Journal of Plant Research
https://doi.org/10.1007/s10265-021-01279-3

REGULAR PAPER – GENETICS/DEVELOPMENTAL BIOLOGY

Differential DNA methylation and gene expression during development of reproductive and vegetative organs in *Ilex* species

Jimena Cascales[1,4] · Raúl Maximiliano Acevedo[2,4] · Daniela Ivana Paiva[3,4] · Alexandra Marina Gottlieb[1,4]

Received: 13 January 2021 / Accepted: 5 March 2021
© The Botanical Society of Japan 2021

Abstract
Differential epigenetic (DNA cytosine methylation) and gene expression patterns were investigated in reproductive and vegetative organs from *Ilex paraguariensis* and *I. dumosa*, at distinct developmental stages. We aimed at contributing towards elucidating major molecular changes underlying the sexual differentiation processes which, in these dioecious species, are completely unknown. Simultaneously, as a first step towards the development of an early sexing system, we searched for promising molecular markers. This was assessed through Methylation Sensitive Amplified Polymorphism (MSAP) and Amplified Fragment Length Polymorphism on cDNA (cDNA-AFLP) techniques, applying discriminant multivariate analyses, and bioinformatic characterization of differential fragments. A significant positive correlation was found between epigenetic and indirect 'genetic' information for both species, indicating influence of the genetic background on the epigenetic variation. Higher epigenetic than genetic diversities were estimated. Our outcomes showed up to 1.86 times more representation of mCG subepiloci than mCCG in all organs sampled. Along the maturing stages of floral buds, the frequency of mCG evidenced an incremental trend, whereas mCCG and unmethylated conditions showed opposite tendencies. Reproductive and vegetative samples tended to cluster apart based on epigenetic patterns; at gene expression level, organs exhibited clear-cut distinctive patterns, nonetheless profiles of young leaves and floral primordia resemble. Epigenetic and expression data allowed discrimination of *I. dumosa´*s samples according to the gender of the donor; more elusive patterns were observed for *I. paraguariensis*. In total, 102 differentially methylated and expressed fragments were characterized bioinformatically. Forty-three were annotated in various functional categories; four candidate markers were validated through qPCR, finding statistical differences among organs but not among sexes. The methylation condition of epilocus C13m33 appears as indicative of gender in both species. Thirty-three organ-specific and 34 gender-specific methylated markers were discriminated and deserve further research, particularly those expressed in leaves. Our study contributes concrete candidate markers with potential for practical application.

Keywords cDNA-AFLP · Dioecy · DNA methylation · Molecular markers · MSAP · Multivariate analyses

✉ Alexandra Marina Gottlieb
gottlieb@ege.fcen.uba.ar

1 Laboratorio de Citogenética y Evolución, Departamento de Ecología, Genética y Evolución, Facultad de Ciencias Exactas y Naturales, Instituto de Ecología, Genética y Evolución de Buenos Aires (IEGEBA, CONICET-UBA), Universidad de Buenos Aires, Intendente Güiraldes 2160, Pabellón II, Ciudad Universitaria, C1428EHA Ciudad Autónoma de Buenos Aires, Argentina

2 Laboratorio de Biotecnología Aplicada y Genómica Funcional, Facultad de Ciencias Agrarias, Instituto de Botánica del Nordeste (IBONE, UNNE-CONICET), Universidad Nacional del Nordeste, Sargento Juan Bautista Cabral 2131, Corrientes W3402BKG, Argentina

3 Instituto Nacional de Tecnología Agropecuaria, Estación Experimental Agropecuaria Montecarlo (INTA EEA Montecarlo), Av. El Libertador 2472, Misiones N3384, Argentina

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

Published online: 23 March 2021

Introduction

Ilex paraguariensis A. St.-Hil. and *I. dumosa* Reissek are perennial dioecious trees native to subtropical southern South America, with partially overlapping distribution areas (Giberti 1979, 1989, 2001, 2011; Grela 2004). Even though their phylogenetic relationship has been disputed (Cascales et al. 2017; Gottlieb et al. 2005; Manen et al. 2010), their divergence was estimated in around 28 million years ago (Yao et al. 2020). The leaves and twigs of *I. paraguariensis* ('yerba mate') are industrialized to prepare the 'mate', an infusion with nutraceutical properties conferred by high concentrations of secondary metabolites (Bracesco et al. 2011; Cardozo Junior and Morand 2016; Filip 2009; Souza et al. 2015). The leaves of the 'yerba señorita' tree (*I. dumosa*) contain significantly lower amounts of metabolites (Filip et al. 1998, 2001), and were used to adulterate the 'yerba mate' until late 1980s (Giberti 2011). Also, *I. dumosa* shows greater tolerance to drought and low temperatures, and resistance to the *I. paraguariensis*' psilid plague (Prat Kricun 2009). Due to these features, *I. dumosa* represents a source of valuable alleles and regulatory regions for the improvement of the 'yerba mate' crop (Gottlieb and Poggio 2014; Gottlieb et al. 2011).

In *Ilex* species, the flowers start developing bisexually and end as functionally unisexual organs composed by a fertile cycle and rudiments of the other gender (Giberti 1994a, b; Giberti and Gurni 2008), that is, they are type I diclinous flowers (Diggle et al. 2011). The molecular mechanisms triggering development of the gynoecium or of the androecium, and the arrestment of the counter cycle, are utterly unknown in this genus. As well, the genetic bases of the operating system are unexplored.

Within the small percentage of dioecious sexual systems that have been elucidated in angiosperms, a broad spectrum of mechanisms has been described (Charlesworth 2016; Filatov 2015; Henry et al. 2018; Kersten et al. 2017; Yang et al. 2019). Hormonal and/or epigenetic controllers were found to have important roles in various genera (Akagi et al. 2016; Bräutigam et al. 2017; Khadka et al. 2019); in particular, sex-specific methylation appeared independently multiple times in the evolution of dioecy (Bräutigam and Cronk 2018). As to *I. paraguariensis* and *I. dumosa*, there is no evidence of sexual chromosomes (Greizerstein et al. 2004) or sex-linked genomic regions (Gottlieb and Poggio 2010). Moreover, these species cannot be sexed before anthesis, which can take 3–10 years since germination (Giberti 2001).

Epigenetic marks can alter gene expression (Bartels et al. 2018; Heikrujam et al. 2015; Piferrer 2013; Song et al. 2015), regulate leaf morphology, flowering time, floral organ identity, fertility and embryogenesis, among other processes (Bartels et al. 2018). The global content of DNA methylation in angiosperms is thought to vary across species, organs, and developmental stages (Bartels et al. 2018; Vanyushin and Ashapkin 2011). Knowledge on epigenetics and gene expression is still scarce in *Ilex*. Investigations on epigenetics are limited to natural populations of the common holly, *I. aquifolium* L. (Herrera and Bazaga 2013), for which epigenetic variation could be a major source of organ-level phenotypic variance. Recent transcriptomic studies on *I. paraguariensis* have focused on elucidating abiotic stress responses, depicting global female gene expression panels, and characterizing the phenylpropanoid pathway (Acevedo et al. 2013, 2016, 2019; Debat et al. 2014; Fay et al. 2018). Gene expression studies on other *Ilex* species, *I. asprella* (Zheng et al. 2014, 2015) and *I. pubescens* (Wen et al. 2017), characterized the triterpenoid saponin biosynthetic pathway.

The phenotypic similarities shown by *I. paraguariensis* and *I. dumosa* flowers (Giberti 1994a, b; 2001; Giberti and Gurni 2008) led us to assume a common sexual system. Also, we postulate that sexual differentiation in these species could be based on differential gene expression, including sex-specific epigenetic regulation. In this context, the following questions guided our study: how differentiated are these species at their epigenetic and gene expression profiles?, are there organ-specific patterns? is the epigenetic control over gene expression involved in sexual differentiation?, which differentially methylated and/or expressed genes maximize distinction among organs and sexes?, in which pathways do they intervene? Thus, we analyzed flowers at various developmental stages and, also, in comparison with vegetative organs, aiming at detecting major distinctive epigenetic (DNA cytosine methylation, mC) and gene expression patterns, that could aid in elucidating the subjacent mechanisms. In addition, we expected to reveal gender-associated markers that could be examined from leaves, as a first step towards the development of a sexing procedure. Such a method would allow 'yerba mate' producers to manage fruit producing female plants more efficiently.

We implemented Methylation Sensitive Amplified Polymorphism (MSAP; Reyna-Lopez et al. 1997) and cDNA-AFLP (Bachem et al. 1996) techniques to answer the above questions. MSAP has been successfully employed to isolate sex-specific fragments in *Populus tomentosa* (Song et al. 2012); cDNA-AFLP allowed investigation of a transcript population involved in sexual differentiation in *Cannabis sativa* (Moliterni et al. 2004), and the differential gene expression analysis among reproductive and vegetative organs of *Angelica sinensis* (Yu et al. 2011). These techniques reduce the complexity of both the genome and the transcriptome by selective amplification of a subset of enzymatically restricted fragments, allowing the concurrent screening of multiple anonymous dominant markers.

As well, they are cost-effective and require minimal start-up. Thus, they facilitate the investigation of the methylomes and transcriptomes of non-model species (Schrey et al. 2013).

Materials and methods

Materials

Floral buds, flowers and leaves were sampled from plants held at the "Jardín Botánico Carlos Thays" (Ciudad Autónoma de Buenos Aires, Argentina; *I. paraguariensis*, female accessions: JB1839, JB1845, JB1848, JB1849, JB1851, JB1852; male accessions: JB1835, JB1837, JB1841, JB1842, JB1856; *I. dumosa*, female accession: JB1091). All plants had the same age, as plantlets -donated by the "Establecimiento Las Marías" (Corrientes, Argentina)- were implanted simultaneously in 2010, in a plot especially fitted (as specified by Panigatti 2010). We also used other materials from the "Establecimiento Las Marías" (*I. dumosa*, female specimens: CRV7/4, CRV7/12; male specimens: CRV48/10, 42/92/19) kindly provided by Dr. PA Sansberro (IBONE, UNNE-CONICET).

For MSAP, four to five distinct developmental stages of reproductive organs and two stages of vegetative organs from both species and sexes were analyzed (34 samples; Table S1). Floral primordia were categorized in utilitarian stages according to their sizes (bud-I, bulk of incipient non-pedicellate green buds; bud-II, pedicellate green buds < 1 mm in diameter; bud-III, greenish primordia of 1–2.5 mm in diameter). A fourth stage was established solely for male floral buds of *I. dumosa* (bud-IV, pre-anthesis whitish buds); as for this species female buds' samples were scanty, two reproductive stages were considered: floral primordia and flowers (Table S1). A single category was used to account for mature flowers. Newly emerged juvenile leaves, referred to as young, were less than 2 cm^2; adult leaves referred to completely expanded leaves from the former season (Table S1). As floral ontogeny is asynchronous (Giberti and Gurni 2008), sampled individuals were those that exhibited the floral primordia and leaves in comparable developmental stages at the collection dates.

The cDNA-AFLP analysis involved floral primordia in bud-III stage, and young and adult leaves from both species and sexes (28 samples; Table S2).

After collected, samples were immediately submerged in liquid nitrogen or in RNAlater (Thermo Fisher Scientific), and then stored at − 80 °C until processing.

DNA extraction

Total genomic DNA was isolated from ca. 100 mg of each sample (bulks of floral primordia at the same stage, flowers, or leaves), using the DNeasy Plant Quick Extraction kit (QIAGEN Inc., Duesseldorf, Germany) following manufacturer's instructions.

