High-throughput sequencing analysis of *Eugenia uniflora*: insights into repetitive DNA, gene content and potential biotechnological applications

Valdir Marcos Stefenon¹ · Deise Schröder Sarzi² · Luiz Fernando Würdiger Roesch¹

Received: 13 November 2018 / Accepted: 24 April 2019 / Published online: 3 May 2019
© King Abdulaziz City for Science and Technology 2019

Abstract

In this study, the Ion Torrent PGM platform was employed to conduct a whole-genome sequencing analysis of *Eugenia uniflora* L., aiming to generate new genomic information for this non-model species. The whole-genome set of assembled sequences (WGSAS) generated 2601 contigs, with a cumulative length of 3.15 Mbp. About 2.60% of the WGSAS was characterized by repetitive sequences, while 71.66% of the WGSAS was represented by predicted genes. A total of 147 metabolic pathways related to enzymes characterized in the *E. uniflora* WGSAS were detected. Proteins related to antifungal activity and to bacterium and fungi defense response as well as transcriptions factors related to abiotic stress responses were also identified. This sequencing provides important genomic resources for future breeding and selection regarding plants with higher production of such metabolites, as well as source of genes for genetic engineering of cultivated species, supporting a wide application of marker-assisted and genomic selection, towards the establishment of commercial orchards with improved cultivars of *E. uniflora*.

Keywords Pitangueira · Functional annotation · Medicinal plant · Agriculture · Biotechnology

Introduction

Eugenia uniflora L. is a South American tree species of the Myrtaceae family, with recognized importance into pharmaceutical, fruticulture, and environmental sectors. The infusion of its leaves is used against gastrointestinal illnesses, while its secondary metabolites have demonstrated several biotechnological properties. *Eugenia uniflora* has been exploited by the pharmaceutical and cosmetic industries with attested anti-inflammatory functions (Costa et al.

2010), while its essential oils have been proposed as a potential natural insecticide (Cunha et al. 2015), and its extract demonstrated cytoprotective effect against mercury chloride and antioxidant activity (Cunha et al. 2016a, b). The fruits of *E. uniflora* are consumed fresh, as juice and ice cream (Coradin et al. 2011). It is also used for recovering mine-degraded soils and is widely used by rural populations for heating residences and manufacturing poles for fencing. Although still underexplored, local market in Southern Brazil is growing and the species is listed among the Brazilian fruit species with high commercial potential (Coradin et al. 2011).

Despite its importance, few initiatives have been developed towards generating genetic and genomic information on *E. uniflora*. Genetic studies about *E. uniflora* have mainly described the characterization of SSR molecular markers (Ferreira-Ramos et al. 2008, 2014; Sarzi et al. 2019) and the application of AFLP and cpSSR markers in population genetic studies (Margis et al. 2002; Salgueiro et al. 2004). Genomic data about identification of micro-RNAs (Guzman et al. 2012) and a de novo assembly of the transcriptome aiming to identify genes from the terpenoid biosynthesis pathway (Guzman et al. 2014) were just recently published for *E. uniflora*. At the time of writing, a project on

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s13205-019-1729-1>) contains supplementary material, which is available to authorized users.

✉ Valdir Marcos Stefenon
valdirstefenon@gmail.com

¹ Centro Interdisciplinar de Pesquisas em Biotecnologia, Universidade Federal do Pampa, Rod. BR290, Campus Universitário, São Gabriel, RS 97300-000, Brazil

² Universidade Federal Rural do Rio de Janeiro, Rod. BR 465, Km 07, s/n-Zona Rural, Seropédica-RJ, Rio de Janeiro, RJ 23890-000, Brazil

the sequencing of the genome of *Eugenia dysenterica* has been listed in GenBank under Bioproject PRJNA445709, but sequence data have not yet been submitted.

Based on flow cytometry, *E. uniflora* has a predicted haploid genome of about 244.99 Mbp (Costa et al. 2008). With the advent of the next generation sequencing (NGS) platforms, drafting small genomes becomes an attractive initiative towards generating genomic information for biotechnological exploration, conservation, and breeding of non-model tree species. Comprehensive information about the genome of non-model organisms can be achieved even with very low-coverage genome sequencing, including annotation of the main gene families and characterization of repetitive DNA regions. Following this rationale, the primary goal of this study was to build primary genomic resources for *E. uniflora*, characterizing the repetitive DNA distributed across the genome and identifying the overall gene content of the species through data from partial sequencing of the whole genome. Retrieving these data, we searched for general trends that could be extrapolated to the whole genome of *E. uniflora*.

