from *I. paraguariensis* (Gottlieb and Poggio, 2010) is also present in *I. dumosa* (sequence type 2.16). This is somehow expectable because the Chromoviridae branch is ancient and

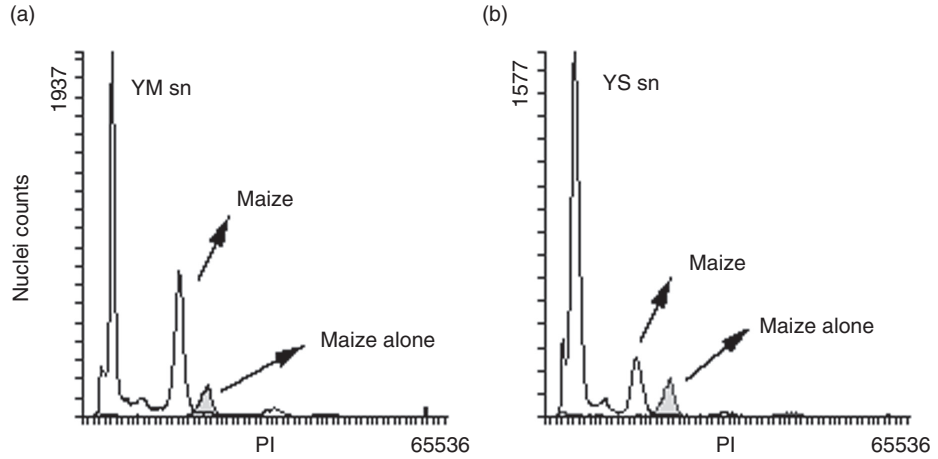

Fig. 3. Overlay flow cytometry histograms [nuclei counts vs. relative fluorescence (propidium iodide, PI) intensity]. The shaded area shows the fluorescence peak derived from nuclei preparations of *Zea mays* (cultivar CE-777, 2C = 5.43 pg) when measured alone. The white area shows the fluorescence peaks derived from the simultaneous analysis of nuclei preparations from the leaves of *Ilex* spp. and leaves of maize (as the internal standard), showing the impact that the cytosolic leaf components of *Ilex* has on the fluorescence of the standard. (a) Overlay histogram of *Ilex paraguariensis* (YM sn) plus the internal standard (maize), and that of maize alone (shaded area). (b) Overlay histogram of *Ilex dumosa* (YS sn) plus the standard, and that of maize alone (shaded area).

ubiquitous (Lloréns *et al*., 2008, 2011); yet, this is the first report of retroelements isolated from *I. dumosa*. As in many plant species, it seems plausible that a vast proportion of the genomes of 'yerba mate' and 'yerba señorita' are represented by repetitive DNA. The pervasiveness of the repetitive DNA, a lack of sufficient driver competition during the hybridization step, a linear increase in contaminating heteroduplexes, an inefficient annealing in the complex mixture after the subtraction step, and/or a trapping of a fraction of the tester molecules in heteroduplexes (Kuvachieva and Goffinet, 2002) would explain the absence of species-specific fragments among the products isolated. Regarding the sequences identified as organellar or ribosomal DNA (group C), albeit they represent a minor fraction of the isolated products (19–31%), they are non-differential products that could have been generated due to the ubiquity of organellar DNA in preparations from leaf tissue (Gottlieb and Poggio, 2010; Shi *et al*., 2012).

In addition, we detected protein domains from enzymes involved in plants' primary and secondary metabolism and signal transduction and regulation (Grabarek, 2006; Hammargren *et al*., 2007; Tonfack *et al*., 2011; Bienert *et al*., 2012; Moummou *et al*., 2012). Particularly, two fragments show similarity to a flavonoid-reductase involved in lignin biosynthesis, and to a serine carboxypeptidase that may contribute to defensive chemistry towards insects (Mugford *et al*., 2009), as was observed in wild *Solanaceae* (Li and Steffens, 2000). These sequences deserve a close examination, for instance by comparing their expression levels and investigating any relationship with psilid resistance.