RNA extraction and cDNA synthesis

Total RNA was extracted from ca. 100 mg of floral buds or leaves, with Trizol Reagent (Thermo Fisher Scientific, Waltham, Massachusetts, USA) following manufacturer's instructions, and re-purified using either RNeasy Plant Mini Kit (QIAGEN) or EasyPure Plant RNA kit (Transgen Biotech, Beijing, China). After verification of RNA integrity by agarose gel electrophoresis (1 % w/v), RNA was treated with 4 U of DNase I (Promega, Madison, Wisconsin, USA). Simple-stranded cDNA was synthesized using the SuperScript III First-Strand Synthesis System (Thermo Fisher Scientific) and oligo dT_{20} as the primer, following the manufacturer's protocol. Then, double-stranded cDNA was obtained in a 150 µl-reaction containing 1X Second Strand Synthesis Reaction Buffer (New England Biolabs, NEB, Hitchin, Hertfordshire, UK), 2 mM dNTP mix, 10 U of *E. coli* ligase (NEB), 40 U of *E. coli* DNA polymerase I (NEB), 2 U of *E. coli* RNase H (NEB) and DEPC-treated water. The mix was incubated for 2 h at 16 ºC. Double-stranded cDNA was purified with one volume of phenol:chloroform:isoamyl alcohol (25:24:1), precipitated with NH_4OAc 7.5 M and resuspended in 10 mM Tris-HCl.

MSAP procedure

The technique was carried out employing the restriction enzyme *EcoRI* as the rare cutter, in combination with either of the frequent cutters *HpaII* or *MspI*. These isoschizomers recognize and cut the 5'-CCGG-3' double stranded motif when unmethylated but are sensitive to the methylation state of the cytosines (McClelland et al. 1994), thus allowing inference of methylation status of target sequences (Pérez-Figueroa 2013). *HpaII* also cuts externally hemimethylated targets (abbreviated as mCCG), and *MspI* also cuts when the internal cytosine is fully or hemimethylated (abbreviated as mCG). Both enzymes are unable to cut when the target is hypermethylated, that is, when both strands show external methylation or any other combination of methylated cytosines (Guevara et al. 2017; Pérez-Figueroa 2013; Schulz et al. 2013).

DNA aliquots (500 ng) were digested with 5 U of *EcoRI* (Thermo Fisher Scientific) in a 30-µl reaction for 2 h at 37 °C. After a chloroform:isoamyl alcohol extraction, half volume was treated with 2.5 U of *MspI* (NEB) and the other half with 2.5 U of *HpaII* (NEB), in a final volume of 30 µl. Digestions were incubated for 1.5 h at 37 °C. For oligonucleotide adapter ligation, 10 µl of a mix containing 0.5 U of T4 DNA ligase (Thermo Fisher Scientific), 1X ligase

buffer, 5 µM of *EcoRI* adapter and 50 µM of *MspI/HpaII* adapter were added to 15 µl of each digested sample. Ligation reactions were incubated for 2 hs at 37 ºC and kept at 4 ºC overnight. Adapter and primer sequences are listed in Table S3. Ligated DNA was diluted 100-fold, and 5 µl aliquots were used for 50-µl pre-selective PCR amplifications using 90 ng of each pre-selective primer (Table S3), 1X PCR buffer, 1.5 mM $MgCl_2$, 0.2 mM of each dNTP and 1 U of *Taq* DNA polymerase (Thermo Fisher Scientific). Cycling consisted of 20 cycles of 30 s at 94 ºC, 1 min at 56 ºC and 1 min at 72 ºC. The products were diluted fivefold and 5 µl were selectively amplified in 20-µl reactions containing 1X PCR buffer, 1.5 mM $MgCl_2$, 0.25 mM of each dNTP, 36 ng of each selective primer (Tables S3–S5) and 0.75 U of *Taq* DNA polymerase (Thermo Fisher Scientific). Thermal cycling conditions were: 13 cycles of 30 s at 94 ºC, 30 s at 65 ºC (with a touchdown of 0.7 ºC) and 30 s at 72 ºC, following 23 cycles of 30 s at 94 ºC, 30 s at 56 ºC and 1 min at 72 ºC, and a final extension of 10 min at 72 ºC.

The resulting fragments were separated through 5 % (w/v) high resolution denaturing PAGE, as in Cascales et al. (2014). Bands were visualized using the Silver Sequence DNA Sequencing System (Promega) following the manufacturer's protocol; dried gels were digitized for analysis. To gather reliable data from each PAGE gel, we took as a reference the intensity of the clearest band from the molecular weight ladder, measured through the histogram tool of Adobe Photoshop. Bands with an intensity over that threshold were excluded from the analysis (as the higher the value, the lower the intensity). In addition, in every PAGE run amplicons were included in replicates. When needed, internal standards of known patterns were also included to facilitate comparisons among gels. Banding patterns were blindly scored by the same person as binary data (presence [1] /absence [0]).

Thirty-seven selective primer combinations were explored in a subset of six representative samples; as a result, 13 combinations were selected for screening the entire set (Table S4); remaining 24 combinations yielded ≤ 35 total bands, showed repeated amplification failures or unreliable banding patterns, and were thus discarded (Table S5). Genotyping error rates per primer pair and species were estimated based on randomly chosen samples (13 %) run in duplicates; errors were computed following Herrera and Bazaga (2010).

cDNA-AFLP procedure

Purified double-stranded cDNA (300 ng) was doubly digested in a 30-µl reaction mix containing 3 U of *EcoRI-HF* (NEB), 3 U of *MseI* (NEB) and 1X CutSmart Buffer (NEB), for 2 h at 37 °C. Adapter ligation, amplification steps, PAGE and data recordings were performed as described above, except that *MseI*'s adapter and primers were used (Tables S3–S5). Thirty-one primer combinations were tested in a subset of six representative samples and 15 primer combinations (Table S4) were selected for analyzing the entire set. Sixteen combinations were discarded for yielding ≤ 25 bands or for evidencing inconsistent PCR amplifications (Table S5). Genotyping error rates for each primer pair and species were estimated, as in Bonin et al. (2004), from 30 % of randomly chosen samples run in duplicates.

MSAP data analyses

Two binary data matrices were manually constructed, one for *I. paraguariensis* (M_{YM}) and another for *I. dumosa* (M_{YS}). These were processed with the *msap* 1.1.8 package (Pérez-Figueroa 2013) of R 3.1.3 (R Core Team 2019), including the corresponding Error Rate-based Thresholds (ERT) in the command line, i.e., the estimated genotyping error rates per primer pair and species mentioned above. This software classifies each restriction site as either '**m**ethylation-**s**usceptible **l**ocus' (MSL) or '**n**on-**m**ethylated **l**ocus' (NML, that is, a site not susceptible of being methylated), depending on whether the proportion of methylated states across all samples exceeded the ERT specified for each primer combination. Thus, the program automatically splits and transforms the input matrix into the corresponding MSL and NML sub-matrices, excluding all monomorphic loci (Pérez-Figueroa 2013). To discriminate among methylation conditions or subepiloci (mCG, mCCG and unmethylated loci, the latter denoting a methylation-susceptible site but without any mCs), another transformation was performed on input matrices (M_{YM} and M_{YS}). After extracting the recognized methylation-susceptible loci from raw matrices, such transformation was accomplished under the Mixed Scoring 2 (MS2) approach with the R script *MSAP_calc.r* (Schulz et al. 2013). The two output matrices (i.e., $MS2_{YM}$ and $MS2_{YS}$), hereafter considered as epigenetic data, circumvent the inclusion of uninformative hypermethylated or mutated targets sites, which could conduct to confounding effects (Schulz et al. 2013). The NML matrices (NML_{YM} and NML_{YS}) were considered as 'genetic' data as negligible frequencies of methylated states were inferred for those loci by *msap* (Pérez-Figueroa 2013). Input matrices and scripts are presented in Online Resource 1.

Diversity indices were estimated from MS2 and NML matrices based on the Shannon's information index (I; Shannon and Weaver 1949) manually computed in a spreadsheet, and were calculated within each locus as $I = -P_i * \log_2(P_i)$ where P_i is the relative frequency of the presence or absence of the ith band, and then computing the average. Estimated values were compared through a Wilcoxon Rank Sum test in R. Additionally, correlations between matrices were investigated via paired Mantel tests (999 permutations) in GenAlex 6.501 (Peakall and

Smouse 2012). For this, distance matrices were constructed by calculating the complement of the Dice similarity coefficient (Sneath and Sokal 1973), using Famd 1.31 software (Schlüter and Harris 2006). Differentiation among organs, developmental stages and sexes were analyzed for each species, based on the frequency of the three methylation conditions. These frequencies and the Shannon diversity indices were calculated with the script *MSAP_calc.r*.

Sparse DIABLO (Data Integration Analysis for Biomarker discovery using Latent variable approaches for 'Omics studies) conduct a Sparse Partial Least Squares-Discriminant Analysis (sPLS-DA) which reduces matrix dimensions while capturing and highlighting the largest sources of variation from the data, selecting the most discriminant subset of variables (Rohart et al. 2017). Although this bioinformatic tool was developed to analyze 'omics datasets, it also allows the multivariate analysis of binary data. Thus, this tool was employed to investigate associations between the subepiloci (the variables) and the set of organs (buds, flowers, and leaves) or between the subepiloci and the sexes; also, it was used to identify the most discriminant subset of subepiloci. sPLS-DA analyses were conducted on $MS2_{YM}$ and $MS2_{YS}$ matrices using the R package *mixOmics* (Le Cao et al. 2016; Rohart et al. 2017). For each matrix, three data blocks, containing either unmethylated, mCG or mCCG subepiloci, were generated. Clustering patterns were analyzed through heatmaps using the *cimDiablo* function. Correlations between the most informative subepiloci and their 'expression level' (frequency, in our case) with the set of samples, were assessed using the *circosPlot* function of *mixOmics*. Data matrices, settings and scripts are presented in Online Resource 1.

cDNA-AFLP data analysis

A binary (presence [1] /absence [0]) data matrix was constructed for each species (C_{YM} for *I. paraguariensis* and C_{YS} for *I. dumosa*). These were used for computing the percentage of polymorphism in *msap* package by setting the argument *meth* = *F* which treats matrices as genetic data; Shannon diversity indices were computed as previously explained. In addition, a Partial Least Squares-Discriminant Analysis (PLS-DA) was performed in *mixOmics*; this is a supervised analysis that classifies samples in groups according to a hierarchical clustering based on fingerprinting similarity (Rohart et al. 2017). A heatmap was generated for each species' matrix using the *cim* function, with the same program. Matrices and scripts used are presented in Online Resource 1.

Characterization of differential fragments

Bands recognized as differentially methylated regions or differentially expressed genes (that is, those indicating that a biological group of samples have a particular methylation state or transcript) were excised from polyacrylamide gels, and PCR re-amplified as in Cascales et al. (2014). PCR products were purified and sequenced at Macrogen INC. (Seoul, Korea). Sequence identification was performed through BLAST searches against NCBI nucleotide and protein databases. Also, the Gypsy DataBase 2.0 was inspected for retroelements. In all searches, the first significant hit (E-value $\leq E^{-5}$) with the most complete annotation was registered. Additionally, to retrieve full-length transcripts, our differential fragment sequences were used as query in BLASTn searches performed against *I. paraguariensis* transcriptomes from Debat et al. (2014) and Fay et al. (2018) (held at the NCBI's Transcriptome Shotgun Assembly Sequence Database, TSA) and from Acevedo et al. (2019) (held at the repository "MATEANDO" available at) using a more stringent condition (E-value $\leq E^{-20}$). As these *I. paraguariensis* specific databases lack sequence annotation, best hits were used as query in new searches at NCBI to attempt putative identification. Sequences obtained herein were deposited in Genbank (accessions MT495507–MT495599).

Validation of differential fragments from *I. paraguariensis*

Quantitative PCR (qPCR) assays were implemented to validate a set of differentially expressed transcripts and to evaluate the expression of a differentially methylated region. For this, total RNA was newly obtained from fresh 'yerba mate' early and mature floral buds (bud-II and III, respectively) and young and adult leaves. Samples were taken from three female plants (JB1845, JB1849 and JB1851) and three male plants (JB1841, JB1842 and JB1856). After the Trizol and DNase I treatment, RNA was cleaned-up by a second extraction round with Trizol. Simple-stranded cDNA was obtained as explained before. Primer pairs used for qPCR (Table S6) were designed using Primer3Plus (Untergasser et al. 2012) and Primer BLAST (Ye et al. 2012). qPCR reactions were performed in a StepOne Plus apparatus (Thermo Fisher Scientific); the cycling program was set as recommended by the PowerUp SYBR Master Mix' manufacturer (Thermo Fisher Scientific). Each 15 µl reaction consisted of 5 ng of cDNA, 900 nM of each primer and 1X PowerUp SYBR Master Mix; all reactions were performed in triplicate, and the mean cycle threshold value (C_t) was calculated for each sample using the StepOne software 2.3 (Thermo Fisher Scientific). For each primer pair and reaction, amplification specificity was verified by performing a melt curve at a 60–95 °C ramp. Also, the efficiency was calculated for each primer

pair by measuring C_t values of a 1:5 serial dilution (range: 50–0.08 ng) of a representative pool of cDNAs. Reaction efficiencies ranged between 87 and 100%, thus validating the adequacy of the primer pairs designed herein. Expression stability was assessed for three reference genes following Acevedo et al. (2018) (i.e., *actin* –ACT–, *RNA polymerase associated protein rtf1 homolog* –RTF– and *elongation factor 1-alpha* –EF1a–). This was done by registering C_t values over all samples; as a result, EF1a was selected as the reference gene (data not shown). For inter-plate calibration, one sample (JB1841-bud-II) was chosen as calibrator and was loaded in every plate; thresholds were set to obtain the same average C_t for the calibrator. ΔC_t values were calculated for each sample following Livak and Schmittgen (2001) as $\bar{c}_{t\ \text{gene of interest}} - \bar{c}_{t\ \text{reference gene}}$, where 'gene of interest' refers to each target fragment; the lower the value of ΔC_t, the greater the expression, and vice versa. The usage of ΔC_t method was validated for each target fragment, by plotting ΔC_t vs. log of the input cDNA quantity (50–0.08 ng) in a spreadsheet. In all cases, the slope of the linear regression was < 0.1, indicating that an equal qPCR efficiency could be assumed for the targets and EF1a (Livak and Schmittgen 2001).

To estimate the main effects of gender, organ types and their interactions, a linear mixed model was selected and applied on ΔC_t values using the REstricted Maximum Likelihood (REML) method (Faraway 2016) as implemented in the R package *lmerTest* (Kuznetsova et al. 2017). To assess the correlation among ΔC_t values derived from samples taken from the same individual, a *plant* factor was included as a random effect. Model selection was performed by backward elimination of non-significant fixed effects, always retaining the *plant* factor. In all cases, the final model solely included *organ* as an explanatory variable of fixed effect. To evaluate differences between groups, post hoc comparisons were performed using the Tukey's test at 5% level of significance, with the R package *multcomp*. R script is presented in Online Resource 1.

Results

MSAP analyses

A total of 997 bands were recorded (1.51% missing data), of which 20.36% were shared by *I. paraguariensis* and *I. dumosa*. An average of 76.7 loci per primer combination was registered. Estimated genotyping error rates per primer combination and species ranged from 0 to 0.24 (Table S4). The raw data matrix for *I. paraguariensis* (M_{YM}) involved 1.23 times the number of loci registered for *I. dumosa* (M_{YS}). Both matrices had more loci non-susceptible to methylation -NML- than susceptible loci -MSL- (*I. paraguariensis* 2.43 times; *I. dumosa* 1.23 times); also, NML showed higher polymorphism (Table S7). Transformed matrices $MS2_{YM}$ and $MS2_{YS}$ had similar number of subepiloci (Table 1). On average, more subepiloci (> 1.18 times) and private subepiloci (> 1.36 times) were detected in floral buds of both species in comparison to flowers and leaves, whereas the lowest values were registered in flowers (Table 2).

Global Shannon indices revealed for both species that transformed MS2 matrices hold more diversity (1.26–1.36 times) than NML matrices and that this difference is statistically significant (Table 1). Mantel tests indicated the existence of a significant positive correlation between epigenetic and 'genetic' matrices ($P = 0.001$; Table 1).

Ilex paraguariensis' floral primordia showed the highest absolute frequencies of subepiloci and private subepiloci, and the highest diversity (Table 2). The frequency of methylated subepiloci (mCG + mCCG) exceeds around 1.65–2.44 times that of unmethylated sites (Table 2). Still, mCCG subepiloci were the least frequent. Noteworthy, the flowers of *I. paraguariensis* showed higher counts of mCG subepiloci, whereas actively growing organs (floral buds and young leaves) had similar higher frequencies of unmethylated and mCG subepiloci. Adult leaves showed most of the subepiloci unmethylated. The floral buds of *I. dumosa* also showed the highest total number of subepiloci, private subepiloci and diversity index. As well, in this species the methylated subepiloci were 1.23–1.80 times more frequent than the unmethylated, but when methylated states were distinguished, the unmethylated were the most abundant sites (Table 2).

Table 1 Dimensions and diversity indices for MS2-transformed epigenetic matrices and genetic (NML) matrices

	$MS2_{YM}$	NML_{YM}	$MS2_{YS}$	NML_{YS}
No. loci × No. samples	207 × 20	459 × 20	202 × 14	281 × 14
I ± SD	0.680 ± 0.258	0.501 ± 0.225	0.700 ± 0.223	0.557 ± 0.217
W (*P*-value)	64,685 ($3.53E^{-14}$)		36,628 ($3.25E^{-08}$)	
r (*P*-value)	0.631 (0.001)		0.582 (0.001)	

I Shannon's information index, *SD* standard deviation, *W* Wilcoxon Rank Sum test statistic, *r* Mantel correlation coefficient

Table 2 Descriptors and statistical estimates of MS2-transformed matrices

Matrix	Group	Samples	Subepiloci					Shannon diversity index
			N	Private	Frequency (in percentage)			
					unmet	mCG	mCCG	
$MS2_{YM}$	Floral buds (stages I–III)	8	162	22	36.42	37.65	25.93	0.590
	Flowers	4	110	4	29.09	40.91	30.00	0.418
	Young leaves	4	130	9	37.69	37.69	24.62	0.513
	Adult leaves	4	142	13	36.62	33.10	30.28	0.588
$MS2_{YS}$	Floral buds (stages I–IV)	6	145	25	44.83	30.34	24.83	0.566
	Flowers	2	72	2	44.44	36.11	19.44	0.267
	Young leaves	3	115	24	35.65	31.30	33.04	0.505
	Adult leaves	3	118	11	44.92	33.90	21.19	0.505

unmet unmethylated, *mCG* internally methylated, *mCCG* externally hemimethylated

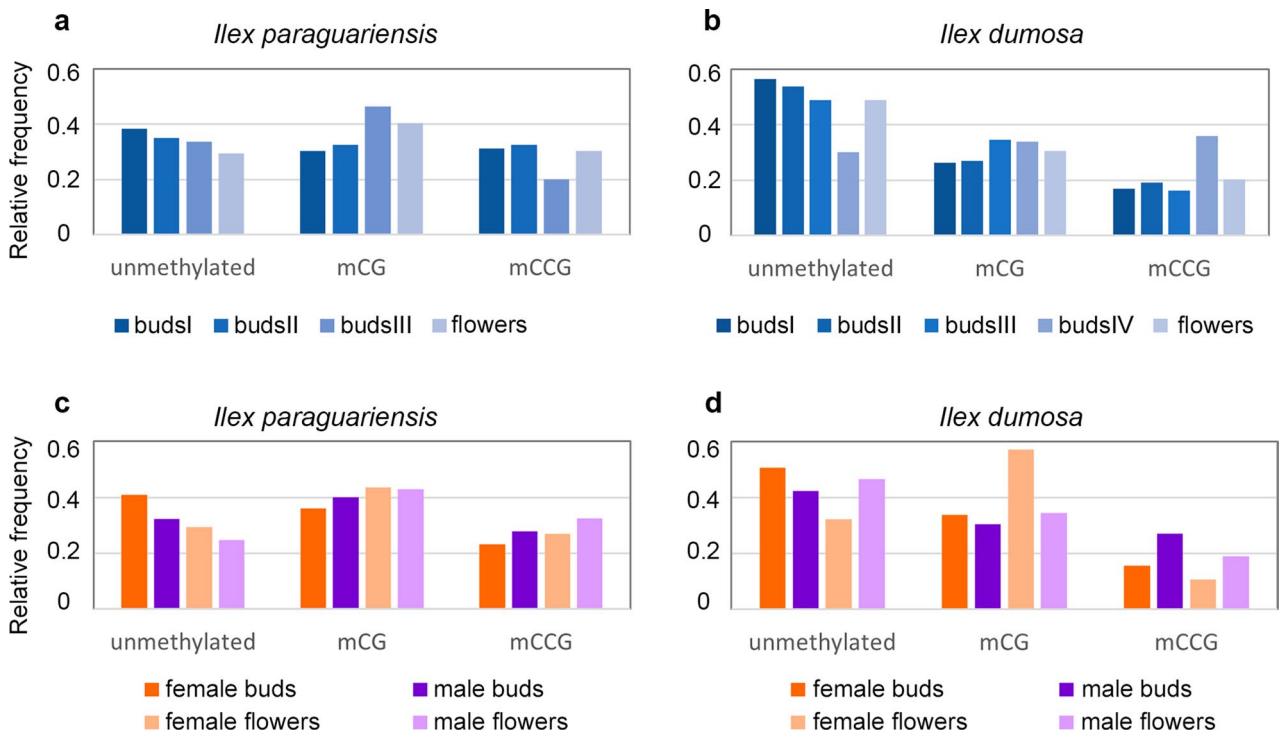

Fig. 1 Relative frequencies of subepiloci discriminating among reproductive stages (**a**, **b**) and between genders (**c**, **d**). Data derived from MS2-transformed matrices, $MS2_{YM}$ (**a**, **c**) and $MS2_{YS}$ (**b**, **d**)

A concurrent incremental tendency along the developmental stages of the floral buds was observed in the frequency of mCG methylation in both species, while the unmethylated and mCCG subepiloci showed opposite trends (Fig. 1a, b). When the gender was considered, female floral buds had higher frequency of unmethylated subepiloci than male floral buds, for both species (Fig. 1c, d). For *I. paraguariensis*, female floral buds differ from male buds in having lower mCG and mCCG frequencies, while female flowers had lower mCCG but similar mCG frequencies (Fig. 1c). For *I. dumosa*, female flowers had higher frequency of mCG sites (Fig. 1d).