It is worth remarking that *Ilex* spp. studied here are not closely related, nor are their sister taxa. Their phylogenetic relationship has not been clearly established yet because *I. paraguariensis* shows an unstable phylogenetic position (Gottlieb *et al*., 2005). Moreover, *I. dumosa* shows either a relationship with the Asian species *Ilex crenata* and *Ilex mutchagara* (Gottlieb *et al*., 2005) or with *Ilex argentina* (Manen *et al*., 2002), but not with 'yerba mate'. Our hybridization outcomes suggest that the genomes of both species have less than 80% similarity. This value is a little lower than the value reported previously, on the basis of amplified fragment length polymorphism analysis, which points to an approximately 23% difference (Gottlieb *et al*., 2011). That study has also shown that, in comparison with other southern South American *Ilex* spp. (i.e. *I. argentina*, *Ilex brasiliensis*, *Ilex brevicuspis*, *Ilex theezans* and *Ilex integerrima*), *I. paraguariensis* shows a range of interspecific genetic differences of 22–24.5%, whereas *I. dumosa* shows a range of higher interspecific differences (24–27%) (Gottlieb *et al*., 2005). In addition, the genetic distance between the two species duplicates and quadruplicates their respective intraspecific value (Gottlieb *et al*., 2005). Thus, the overall genetic difference between *I. paraguariensis* and *I. dumosa* obtained here is concurrent with previous results. Additionally, we detected contrasting performances through filter hybridizations that could be explained by the occurrence of a class (or classes) of homologous but slightly divergent repetitive DNA, highly amplified in the 'yerba mate' genome but not in 'yerba señorita'. In support of our proposal, the repetitive sequences detected in *I. paraguariensis* outnumber

(by up to 40-fold) those found in *I. dumosa*. These repetitive DNA sequences could include any of the telomeric or centromeric sequences, long or short repeats, micro- or minisatellites, satellite DNA, ribosomal tandem repeats, and transposons or retrotransposons (Sharma and Raina, 2005). This deserves a careful verification through physical mapping.

As mentioned previously, the leaves of 'yerba mate' have abundant caffeine and chlorogenic acids (Filip *et al*., 1999, 2001; Reginatto *et al*., 1999). The effect of cytosolic chlorogenic acids on flow cytometry fluorescence was demonstrated by Noirot *et al*. (2003, 2005) in *Coffea*. The presence of a complex mixture of secondary metabolites in cytosolic preparations of *I. paraguariensis* could account for the failure in measuring its nuclear content under conventional procedures. By using a highly reducing isolation buffer and short incubation times, we were able to estimate the genome size via flow cytometry for both species, for the first time. For *I. paraguariensis*, it has been shown that the estimate value results in 1.33 times lower than the one previously inferred by microdensitometry (2C = 2200 Mbp; Barral *et al*., 1995). Notwithstanding, the value obtained herein represents a more precise estimate, being an average of 24,119 nuclei, compared with the 384 nuclei measured by Barral *et al*. (1995). In *I. dumosa*, we observed a variable performance that may be due to an intraspecific variation in the abundance of cytosolic compounds; a fact not previously reported for this species. The genome size reported herein for *I. dumosa* constitutes a novel contribution. The difference in the DNA content between *I. paraguariensis* and *I. dumosa* (approximately 10%) does not appear to be related to the chromosome number since both species have $2n = 40$ chromosomes (Barral *et al*., 1995; Greizerstein *et al*., 2004); a correlation with other cytogenetic parameters remains to be investigated.

The genetic information gathered herein contributes to the scientific knowledge of the two species considered, offers a basis for developing suitable probes for physical mapping and genome organization studies, and certainly will facilitate the design of future genomic projects.

Supplementary material

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S1479262114000756

Acknowledgements

The authors are indebted to S. D. Prat Kricun, L. D. Belingheri and R. M. Mayol from EE INTA Cerro Azul (Misiones, Argentina), to the Establecimiento Las Marías (Corrientes, Argentina) and to P. A. Sansberro from IBONE-UNNE (Corrientes, Argentina), who kindly provided specimens from living collections. We are also grateful to G. C. Giberti who identified the plants and to M. S. Rodriguero and anonymous referees who helped improve the present and earlier versions of the manuscript. The authors gratefully acknowledge the 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).