When maximizing the distinction among *I. paraguariensis*' organs using matrix $MS2_{YM}$, 17 subepiloci were retained as the most discriminating variables or markers, among which all methylation conditions were equally represented. Opened flowers and floral buds I and II tended to differentiate, while the remaining biological samples appeared intermingled (Fig. S1a). The circos plot ($r \geq 0.7$), in addition, showed that floral buds could be discriminated

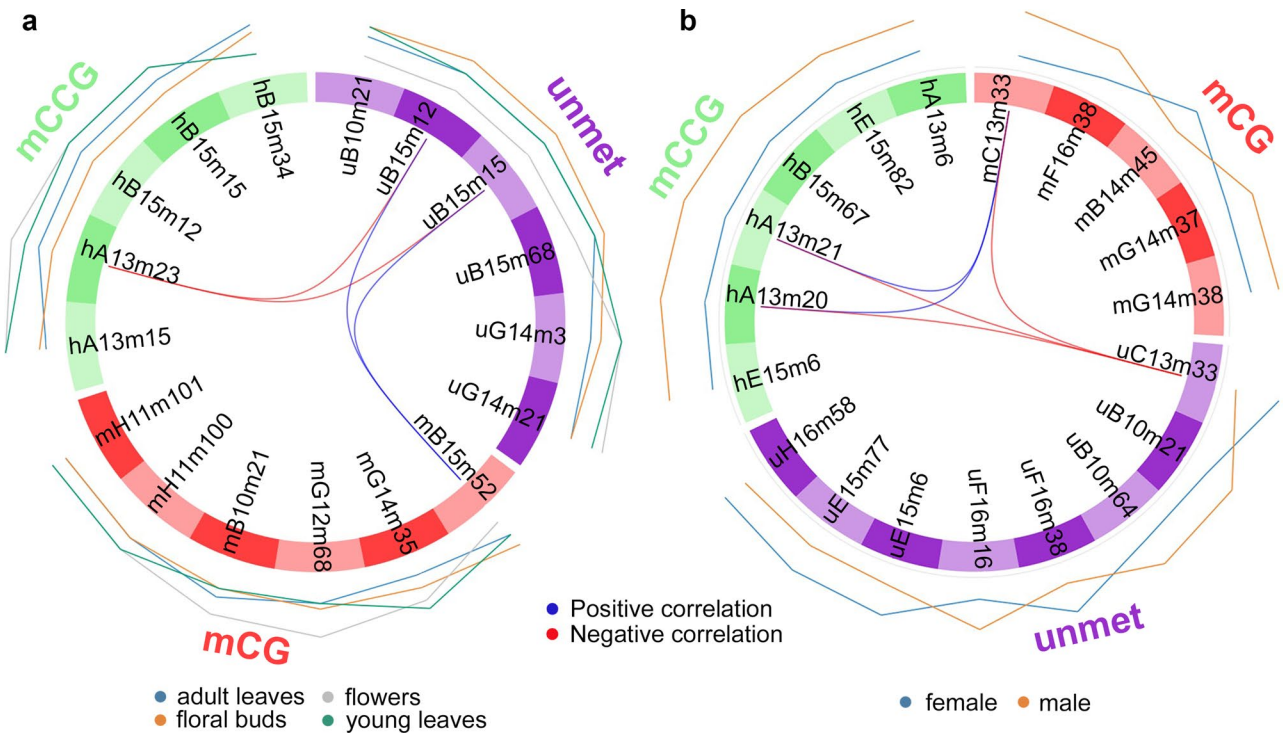

Fig. 2 Circos plot of *Ilex paraguariensis* epigenetic MS2 transformed data ($MS2_{YM}$ matrix). A single principal component was retained in each case. The variables are indicated in the middle colorful ring. Unmethylated variables (unmet) are indicated with violet shades, externally hemimethylated (mCCG) with green shades, and internally methylated (mCG) with red ones. Correlations among variables are indicated in the central zone by blue lines (when positive) and red lines (when negative). Threshold: $r = 0.7$. **a** Discriminating by organs. Outer lines depict the 'expression level' (frequency) of each subepilocus on each organ (light blue, adult leaf samples; green, young leaves; orange, floral bud samples; grey, flowers). **b** Discriminating by genders. Outer lines depict the 'expression level' of each subepilocus on each gender (female in light blue and male in orange)

by the simultaneous presence of three markers, a mCG site (mB15m52) and two unmethylated subepiloci (uB15m12 and uB15m15). In turn, these two were negatively correlated with a mCCG marker (hA13m23) that is highly represented in flowers (Fig. 2a). Differentiation between the genders was maximized by 19 subepiloci (58 % methylated). Three clusters could be recognized, one exclusively composed by female samples, other by male samples, and the third with a mixed composition (67 % males; Fig. S1b). The circos plot revealed that male samples were characterized by 68.42 % of the retained variables; positive correlations ($r \geq 0.7$) were evidenced between the mCG subepilocus mC13m33 and two mCCG subepiloci (hA13m20 and hA13m21). Epilocus C13m33 appeared internally methylated in male samples but unmethylated in females, and these states were negatively correlated (Fig. 2b).

Differences among organ types taken from *I. dumosa* were explained by 16 subepiloci (62.5 % methylated). Roughly, reproductive and vegetative samples formed coherent clusters, although male flowers appeared akin to vegetative organs (Fig. 3a). Seven correlated markers characterized vegetative samples (two exclusive of adult leaves), and four distinguished reproductive samples (two appeared only in floral buds; Fig. 3a, S2a). A mCG marker (mF16m28) set apart the male floral buds (Fig. 3a). All significant correlations ($r \geq 0.7$) involved subepiloci highly represented in adult leaves (Fig. S2a). Additionally, differentiation between the genders was maximized via 15 subepiloci, where methylation conditions were equally represented. Samples from male and female plants were clearly differentiated, except for a sample of young leaves from a male plant that appeared nested among samples taken from females (Fig. 3b). 80 % of retained variables showed strong correlation ($r \geq 0.9$) in those samples gathered from males. Epilocus C13m33 was also retained in *I. dumosa*; here, organs sampled from females were characterized by mCG and those sampled from males by the unmethylated condition (Fig. S2b).

cDNA-AFLP analyses

The complete dataset consisted of 1157 loci derived from 15 selective primer combinations (average number of loci per primer pair: 47.3 for *I. paraguariensis* and 47.1 for *I. dumosa*). Estimated genotyping error rates per primer

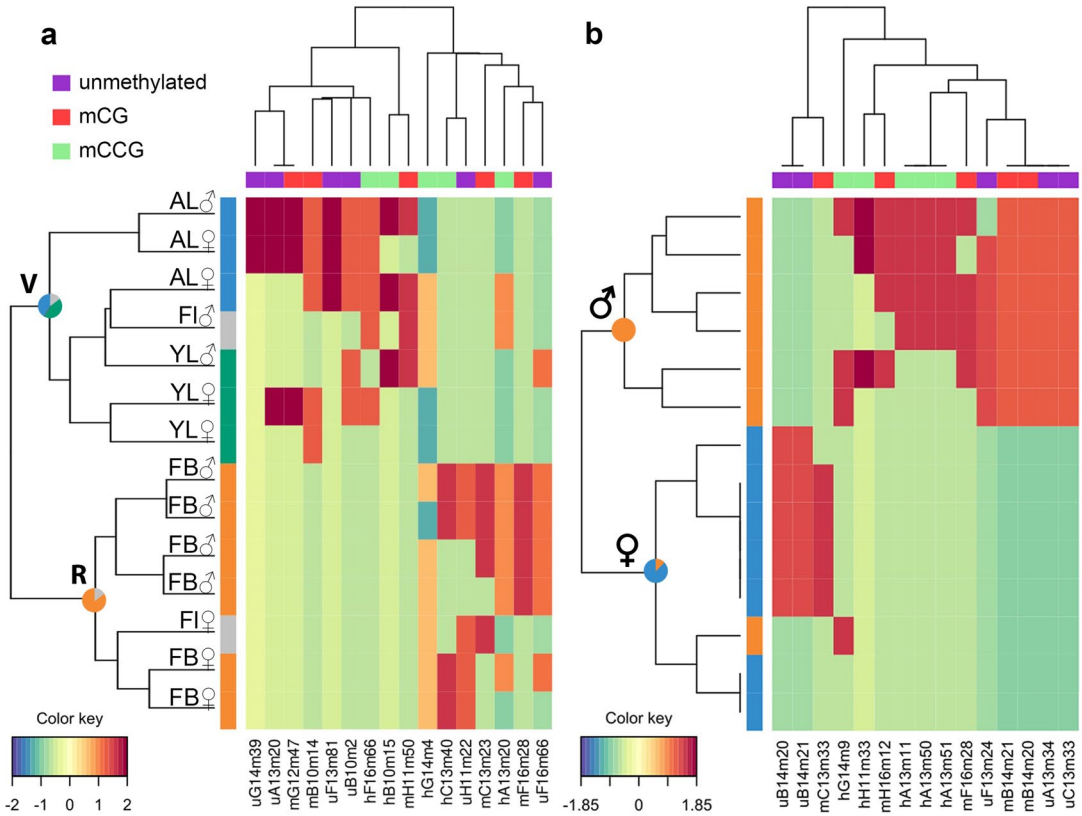

Fig. 3 sPLS-DA heatmaps from *Ilex dumosa*'s epigenetic $MS2_{YS}$ matrix. The reduced sets of discriminant subepiloci are indicated, in the upper part, as follows: unmethylated in violet, mCCG in light green, and mCG in red. A single principal component was retained in each heatmap. Correlation coefficients are displayed through a color gradation. The left dendrogram displays relationships among samples; the upper dendrogram displays relationships among the variables. Relevant nodes were labelled with a pie graph denoting cluster composition. **a** Maximization of sample differentiation according to organ types (shown with colors on the left). Floral buds (FB) are denoted in orange, flowers (Fl) in light grey, young leaves (YL) in dark green and adult leaves (AL) in light blue. Samples taken from female (♀) or male (♂) plants are indicated. V, vegetative organs; R, reproductive organs. **b** Maximization of gender differentiation. Female samples (♀) are shown in light blue; male samples (♂) are displayed in orange

combination ranged from 0.01 to 0.17 for *I. paraguariensis* and 0.03 to 0.15 for *I. dumosa* (Table S4). Both matrices (C_{YM} and C_{YS}) had equal amounts of variable loci. Species shared 23.34 % of bands, but C_{YS} was significantly more diverse (Table S8).

For each species, sPLS-DA analyses considering organ type and gender roughly indicated that each sample had a unique combination of highly correlated markers (Fig. 4a, b). Yet, coherent clusters according to the organ type were apparent; the expression patterns of floral buds resemble those of young leaves, and in a lesser extent to that of adult leaves (Fig. 4a, b). Samples taken from *I. dumosa* tended to form groups according to gender (Fig. S3a, b).

Characterization of differential fragments

In total, 102 differentially methylated regions and expressed genes were characterized with bioinformatic tools. Seven sequences were recovered in duplicate and one in triplicate, thus, 93 unique regions or genes could be tentatively identified (Tables S9, S10). Of these, 68.42% gave significant hits when blasted against *I. paraguariensis* transcriptomic databases, allowing the recovery of 65 full-length transcripts. All in all, 43 regions and transcripts were annotated, and further classified in four categories (Tables S9, S10): 15 genes code proteins or RNAs that participate in the regulation of gene expression at DNA, mRNA, and protein levels; 12 genes products are related to metabolism; eight to biotic and/or abiotic stress responses (such as pathogen defenses, disease resistances, etc.); and eight are associated to functions such as signaling, protein transport and cell structure (Fig. 5). Among the unannotated, 64% yielded non-significant hits with sequences of NCBI's databases; 28% gave significant hits but the subjects were anonymous (genomic or unannotated); 6% gave no hits in each search performed; a single fragment (2%) emerged as an artifactual chimera (Table S11). No significant similarities with transposons were found.

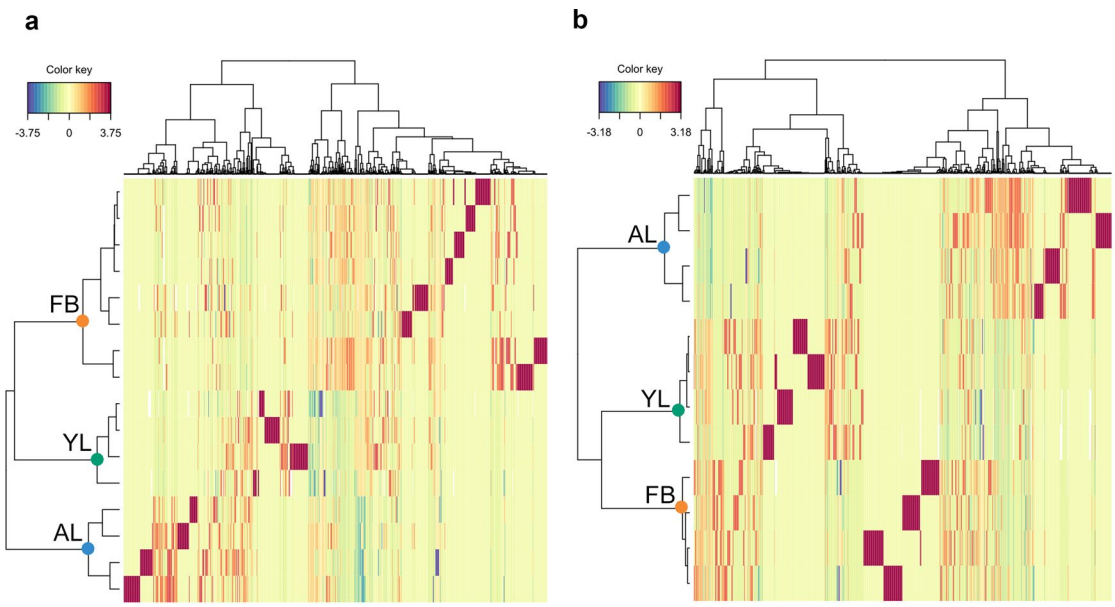

Fig. 4 PLS-DA heatmaps maximizing differences among organ types using cDNA-AFLP matrices. Floral buds' (FB) clusters are denoted with orange dots, young leaves' (YL) clusters with green dots and adult leaves' (AL) clusters with light blue dots. Heatmaps display correlation coefficients through a color gradation. **a** Heatmap obtained for *Ilex paraguariensis* employing the C_{YM} matrix with 701 total markers. **b** Heatmap obtained for *I. dumosa* employing the C_{YS} matrix with 687 total markers

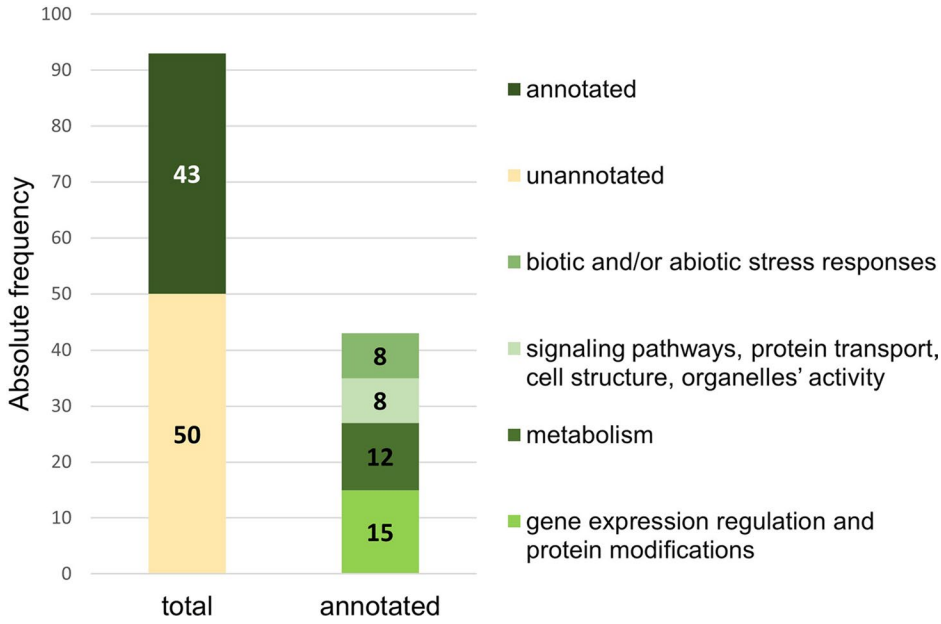

Fig. 5 Characterization of the sequenced fragments, detailing the absolute frequency of genes belonging to each of the proposed functional categories

Validation of differential fragments from *I. paraguariensis*

Based on their observed methylation or gene expression patterns, and on their deduced functional category, four fragments were selected for validation. These fragments were identified as: phosphatidylinositol 4-kinase gamma 8-like (PI4KG8-like; Genbank accession MT495509), AGAMOUS (AG; MT495520), flavonol synthase (FLS; MT495535) and chlorophyll a/b-binding protein 1 (CAB1; MT495536).

Neither gender nor organ-gender interaction effects were detected (Online Resource 1). Gene PI4KG8-like showed a statistically significant higher expression (i.e., lowest ΔC_t value) in late floral primordia (buds-III; $P < 0.01$; Fig. 6a) which also had the highest quantified expression of AG. This gene was also highly expressed in early stages of floral development (bud-II) and was found in low quantities

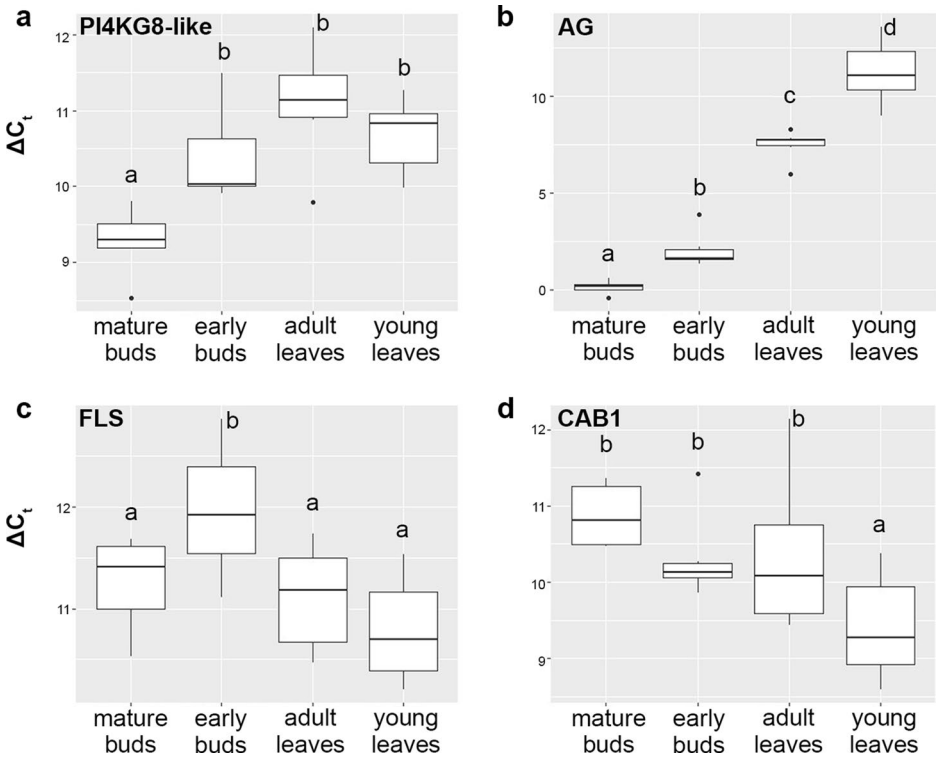

Fig. 6 ΔC_t values calculated for differential fragments tested on reproductive and vegetative samples. **a** *phosphatidylinositol 4-kinase gamma 8-like* (PI4KG8-like). **b** *AGAMOUS* (AG). **c** *flavonol synthase* (FLS). **d** *chlorophyll a/b-binding protein 1* (CAB1). Lower ΔC_t values indicate higher expression. The lower and upper hinges correspond to the first and third quartiles (the 25th and 75th percentiles); the median is denoted as a horizontal line. Whiskers represent 1.5 times the inter-quartile range. Individual points represent outlier data. Statistically significant differences between biological groups are indicated with different letters. Early buds, corresponds to the bud-II stage; mature buds refers to the bud-III stage

in leaves. All comparisons for AG expression were significant ($P < 0.001$; Fig. 6b). The expression of FLS was significantly lower in buds-II ($P < 0.05$; Fig. 6c). CAB1 was highly expressed in young leaves, and this was statistically significant ($P < 0.05$; Fig. 6d).

Discussion

Herein, we characterized the epigenomes and transcriptomes of reproductive and vegetative organs from individuals of *I. paraguariensis* and *I. dumosa* of both sexes, as a contribution towards elucidating their operative system of sexual differentiation. All in all, present results allowed the recognition of some candidate regions that are worth studying as gender associated markers, scorable in vegetative organs.

The two species were clearly distinguished by both their epigenetic and gene expression fingerprinting profiles; previous works reported differences in their genetic fingerprints (Gottlieb and Poggio 2014; Gottlieb et al. 2005, 2011). Our results evidenced that for most of the intraspecific comparisons performed (60%), the unmethylated subepiloci prevailed over mCG sites, and that these were more frequent than mCCG loci. Also, herein a higher epigenetic than 'genetic' diversity was revealed within the limited sampling of each species. It is worth noting that present 'genetic' results are an approximation as they relied on the NML chosen by the *msap* algorithm (Perez Figueroa 2013). In a previous work, we used dominant markers to genotype all the plants of *I. paraguariensis* kept at the "Jardín Botánico Carlos Thays de Buenos Aires" (Cascales 2018). These plants (13 males and six females) showed a Shannon's information index (I) of 0.726 ± 0.245. A subset of those plants (two males and two females) was examined here, yielding a comparable value ($I = 0.501 \pm 0.225$) even though it derives from NML. Another genotyping analysis of dominant markers from *I. paraguariensis* plants from a breeding population ($N = 128$) reported a genetic diversity value of 0.569 (Paiva et al. 2020). Then, it is reasonable to conclude that, at least for this species, the NML render a decent indicator of genetic diversity. Consequently, the higher Shannon's index value obtained with epigenetic data ($I = 0.680$) suggests that *I. paraguariensis´* sampled organs have high epigenetic diversity.

This higher epigenetic than 'genetic' diversity was revealed for both species. Studies on non-clonal plant species also observed a prevalence of epigenetic diversity (Herrera and Bazaga 2010; Lira-Medeiros et al. 2010; Medrano et al. 2020), suggesting a rather usual pattern as proposed by Schulz et al. (2014). However, as different experimental designs and analytical criteria are documented in the MSAP literature (Alonso et al. 2016; Fulneček and Kovařík 2014; Schulz et al. 2013), generalizations should be taken cautiously. For both *Ilex* species, the epigenetic and the indirect 'genetic' information appeared significantly and positively correlated. The

correlation is interpreted as indicative of influence of the genetic background on the epigenetic variation, following Richards (2006). Likewise, such correlation was reported in natural populations of *Viola elatior* (Schulz et al. 2014) and of wild potato (*Solanum* sp; Cara et al. 2013).

Methylation patterns tended to set apart reproductive and vegetative samples from *I. dumosa*. Organ-specific MSAP patterns were detected in leaves and flowers of the turnip (*Brassica rapa*; Kellenberger et al. 2016), and in adult and young leaves and root tips of cassava (*Manihot esculenta*; Kitimu et al. 2015). In addition, changes in methylation during the floral ontogeny were evidenced in 'yerba mate'; it was shown an incremental trend in mCG sites that reverts towards maturation, whereas for mCCG sites the opposite trend was apparent. As external methylations (mCCG) are associated with transcriptional suppression and transposon silencing (Bouyer et al. 2017; Crisp et al. 2019; Niederhuth et al. 2016; Zhang et al. 2018), present results would suggest, altogether, a generalized gene repression before anthesis. Similar changes were visualized in 'yerba señorita', but due to the current reduced sampling, interpretations are merely tentative. Moreover, the epigenetic—'genetic' data correlation represents an issue when faced with detecting organs' differences, as discrepancies among biological replicates may prevail. Unfortunately, inbred lines are unavailable for these two species, and the dioecy and the long generation times of *Ilex* obligates comparing distinct genetic backgrounds (at best, from half siblings).

Also, differential gene expression patterns were detected here among floral primordia, young and adult leaves. Particularly, the patterns displayed by young leaves appeared more akin to stage III floral primordia than to fully expanded leaves. This suggest that, to some extent, actively growing tissues would have similar expression patterns, and that subjacent transcriptomes do differ even when methylation patterns are not so discernible. Similar relationships among patterns of gene expression of reproductive and vegetative organs were observed in *Eucalyptus grandis* (Vining et al. 2015). In addition, present results validate the use of cDNA-AFLP for fishing active genes.

Four organ-specific epigenetic markers were detected in *I. dumosa*, two from floral buds -hC13m40 and mF16m28- and two from adult leaves -uG14m39 and uF13m81-. In addition, for this species we found two private subepiloci in reproductive organs (uH11m22 and mC13m23) and other five private subepiloci in the leaves (uA13m20, mG12m47, mB10m14, uB10m2 and hB10m15). Previous MSAP surveys in rice (Xiong et al. 1999) and maize inbred lines (Lu et al. 2008) also yielded organ-specific fragments. Among the fragments selected here, the mCG marker mF16m28 distinguished male floral buds, and thus, deserves further identification and characterization. Likewise, additional research should be done to confirm the organ- and gender-specificity of these candidate markers.