References

Argout X, Salse J, Aury JM, Guiltinan MJ, Droc G, Gouzy J, Allegre M, Chaparro C, Legavre T, Maximova SN, Abrouk M, Murat F, Fouet O, Poulain J, Ruiz M, Roguet Y, Rodier-Goud M, Barbosa-Neto JF, Sabot F, Kudrna D, Ammiraju JS, Schuster SC, Carlson JE, Sallet E, Schiex T, Dievart A, Kramer M, Gelley L, Shi Z, Bérard A, Viot C, Boccara M, Risterucci AM, Guignon V, Sabau X, Axtell MJ, Ma Z, Zhang Y, Brown S, Bourge M, Golser W, Song X, Clement D, Rivallan R, Tahi M, Akaza JM, Pitollat B, Gramacho K, D'Hont A, Brunel D, Infante D, Kebe I, Costet P, Wing R, McCombie WR, Guiderdoni E, Quetier F, Panaud O, Wincker P, Bocs S and Lanaud C (2011) The genome of *Theobroma cacao*. *Nature Genetics* 43: 101–108.

Barral G, Poggio L and Giberti GC (1995) Chromosome numbers and DNA content from *Ilex argentina* (Aquifoliaceae). *Boletín de la Sociedad Argentina de Botánica* 30: 243–248.

Belingheri LD, Prat Kricum SD (2000) Programa de mejoramiento genético de la yerba mate en el INTA. II Congresso Sul-americano da Erva Mate. Encantado-Brasil. *Actas* pp 60–63.

Berté KA, Beux MR, Spada PK, Salvador M and Hoffmann-Ribani R (2011) Chemical composition and antioxidant activity of yerba-mate (*Ilex paraguariensis* A. St.-Hil., Aquifoliaceae) extract as obtained by spray drying. *Journal of Agricultural and Food Chemistry* 59: 5523–5527.

Bienert MD, Delannoy M, Navarre C and Boutry M (2012) NtSCP1 from tobacco is an extracellular serine carboxypeptidase III that has an impact on cell elongation. *Plant Physiology* 158: 1220–1229.

Bracesco N, Sanchez AG, Contreras V, Menini T and Gugliucci A (2011) Recent advances on *Ilex paraguariensis* research: mini review. *Journal of Ethnopharmacology* 136: 378–384.

Cascales J (2013) Caracterización microsatélite en poblaciones silvestres de yerba mate (*Ilex paraguariensis*). MSc Thesis, Universidad de Buenos Aires, Argentina, 61 pp.

Chen ZJ, Phillips RL and Rines HW (1998) Maize DNA enrichment by representational difference analysis in a maize chromosome addition line of oat. *Theoretical and Applied Genetics* 97: 337–344.

Cossu RM, Buti M, Giodani T, Natali L and Cavallini A (2012) A computational study of the dynamics of LTR retrotransposons in the *Populus trichocarpa* genome. *Tree Genetics and Genomes* 8: 61–75.

Doležel J, Binarová P and Lucretti S (1989) Analysis of nuclear DNA content in plant cells by flow cytometry. *Biologia Plantarum* 31: 113–120.

Doležel J, Bartoš J, Voglmayr H and Greilhuber J (2003) Nuclear DNA content and genome size of trout and human. *Cytometry* 51: 127–128.

Filip R, López P, Coussio J and Ferraro G (1998) Mate substitutes or adulterants: study of xanthine content. *Phytotherapy Research* 12: 129–131.

Filip R, López P, Coussio J and Ferraro G (1999) Phytochemical study of *Ilex dumosa*. *Acta Horticulturae* 501: 333–335.

Filip R, López P, Giberti GC, Coussio J and Ferraro G (2001) Phenolic compounds in seven South American *Ilex* species. *Fitoterapia* 72: 774–778.