Dioecy is a feature shared by > 600 accepted species of *Ilex* (Loizeau et al. 2016), thus, whatever the operating system, it is expected to predate the divergence of extant taxa. Furthermore, in this genus the androecium and gynoecium begin to develop early in every floral primordium and, at a certain point, the development of one whorl become arrested to yield a unisexual flower (showing either pistillodia or staminodia; Giberti 2001; Giberti and Gurni 2008). These processes suggest that every individual holds the genetic information for the generation of male and female whorls, and support the hypothesis that an epigenetically regulated gene expression network, including the MADS-box homeotic genes or their downstream genes, would be involved (Cronk and Müller 2020; Khadka et al. 2019; Minelli 2018; Pajoro et al. 2014). In this context, we assumed that *I. paraguariensis* and *I. dumosa* have similar sexual differentiation systems, not overlooking that some variations could occur within a single genus. For instance, in *Silene* XY and ZW sex determination systems were inferred (Martin et al. 2019), and distinct architectures of the sex-determining region were reported in *Populus* (Geraldes et al. 2015).

Our MSAP survey allowed discrimination of samples according to the gender of the donor, especially for *I. dumosa*. For this species, the sexes were also accurately distinguished through PLS-DA of cDNA-AFLP patterns. For *I. paraguariensis*, a more elusive relationship among the MSAP-derived clusters was retrieved, and the expression patterns were ambiguous. Further investigations would benefit from using genetically homogeneous, full sibling, and sexually mature plants to focus on gender-related transcripts. All in all, DIABLO analyses revealed 19 potential gender-specific markers for *I. paraguariensis* and 15 for *I. dumosa*. Noteworthy are the markers detected in *I. paraguariensis*, namely a female specific marker (the unmethylated version of epilocus C13m33) and three subepiloci that were deemed as male-specific markers (the mCG-methylated version of C13m33, and the mCCG subepiloci hA13m20 and hA13m21). These four markers are interesting because they not only are detectable in reproductive organs but also in leaves, an essential requisite for developing an early sexing system. Then, most of the retained markers in *I. dumosa* (14 out of 15) are present in both reproductive and vegetative organs. Particularly, epiloci B14m20 and B14m21 appear unmethylated in females but mCG-methylated in male plants. Interestingly, the methylation condition of epilocus C13m33 seems to be indicative of gender in both species, but in opposite ways. Functionally male plants of *I. paraguariensis* have this site internally methylated and it arises unmethylated in *I. dumosa*, whereas female plants of *I. paraguariensis* have this site unmethylated and it is mCG-methylated in those of *I. dumosa*. Gender-specific methylations

have been reported in poplars (*Populus tomentosa*, Song et al. 2012; *P. balsamifera*, Bräutigam et al. 2017), persimmons (*Diospyros kaki*; Akagi et al. 2016) and papaya (*Carica papaya*; Liu et al. 2018), among other angiosperms.

It was reported that gene-associated DNA methylation could influence gene expression whether the methylated cytosine falls within the promoter region or the transcribed gene body (Zhang et al. 2018). Thus, the epiloci detected here deserve further study to achieve their characterization. In turn, when the reference genome of *I. paraguariensis* becomes available, a thorough epigenetic profiling could be accomplished, for instance, through a Whole Genome Bisulfite Sequencing (Frommer et al. 1992; Taudt et al. 2018) or a Reduced Representation Bisulfite Sequencing (Meissner et al. 2005). We are aware that gender-specific expression of a gene is not indicative of its involvement in sexual differentiation as it could co-segregate with the gender. Yet, such markers could be capitalized to launch a sexing procedure on plantlets, of which commercial producers and improvement programs of 'yerba mate' would benefit.

Among differentially methylated and expressed fragments that were characterized bioinformatically for both species, 70% of full-length transcripts were recovered with confidence and a rather high proportion of genes were annotated. Most products appeared involved in the regulation of gene expression at diverse levels (i.e., transcription, translation, post-translational modifications, etc.), and in cellular metabolism. The differential expression of four interesting markers were validated in 'yerba mate'. One of these, the epilocus G14m39, changes its methylation condition during flower development in male plants from unmethylated to mCG-methylated. This gene was identified as a kinase-like enzyme (PI4KG8) that catalyzes the phosphorylation of the phospholipid phosphatidylinositol leading to the synthesis of the second messenger inositol 1,4,5-trisphosphate (InsP3) (The UniProt Consortium 2019). The phosphatidylinositol may participate in the regulation of plant cytoskeletal structures and in polar pollen tube growth (Mueller-Roeber and Pical 2002), whereas InsP3 participates in responses to biotic and abiotic stress conditions (Hung et al. 2014; Na and Metzger 2020). Our results also suggest that PI4KG8 is more actively expressed in floral primordia (ca. 1.2–3.2 times), being more abundant in mature buds. Taking together, the methylation of internal cytosines could enhance PI4KG8 expression during staminate (male) flower development. It is suspected that mCG methylation within transcribed regions (i.e., gene body methylation) is associated with constitutive expression (Niederhuth et al. 2016; Zhang et al. 2018), but no direct evidence supports such a causal link (Bewick et al. 2019). Therefore, the analysis of methylation status and gene expression of this gene on a wider sampling will warrant the verification of our results.

Then, we selected for validation a partial K-box domain of the floral homeotic gene AGAMOUS (AG), which was isolated from both species and genders. This AG gene encodes a transcription factor of the MADS-domain protein family. According to the ABC(DE) model that explains flower development (Minelli 2018; Pajoro et al. 2014), AG participates in stamens and carpels identity (C-function; Dreni and Kater 2014; Zhu et al. 2018). In accordance, our results showed that in *I. paraguariensis* this gene is highly expressed in floral buds, and even more in mature stages where carpels or stamens were fully developed. The maturity of sexual structures was verified through histological preparations (Garberoglio Mariana Julia, Rosenfeldt Sonia, González Graciela Esther, Gottlieb Alexandra Marina unpublished). Due to its conserved C-function across angiosperms (Dreni and Kater 2014), it could be safely assumed similar expression patterns for *I. paraguariensis* and *I. dumosa*. Still, this transcript was also detected in the leaves of *I. paraguariensis*, though in very low quantities. A similar result was reported for *Malus* x *domestica* (van der Linden et al. 2002) and for *Aristolochia fimbriata* (Pabón-Mora et al. 2015). The presence of transcripts in leaves could be related to sub-functionalization or neo-functionalization of AG homologues, as reported in a diverse array of taxa (Ma et al. 2018).

The third gene characterized codes for the enzyme flavonol synthase (FLS) that catalyzes the production of flavonoids such as kaempferol, rutin and quercetin (Ayabe et al. 2010; Prescott et al. 2002). The leaves of *I. paraguariensis* contain higher concentrations of rutin (11x) and quercetin (8x) than those of *I. dumosa*, but both species produce similar levels of kaempferol (Filip et al. 2001). These high contents of flavonoids are responsible of the antioxidant properties of the "mate" infusion (Anesini et al. 2012; Filip et al. 2001). Our transcript quantification assay showed that the highest concentration of FLS occur in leaves and mature buds; altogether, current results suggest an increment of this enzyme during floral development. More research is needed to explore the regulatory mechanisms of flavonoids biosynthesis and its use for improvement programs towards quality and nutraceutical values.

The chlorophyll a/b-binding protein 1, also called light-harvesting chlorophyll a/b-protein 1.3 (LHCB1.3), was identified here from a fragment isolated from female reproductive green organs (i.e., in initial stages). This protein is a subunit of the light-harvesting complex II, which absorbs light and transfers energy to the photosynthetic reaction center (TAIR 2020, Umate 2010). Although no gender-specific expression was verified, refuting the female-specific expression suggested by cDNA-AFLP patterns, we demonstrated that its expression level in young leaves was nearly twice that of adult leaves. This result

could be explained by an increased photosynthetic activity of growing tissues (Wolf 1993).

AFLP-based techniques are highly efficient for detecting enzymatically restricted polymorphic fragments. The difficulty in identifying homologous markers, the potential non-independence of bands and the dominant nature of these anonymous markers, are the main limitations and were discussed elsewhere (Mueller and Wolfenbarger 1999; Schrey et al. 2013). However, fingerprinting techniques allow the simultaneous screening of many high-resolution markers randomly distributed throughout the genome and are especially suitable for non-model species when genomic reference data is lacking (Herrera and Bazaga 2010). In addition, implementation of AFLP-based techniques still implies lower costs compared to NGS-based techniques. As to MSAP fingerprinting, it detects variations in methylation that occur solely at the target sites of the frequent cutter endonucleases, underestimating the overall level of methylation (Perez Figueroa 2013). Still, the combinations of selective primers employed in this study allowed the recognition of several diagnostic bands within each biological group of samples even when a small portion of the genomes were scrutinized (11.89% of *I. paraguariensis* and 8.73% of *I. dumosa;* Gottlieb and Poggio 2014).

Ilex paraguariensis and *I. dumosa* are known to have contrasting phenotypical, physiological, and genetic traits. Here, we add two more layers of differentiation: the epigenetic and gene expression levels. In addition, we have shown that the organs herein studied have distinctive gene expression patterns, and that reproductive and vegetative organs and sexes tend to sustain particular epigenetic marks. Although it cannot be ruled out the involvement of other epigenetic processes such as histone modifications and siRNA pathways (as proposed by Pikaard and Scheid 2014) in organs ontogeny and in sexual differentiation, our approach allowed detecting and characterizing concrete candidate markers that deserve further evaluation. The detection of agronomical and/or metabolic interesting gene variants and the understanding of their regulation constitute valuable information for the application of successful biotechnological improvement programs.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10265-021-01279-3.

Acknowledgements The authors are indebted to Ing. Agr. F. Cano, G. Benito, and G. Barreiro from "Jardín Botánico Carlos Thays de Buenos Aires" and to Dr. P.A. Sansberro (IBONE-UNNE), who kindly provided specimens from living collections. We are also grateful to Drs. L. Babino, D. Kachanovsky, J.C. Vilardi, M.S. Rodriguero, and G.R. Cueto (all from FCEN, UBA) for their helpful and kind technical assistance, and we thank the anonymous reviewers whose suggestions helped to improve this manuscript. Grants from the Universidad de Buenos Aires (UBACYT 20020170100614BA) and the Agencia Nacional de Investigaciones Científicas y Técnicas (PICT 2015–2525) are gratefully acknowledged.

Author contributions JC and AMG conceived and designed the experiments and wrote the manuscript; RMA helped in qPCR assays and results interpretation; JC performed the experiments; JC and DIP analyzed the data; reagents were provided through grants awarded to AMG. All authors read and approved the manuscript.

Declarations

Conflict of interest The authors declare no conflict of interest. The funding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

Acevedo RM, Maiale SJ, Pessino SC et al (2013) A succinate dehydrogenase flavoprotein subunit-like transcript is upregulated in *Ilex paraguariensis* leaves in response to water deficit and abscisic acid. Plant Physiol Biochem 65:48–54. doi:https://doi.org/10.1016/j.plaphy.2012.12.016

Acevedo RM, Ruiz OA, Sansberro PA (2016) Gene expression changes in response to drought stress in *Ilex paraguariensis* leaves. Plant Omics 9:334–343. doi:https://doi.org/10.21475/poj.09.05.16.pne185

Acevedo RM, Avico EH, Ruiz OA, Sansberro PA (2018) Assessment of reference genes for real-time quantitative PCR normalization in *Ilex paraguariensis* leaves during drought. Biol Plant 62:89–96. doi:https://doi.org/10.1007/s10535-017-0754-3

Acevedo RM, Avico EH, González S et al (2019) Transcript and metabolic adjustments triggered by drought in *Ilex paraguariensis* leaves. Planta 250:445–462. doi:https://doi.org/10.1007/s00425-019-03178-3

Akagi T, Henry IM, Kawai T et al (2016) Epigenetic regulation of the sex determination gene *MeGI* in polyploid persimmon. Plant Cell 28:2905–2915. doi:https://doi.org/10.1105/tpc.16.00532

Alonso C, Pérez R, Bazaga P et al (2016) MSAP markers and global cytosine methylation in plants: a literature survey and comparative analysis for a wild-growing species. Mol Ecol Resour 16:80–90. https://doi.org/10.1111/1755-0998.12426

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. doi:https://doi.org/10.1016/j.lwt.2011.06.015

Ayabe S, Uchiyama H, Aoki T, Akashi T (2010) Plant phenolics: phenylpropanoids. In: Mander L, Liu H-W (eds) Comprehensive natural products II. Chemistry and biology, 1st edn. Elsevier, UK, pp 929–976

Bachem CWB, van der Hoeven RS, de Bruijn SM et al (1996) Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development. Plant J 9:745–753. https://doi.org/10.1046/j.1365-313X.1996.9050745.x

Bartels A, Han Q, Nair P et al (2018) Dynamic DNA methylation in plant growth and development. Int J Mol Sci 19:2144. doi:https://doi.org/10.3390/ijms19072144

Bewick AJ, Zhang Y, Wendte JM et al (2019) Evolutionary and experimental loss of gene body methylation and its consequence to gene expression. G3 Genes. Genomes Genet 9:2441–2445. doi:https://doi.org/10.1534/g3.119.400365

Bonin A, Bellemain E, Bronken Eidesen P et al (2004) How to track and assess genotyping errors in population genetics studies. Mol Ecol 13:3261–3273. doi:https://doi.org/10.1111/j.1365-294X.2004.02346.x

Bouyer D, Kramdi A, Kassam M et al (2017) DNA methylation dynamics during early plant life. Genome Biol 18:179. doi:https://doi.org/10.1186/s13059-017-1313-0

Bracesco N, Sanchez AG, Contreras V et al (2011) Recent advances on *Ilex paraguariensis* research: minireview. J Ethnopharmacol 136:378–384. https://doi.org/10.1016/j.jep.2010.06.032

Bräutigam K, Cronk Q (2018) DNA methylation and the evolution of developmental complexity in plants. Front Plant Sci 9:1447. doi:https://doi.org/10.3389/fpls.2018.01447

Bräutigam K, Soolanayakanahally R, Champigny M et al (2017) Sexual epigenetics: gender-specific methylation of a gene in the sex determining region of *Populus balsamifera*. Sci Rep 7:45388. https://doi.org/10.1038/srep45388

Cara N, Marfil CF, Masuelli RW (2013) Epigenetic patterns newly established after interspecific hybridization in natural populations of *Solanum*. Ecol Evol 3:3764–3779. doi:https://doi.org/10.1002/ece3.758

Cardozo Junior EL, Morand C (2016) Interest of mate (*Ilex paraguariensis* A. St.-Hil.) as a new natural functional food to preserve human cardiovascular health—a review. J Funct Foods 21:440–454. https://doi.org/10.1016/j.jff.2015.12.010

Cascales J (2018) Estudios en expresión génica diferencial y genómica comparativa en yerba mate y yerba señorita. Dissertation, Universidad de Buenos Aires

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. doi:https://doi.org/10.1007/s10709-014-9804-3

Cascales J, Bracco M, Garberoglio MJ et al (2017) Integral phylogenomic approach over *Ilex* L. species from southern South America. Life 7:47. doi:https://doi.org/10.3390/life7040047

Charlesworth D (2016) Plant sex chromosomes. Annu Rev Plant Biol 67:397–420. doi:https://doi.org/10.1146/annurev-arplant-043015-111911

Crisp PA, Noshay JM, Anderson SN, Springer NM (2019) Opportunities to use DNA methylation to distil functional elements in large crop genomes. Mol Plant 12:282–284. doi:https://doi.org/10.1016/j.molp.2019.02.006

Cronk Q, Müller NA (2020) Default sex and single gene sex determination in dioecious plants. Front Plant Sci 11:1162. doi:https://doi.org/10.3389/fpls.2020.01162

Debat HJ, Grabiele M, Aguilera PM et al (2014) Exploring the genes of yerba mate (*Ilex paraguariensis* A. St.-Hil.) by NGS and de novo transcriptome assembly. PLoS One 9:e109835. doi:https://doi.org/10.1371/journal.pone.0109835

Diggle PK, Di Stilio VS, Gschwend AR et al (2011) Multiple developmental processes underlie sex differentiation in angiosperms. Trends Genet 27:368–376. doi:https://doi.org/10.1016/j.tig.2011.05.003

Dreni L, Kater MM (2014) MADS reloaded: evolution of the AGAMOUS subfamily genes. New Phytol 201:717–732. https://doi.org/10.1111/nph.12555

Faraway JJ (2016) Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models. Chapman and Hall/CRC, Boca Raton

Fay JV, Watkins CJ, Shrestha RK et al (2018) Yerba mate (*Ilex paraguariensis*, A. St.-Hil.) de novo transcriptome assembly based on tissue specific genomic expression profiles. BMC Genomics 19:891. doi: https://doi.org/10.1186/s12864-018-5240-6

Filatov DA (2015) Homomorphic plant sex chromosomes are coming of age. Mol Ecol 24:3217–3219. doi:https://doi.org/10.1111/mec.13268

Filip R (2009) Mate (*Ilex paraguariensis*). In: Imperato F (ed) Recent advances in phytochemistry research. Signpost, Kerala, pp 113–131

Filip R, López P, Coussio J, Ferraro G (1998) Mate substitutes or adulterants: study of xanthine content. Phyther Res 12:129–131. doi:https://doi.org/10.1002/(SICI)1099-1573(199803)12:2%3c129::AID-PTR191%3e3.0.CO;2-1

Filip R, López P, Giberti GC et al (2001) Phenolic compounds in seven South American *Ilex* species. Fitoterapia 72:774–778. https://doi.org/10.1016/S0367-326X(01)00331-8

Frommer M, McDonald LE, Millar DS et al (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc Natl Acad Sci USA 89:1827–1831. https://doi.org/10.1073/pnas.89.5.1827

Fulneček J, Kovařík A (2014) How to interpret Methylation Sensitive Amplified Polymorphism (MSAP) profiles? BMC Genet 15:2. doi:https://doi.org/10.1186/1471-2156-15-2

Geraldes A, Hefer CA, Capron A et al (2015) Recent y chromosome divergence despite ancient origin of dioecy in poplars (*Populus*). Mol Ecol 24:3243–3256. doi:https://doi.org/10.1111/mec.13126

Giberti GC (1979) Las especies argentinas del género *Ilex* L. (Aquifoliaceae). Darwiniana 22:217–240

Giberti GC (1989) Los parientes silvestres de la yerba mate y el problema de su adulteración. Dominguezia 7:1–22

Giberti GC (1994a) Aquifoliaceae. In: Spichiger R, Ramella L (eds) Flora del Paraguay. Editions des Conservatoire et Jardin botaniques de la Ville de Geneva. Missouri Botanical Garden, Switzerland, p 24

Giberti GC (1994b) Maté (*Ilex paraguariensis*). In: Hernández Bermejo JE, León J (eds) Neglected crops. 1492 from a different perspective. Food and Agriculture Organization of the United Nations, Roma, pp 245–252

Giberti GC (2001) Diferentes aspectos del género *Ilex* (Aquifoliaceae). Corología, arquitectura floral, posición sistemática. Dissertation, Universidad de Buenos Aires

Giberti GC (2011) South American *Ilex* raw materials for regional infusions: Advanced arboreal crop systems versus wild forest extractive production. In: Filip R (ed) Multidisciplinary Approaches on Food Science and Nutrition for the XXI Century. Kerala, pp 17–27

Giberti GC, Gurni AA (2008) Anatomía floral comparada de once especies sudamericanas de *Ilex* L. (Aquifoliaceae) relacionadas con la yerba mate. Dominguezia 24:77–94

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. doi:https://doi.org/10.1007/s10709-010-9449-9

Gottlieb AM, Poggio L (2014) Quantitative and qualitative genomic characterization of cultivated *Ilex* L. species. Plant Genet Resour Characterisation Util 13:142–152. doi:https://doi.org/10.1017/S1479262114000756

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. doi:https://doi.org/10.3732/ajb.92.2.352

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

Greizerstein EJ, Giberti GC, Poggio L (2004) Cytogenetic studies of southern South American *Ilex* species. Caryologia 57:19–23. doi:https://doi.org/10.1080/00087114.2004.10589367

Grela IA (2004) Geografía florística de las especies arbóreas de Uruguay: propuesta para la delimitación de dendrofloras. Dissertation, Universidad de la República

Guevara M, de María N, Sáez-laguna E et al (2017) Analysis of DNA cytosine methylation patterns using Methylation-Sensitive

Amplification Polymorphism (MSAP). In: Kovalchuck I, Zemp FJ et al (eds) Plant epigenetics: methods and protocols, methods in molecular biology. Springer Science + Business Media, New York, pp 99–112

Heikrujam M, Sharma K, Prasad M, Agrawal V (2015) Review on different mechanisms of sex determination and sex-linked molecular markers in dioecious crops: a current update. Euphytica 201:161–194. https://doi.org/10.1007/s10681-014-1293-z

Henry IM, Akagi T, Tao R, Comai L (2018) One hundred ways to invent the sexes: theoretical and observed paths to dioecy in plants. Annu Rev Plant Biol 69:553–575. https://doi.org/10.1146/annurev-arplant-042817-040615

Herrera CM, Bazaga P (2010) Epigenetic differentiation and relationship to adaptive genetic divergence in discrete populations of the violet *Viola cazorlensis*. New Phytol 187:867–876. doi:https://doi.org/10.1111/j.1469-8137.2010.03298.x

Herrera CM, Bazaga P (2013) Epigenetic correlates of plant phenotypic plasticity: DNA methylation differs between prickly and non-prickly leaves in heterophyllous *Ilex aquifolium* (Aquifoliaceae) trees. Bot J Linn Soc 171:441–452. doi:https://doi.org/10.1111/boj.12007

Hung CY, Aspesi P, Hunter MR et al (2014) Phosphoinositide-signaling is one component of a robust plant defense response. Front Plant Sci 5:267. doi:https://doi.org/10.3389/fpls.2014.00267

Kellenberger RT, Schlüter PM, Schiestl FP (2016) Herbivore-Induced DNA demethylation changes floral signalling and attractiveness to pollinators in *Brassica rapa*. PLoS One 11:e0166646. doi:https://doi.org/10.1371/journal.pone.0166646

Kersten B, Pakull B, Fladung M (2017) Genomics of sex determination in dioecious trees and woody plants. Trees Struct Funct 31:1113–1125. https://doi.org/10.1007/s00468-017-1525-7

Khadka J, Yadav NS, Guy M et al (2019) Epigenetic aspects of floral homeotic genes in relation to sexual dimorphism in the dioecious plant *Mercurialis annua*. J Exp Bot 70:6245–6259. doi:https://doi.org/10.1093/jxb/erz379

Kitimu SR, Taylor J, March TJ et al (2015) Meristem micropropagation of cassava (*Manihot esculenta*) evokes genome-wide changes in DNA methylation. Front Plant Sci 6:590. doi:https://doi.org/10.3389/fpls.2015.00590

Kuznetsova A, Brockhoff PB, Christensen RHB (2017) lmerTest package: tests in linear mixed effects models. J Stat Softw 82:1–26. https://doi.org/10.18637/jss.v082.i13

Le Cao K-A, Rohart F, Gonzalez I et al (2016) mixOmics: Omics Data Integration Project. R package version 6.1.1

Lira-Medeiros CF, Parisod C, Fernandes RA et al (2010) Epigenetic variation in mangrove plants occurring in contrasting natural environment. PLoS One 5:e10326. doi:https://doi.org/10.1371/journal.pone.0010326

Liu J, Chatham L, Aryal R et al (2018) Differential methylation and expression of *HUA1* ortholog in three sex types of papaya. Plant Sci 272:99–106. doi:https://doi.org/10.1016/j.plantsci.2018.04.001

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. Methods 25:402–408. doi:https://doi.org/10.1006/meth.2001.1262

Loizeau PA, Savolainen V, Andrews S, Spichiger R (2016) Aquifoliaceae. In: Kadereit JW, Bittrich V (eds) Flowering plants. Eudicots, The families and genera of vascular plants 14. Springer International Publishing, Basel, pp 31–36