Gauer L and 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 (1989) Los parientes silvestres de la yerba mate y el problema de su adulteración. *Dominguezia* 7: 3–22.

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

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

Gottlieb AM and 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, Giberti GC and Poggio L (2005) Molecular analyses of the genus *Ilex* (Aquifoliaceae) in southern South America, evidence from AFLP and ITS sequence data. *American Journal of Botany* 92: 352–369.

Gottlieb AM, Giberti GC and Poggio L (2011) Evaluación del germoplasma de *Ilex paraguariensis* e *Ilex dumosa* (Aquifoliaceae). *Boletín de la Sociedad Argentina de Botánica* 46: 113–123.

Grabarek Z (2006) Structural basis for diversity of the EF-hand calcium-binding proteins. *Journal of Molecular Biology* 359: 509–525.

Greizerstein EJ, Giberti GC and Poggio L (2004) Cytogenetic studies of southern South American *Ilex* species. *Caryologia* 57: 19–23.

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

Hammargren J, Sundström J, Johansson M, Bergman P and Knorpp C (2007) On the phylogeny, expression and targeting of plant nucleoside diphosphate kinases. *Physiologia Plantarum* 129: 79–89.

Heck CI and de Mejia EG (2007) Yerba Mate Tea (*Ilex paraguariensis*): a comprehensive review on chemistry, health implications, and technological considerations. *Journal of Food Sciences* 72: R138–R151.

Kuvachieva AA and Goffinet AM (2002) A modification of Representational Difference Analysis, with application to the cloning of a candidate in the Reelin signalling pathway. *BMC Molecular Biology* 3: 6. Available at http://www.biomedcentral.com/1471-2199/3/6

Li AX and Steffens JC (2000) An acyltransferase catalyzing the formation of diacylglucose is a serine carboxypeptidase-like protein. *Proceedings of the National Academy of Sciences of the USA* 97: 6209–6907.

Lin J, Kudrna D and Wing RA (2011) Construction, characterization, and preliminary BAC-end sequence analysis of a bacterial artificial chromosome library of the tea plant (*Camellia sinensis*). *Journal of Biomedicine and Biotechnology* 2: 476723–476731.

Linares Miquel C (1997) Análisis del genomio A del género *Avena* mediante la variación en secuencias repetidas de ADN. PhD Thesis, Universidad de Alcalá de Henares, Madrid, Spain, 139 pp.

Lisitsyn N, Lisitsyn NM and Wigler M (1993) Cloning the differences between two complex genomes. *Science* 259: 946–951.

Lloréns C, Futami R, Bezemer D and Moya A (2008) The Gypsy Database (GyDB) of mobile genetic elements. *Nucleic Acids Research* 36: 38–46.

Lloréns C, Futami R, Covelli L, Dominguez-Escribá L, Viu JM, Tamarit D, Aguilar-Rodríguez J, Vicente-Ripolles M, Fuster G, Bernet GP, Maumus F, Munoz-Pomer A, Sempere JM, Latorre A and Moya A (2011) The Gypsy Database (GyDB) of mobile genetic elements: release 2.0. *Nucleic Acids Research (NARESE)* 39: D70–D74.

Lysák MA and Doležel J (1998) Estimation of nuclear DNA content in *Sesleria* (Poaceae). *Caryologia* 52: 123–132.

Manen J-F, Boulter MC and Naciri-Graven Y (2002) The complex history of the genus *Ilex* L. (Aquifoliaceae): evidence from the comparison of plastid and nuclear DNA sequences and from fossil data. *Plant Systematics and Evolution* 235: 79–98.

Marcon H, Domingues D and Marino C (2011) Identification of potential transcriptionally active Copia LTR retrotransposons in *Eucalyptus*. *BMC Proceedings* 5: P164.

Moummou H, Kallberg Y, Tonfack LB, Persson B and van der Rest B (2012) The plant short-chain dehydrogenase (SDR) superfamily: genome-wide inventory and diversification patterns. *BMC Plant Biolology* 12: 219.