Lu Y, Rong T, Cao M (2008) Analysis of DNA methylation in different maize tissues. J Genet Genomics 35:41–48. doi:https://doi.org/10.1016/S1673-8527(08)60006-5

Ma J, Shen X, Liu Z et al (2018) Isolation and characterization of *AGAMOUS-like* genes associated with double-flower morphogenesis in *Kerria japonica* (Rosaceae). Front Plant Sci 9:959. doi:https://doi.org/10.3389/fpls.2018.00959

Manen JF, Barriera G, Loizeau PA, Naciri Y (2010) The history of extant *Ilex* species (Aquifoliaceae): evidence of hybridization within a Miocene radiation. Mol Phylogenet Evol 57:961–977. https://doi.org/10.1016/j.ympev.2010.09.006

Martin H, Carpentier F, Gallina S et al (2019) Evolution of young sex chromosomes in two dioecious sister plant species with distinct sex determination systems. Genome Biol Evol 11:350–361. doi:https://doi.org/10.1093/gbe/evz001

McClelland M, Nelson M, Raschke E (1994) Effect of site-specific modification on restriction endonucleases and DNA modification methyltransferases. Nucleic Acids Res 22:3640–3659. doi:https://doi.org/10.1093/nar/22.17.3640

Medrano M, Alonso C, Bazaga P et al (2020) Comparative genetic and epigenetic diversity in pairs of sympatric, closely-related plants with contrasting distribution ranges in southeastern Iberian mountains. AoB Plants. https://doi.org/10.1093/aobpla/plaa013

Meissner A, Gnirke A, Bell GW et al (2005) Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res 33:5868–5877. doi:https://doi.org/10.1093/nar/gki901

Minelli A (2018) Genes and genomes. Plant evolutionary developmental biology. The evolvability of the phenotype. Cambridge University Press, Cambridge, pp 64–90

Moliterni VMC, Cattivelli L, Ranalli P, Mandolino G (2004) The sexual differentiation of *Cannabis sativa* L.: a morphological and molecular study. Euphytica 140:95–106. https://doi.org/10.1007/s10681-004-4758-7

Mueller-Roeber B, Pical C (2002) Inositol phospholipid metabolism in *Arabidopsis*. Characterized and putative isoforms of inositol phospholipid kinase and phosphoinositide-specific phospholipase C. Plant Physiol 130:22–46. doi:https://doi.org/10.1104/pp.004770

Mueller UG, Wolfenbarger LL (1999) AFLP genotyping and fingerprinting. Trends Ecol Evol 14:389–394. https://doi.org/10.1016/s0169-5347(99)01659-6

Na JK, Metzger JD (2020) A putative tomato inositol polyphosphate 5-phosphatase, Le5PT1, is involved in plant growth and abiotic stress responses. 3 Biotech 10:28. doi:https://doi.org/10.1007/s13205-019-2023-y

Niederhuth CE, Bewick AJ, Ji L et al (2016) Widespread natural variation of DNA methylation within angiosperms. Genome Biol 17:194. doi:https://doi.org/10.1186/s13059-016-1059-0

Pabón-Mora N, Suárez-Baron H, Ambrose BA, González F (2015) Flower development and perianth identity candidate genes in the basal angiosperm *Aristolochia fimbriata* (Piperales: Aristolochiaceae). Front Plant Sci 6:1095. doi:https://doi.org/10.3389/fpls.2015.01095

Paiva DI, Cascales J, Rosetti MEN et al (2020) Unraveling the genetic complexity of a cultivated breeding population of "yerba mate" (*Ilex paraguariensis* St. Hil.). An Acad Bras Cienc 92:e20190113. doi:https://doi.org/10.1590/0001-3765202020190113

Pajoro A, Biewers S, Dougali E et al (2014) The (r)evolution of gene regulatory networks controlling *Arabidopsis* plant reproduction: a two-decade history. J Exp Bot 65:4731–4745. https://doi.org/10.1093/jxb/eru233

Panigatti JL (2010) Argentina: 200 años, 200 suelos. Ediciones INTA, Buenos Aires

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. https://doi.org/10.1093/bioinformatics/bts460

Pérez-Figueroa A (2013) msap: a tool for the statistical analysis of methylation-sensitive amplified polymorphism data. Mol Ecol Resour 13:522–527. https://doi.org/10.1111/1755-0998.12064

Piferrer F (2013) Epigenetics of sex determination and gonadogenesis. Dev Dyn 242:360–370. doi:https://doi.org/10.1002/dvdy.23924

Pikaard CS, Scheid OM (2014) Epigenetic regulation in plants. Cold Spring Harb Perspect Biol 6:a019315. doi:https://doi.org/10.1101/cshperspect.a019315

Prat Kricun SD (2009) Antiguo cultivo, nueva infusión: inscriben dos variedades de *Ilex dumosa*. https://www.agrositio.com.ar/noticia/106378-antiguocultivo-nueva-infusion-inscriben-dos-variedades-de-ilex-dumosa. Accessed 09 May 2020

Prescott AG, Stamford NPJ, Wheeler G, Firmin JL (2002) *In vitro* properties of a recombinant flavonol synthase from *Arabidopsis thaliana*. Phytochemistry 60:589–593. doi:https://doi.org/10.1016/S0031-9422(02)00155-3

R Core Team (2019) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Reyna-Lopez GE, Simpson J, Ruiz-Herrera J (1997) Differences in DNA methylation patterns are detectable during the dimorphic transition of fungi by amplification of restriction polymorphisms. Mol Gen Genet 253:703–710. doi:https://doi.org/10.1007/s004380050374

Richards EJ (2006) Inherited epigenetic variation—revisiting soft inheritance. Nat Rev Genet 7:395–401

Rohart F, Gautier B, Singh A, Lê Cao K-A (2017) mixOmics: an R package for 'omics feature selection and multiple data integration. PLoS Comput Biol 13:e1005752. https://doi.org/10.1371/journal.pcbi.1005752

Schlüter PM, Harris SA (2006) Analysis of multilocus fingerprinting data sets containing missing data. Mol Ecol Notes 6:569–572. doi:https://doi.org/10.1111/j.1471-8286.2006.01225.x

Schrey AW, Alvarez M, Foust CM et al (2013) Ecological epigenetics: beyond MS-AFLP. Integr Comp Biol 53:340–350. https://doi.org/10.1093/icb/ict012

Schulz B, Eckstein RL, Durka W (2013) Scoring and analysis of methylation-sensitive amplification polymorphisms for epigenetic population studies. Mol Ecol Resour 13:642–653. doi:https://doi.org/10.1111/1755-0998.12100

Schulz B, Eckstein RL, Durka W (2014) Epigenetic variation reflects dynamic habitat conditions in a rare floodplain herb. Mol Ecol 23:3523–3537. doi:https://doi.org/10.1111/mec.12835

Shannon CE, Weaver W (1949) The mathematical theory of communication. University of Illinois Press, Urbana

Sneath PHA, Sokal RR (1973) Numerical taxonomy: the principles and practice of numerical classification. San Francisco, California

Song Y, Ma K, Bo W et al (2012) Sex-specific DNA methylation and gene expression in andromonoecious poplar. Plant Cell Rep 31:1393–1405. doi:https://doi.org/10.1007/s00299-012-1255-7

Song Y, Tian M, Ci D, Zhang D (2015) Methylation of microRNA genes regulates gene expression in bisexual flower development in andromonoecious poplar. J Exp Bot 66:1891–1905. doi:https://doi.org/10.1093/jxb/eru531

Souza AHP, Corrêa RCG, Barros L et al (2015) Phytochemicals and bioactive properties of *Ilex paraguariensis*: an *in-vitro* comparative study between the whole plant, leaves and stems. Food Res Int 78:286–294. https://doi.org/10.1016/j.foodres.2015.09.032

TAIR (2020) https://www.arabidopsis.org/servlets/TairObject?name=AT1G29930&type=locus. www.arabidopsis.org. Accessed 4 Feb 2020

Taudt A, Roquis D, Vidalis A et al (2018) METHimpute: imputation-guided construction of complete methylomes from WGBS data. BMC Genom 19:444. https://doi.org/10.1186/s12864-018-4641-x

The UniProt Consortium (2019) UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res 47:D506–D515. doi:https://doi.org/10.1093/nar/gky1049

Umate P (2010) Genome-wide analysis of the family of light-harvesting chlorophyll a/b-binding proteins in *Arabidopsis* and rice. Plant Signal Behav 5:1537–1542. doi:https://doi.org/10.4161/psb.5.12.13410

Untergasser A, Cutcutache I, Koressaar T et al (2012) Primer3-new capabilities and interfaces. Nucleic Acids Res 40:e115. doi:https://doi.org/10.1093/nar/gks596

van der Linden CG, Vosman B, Smulders MJM (2002) Cloning and characterization of four apple MADS box genes isolated from vegetative tissue. J Exp Bot 53:1025–1036. doi:https://doi.org/10.1093/jexbot/53.371.1025

Vanyushin BF, Ashapkin VV (2011) DNA methylation in higher plants: Past, present and future. Biochim Biophys Acta - Gene Regul Mech 1809:360–368. doi:https://doi.org/10.1016/j.bbagrm.2011.04.006

Vining KJ, Romanel E, Jones RC et al (2015) The floral transcriptome of *Eucalyptus grandis*. New Phytol 206:1406–1422. doi:https://doi.org/10.1111/nph.13077

Wen L, Yun X, Zheng X et al (2017) Transcriptomic comparison reveals candidate genes for triterpenoid biosynthesis in two closely related *Ilex* species. Front Plant Sci 8:634. doi:https://doi.org/10.3389/fpls.2017.00634

Wolf S (1993) Effect of leaf age on photosynthesis, carbon transport and carbon allocation in potato plants. Potato Res 36:253–262. doi:https://doi.org/10.1007/BF02360534

Xiong LZ, Xu CG, Saghai Maroof MA, Zhang Q (1999) Patterns of cytosine methylation in an elite rice hybrid and its parental lines, detected by a methylation-sensitive amplification polymorphism technique. Mol Gen Genet 261:439–446. doi:https://doi.org/10.1007/s004380050986

Yang HW, Akagi T, Kawakatsu T, Tao R (2019) Gene networks orchestrated by *MeGI*: a single-factor mechanism underlying sex determination in persimmon. Plant J 98:97–111. doi:https://doi.org/10.1111/tpj.14202

Yao X, Song Y, Yang J et al (2020) Phylogeny and biogeography of the hollies (*Ilex* L., Aquifoliaceae). J Syst Evol. doi:https://doi.org/10.1111/jse.12567

Ye J, Coulouris G, Zaretskaya I et al (2012) Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics 13:134. doi:https://doi.org/10.1186/1471-2105-13-134

Yu G, Duan J, Yan H et al (2011) cDNA-AFLP analysis of gene expression differences between the flower bud and sprout-shoot apical meristem of *Angelica sinensis* (Oliv.) Diels. Genet Mol Biol 34:274–279. doi:https://doi.org/10.1590/S1415-47572011000200018

Zhang H, Lang Z, Zhu JK (2018) Dynamics and function of DNA methylation in plants. Nat Rev Mol Cell Biol 19:489–506. doi:https://doi.org/10.1038/s41580-018-0016-z

Zheng X, Xu H, Ma X et al (2014) Triterpenoid saponin biosynthetic pathway profiling and candidate gene mining of the *Ilex asprella* root using RNA-SEq. Int J Mol Sci 15:5970–5987. doi:https://doi.org/10.3390/ijms15045970

Zheng X, Luo X, Ye G et al (2015) Characterisation of two oxidosqualene cyclases responsible for triterpenoid biosynthesis in *Ilex asprella*. Int J Mol Sci 16:3564–3578. doi:https://doi.org/10.3390/ijms16023564

Zhu B, Li H, Wen J et al (2018) Functional specialization of duplicated *AGAMOUS* homologs in regulating floral organ development of *Medicago truncatula*. Front Plant Sci 9:854. doi:https://doi.org/10.3389/fpls.2018.00854

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.