Mugford ST, Qi X, Bakht S, Hill L, Wegel E, Hughes RK, Papadopoulou K, Melton R, Philo M, Sainsbury F, Lomonossoff GP, Roy AD, Goss RJ and Osbourn A (2009) A serine carboxypeptidase-like acyltransferase is required for synthesis of antimicrobial compounds and disease resistance in oats. *Plant Cell* 21: 2473–2484.

Nagarajan N, Navajas-Pérez R, Pop M, Alam M, Ming R, Paterson AH and Salzberg SL (2008) Genome-wide analysis of repetitive elements in papaya. *Tropical Plant Biology* 1: 191–201.

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

Noirot M, Barre P, Duperray C, Louarn J and Hamon S (2003) Effects of caffeine and chlorogenic acid on propidium iodide accessibility to DNA: consequences on genome size evaluation in coffee tree. *Annals of Botany* 92: 259–264.

Noirot M, Barre P, Duperray C, Hamon S and De Kochko A (2005) Investigation on the causes of stoichiometric error in genome size estimation using heat experiments: consequences on data interpretation. *Annals of Botany* 95: 111–118.

Orita M, Iwahana H, Kanazawa H, Hayashi K and Sekiya T (1989) Detection of polymorphisms of human DNA by gel electrophoresis as Single-Strand Conformation Polymorphisms. *Proceedings of the National Academy of Sciences USA* 86: 2766–2770.

Panaud O, Vitte C, Hivert J, Muzlak S, Talag J, Brar D and Sarr A (2002) Characterization of transposable elements in the

genome of rice (*Oryza sativa* L.) using Representational Difference Analysis (RDA). *Molecular Genetics and Genomes* 268: 113–123.

Pereira MF, Ciampi AY, Inglis PW, Souza VA and Azevedo VCR (2013) Shotgun sequencing for microsatellite identification in *Ilex paraguariensis* (Aquifoliaceae). *Applications in Plant Sciences* 1: 1200245.

Prat Kricun S (2009) El registro nacional de propiedad de cultivares de la SAGPYA inscribió dos nuevas variedades de *Ilex dumosa* obtenidas por el INTA Cerro Azul – Misiones. Available at http://www.ife.gov.ar/articulo/716

Reginatto FH, Athayde ML, Gosmann G and Schenkel EP (1999) Methylxanthines accumulation in Ilex species-caffeine and theobromine in erva-mate (*Ilex paraguariensis*) and other Ilex species. *Journal of Brazilian Chemistry Society* 10: 443–446.

Sabot F, Sourdille P and Bernard M (2004) Detecting specific repeated sequences in large, complex genomes by using representative difference analysis and double-probe verification. *Plant Molecular Biology Reporter* 22: 91a–91j.

Sharma S and Raina SN (2005) Organization and evolution of highly repeated satellite DNA sequences in plant chromosomes. *Cytogenetics and Genome Research* 109: 15–26.

Shi C, Hu N, Huang H, Gao J, Zhao Y-J and Gao L-Z (2012) An improved chloroplast DNA extraction procedure for whole plastid genome sequencing. *PLoS ONE* 7: e31468.

Smit AFA, Hubley R and Green P (1996–2010) *RepeatMasker Open-3.0*. Available at http://www.repeatmasker.org

Tonfack LB, Moummou H, Latché A, Youmbi E, Benichou M, Pech JC and Van der Rest B (2011) The plant SDR superfamily: involvement in primary and secondary metabolism. *Current Topics in Plant Biology* 12: 41–53.

Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Déjardin A, Depamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjärvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leplé JC, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouzé P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai CJ, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, Van de Peer Y and Rokhsar D (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. and Gray). *Science* 313: 1596–1604.

Zoldos V, Siljak-Yakovlev S, Papes D, Sarr A and Panaud O (2001) Representational difference analysis reveals genomic differences between *Q. robur* and *Q. suber*: implications for the study of genome evolution in the genus *Quercus*. *Molecular Genetics and Genomes* 265: 234–241.