Materials and methods

Sampling and DNA extraction

Total genomic DNA was isolated from healthy young leaves of one single adult plant of *E. uniflora* collected in a natural population within the Pampa biome in southern Brazil (30°20'05.00"S, 54°21'44.00"W). A voucher of the collected individual was deposited in the Herbarium Bruno Edgar Irgang of the Universidade Federal do Pampa, Campus São Gabriel (voucher HBEI1150). Total DNA was isolated with the DNeasy® Plant mini kit (Qiagen), following the manufacturers' instructions. The quality and the amount of the isolated DNA were evaluated on a NanoVue™ Plus Spectrophotometer (GE Healthcare) and through electrophoresis on 1.0% agarose gel, stained with GelRed® (Biotium Inc.) and visualized under UV-vis.

NGS sequencing and de novo assembling into contigs

For the NGS sequencing, the total genomic DNA was sheared in fragments of about 300 bp using Biorruptor® (Thermo Fisher Scientific) and the genomic libraries were built using the IonChef® (Thermo Fisher Scientific) system following the manufacturers' specifications. DNA fragments were sequenced on Ion 318™ microchip using the Ion Torrent Personal Genome Machine (Thermo Fisher Scientific) and the Ion PGM™ 400 Sequencing Kit following the manufacturers' specifications.

After sequencing, the reads were filtered by the PGM software, removing low quality (Phred score < 20) and polyclonal sequences. The high-quality reads were evaluated for the K-mer number by KmerGenie software (Chikhi and Medvedev 2014) and used for a de novo assembling, using the K-mer number 35. Reads (mean length of 170 bases) were assembled using SPAdes 3.09 (Bankevich et al. 2012) to produce a whole-genome set of assembled sequences (WGSAS) containing single contigs with a minimum size of 1000 nucleotides. This whole-genome Shotgun project was deposited at DDBJ/ENA/GenBank under the accession RQIG000000000 (version RQIG01000000.1), Bioproject PRJNA504673.

Characterization of repetitive elements and overall gene content

Repetitive elements within the WGSAS were identified using the Smith–Waterman alignment algorithm, as implemented in the RepeatMasker web server (<http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker>). RepeatMasker finds repetitive elements by aligning each of the query sequences with each of the repeat consensus sequences in the repeat library file. *Arabidopsis thaliana* was employed as reference repeat database (RepBase Update 20160829).

Gene prediction was performed using an ab initio approach run in Augustus software (Stanke and Morgenstern 2005). *Arabidopsis thaliana* was used as the model species and both complete and partial genes were predicted. Protein-coding genes were predicted using Blast2GPro (Conesa et al. 2005), using the standard workflow for gene ontology annotation and biological process, molecular function and cellular component as GO categories in the functional annotation analysis. Potential homolog sequences from the NCBI database (filtering to use the Magnoliophyta database, 3398 taxa) were searched using the Blastn (-task Megablast) option, and an *e* value threshold of $1.0e^{-5}$. Biological functions were determined based on protein domains and families, using the InterPro routine of Blast2GO, searching within the EMBL-EBI InterPro protein database with the default parameters of the program. Final protein-coding gene prediction was determined merging the Blastn and InterPro searches. Metabolic pathways containing enzymes recovered from the WGSAS were searched within the Kyoto Encyclopedia of Genes and Genomes (KEGG) database.

Results

Sequencing output and genome assemblage

After excluding the low-quality reads, the sequencing effort yielded 7.05×10^6 bases of sequences (Table 1),

Table 1 Data summary of the *E. uniflora* whole genome sequencing

Sequencing and assembling		
Number of sequenced bases		7,052,600
Estimated coverage ^a		0.03×
Average quality per read (Phred score)		30
Number of contigs		2,601
N50 contig length (bases)		1,168
Longest contig (bases)		4477
Average contig size (bases)		1211
Total size of contigs (bases)		3,152,134
GC content		41.17%
Best k-mer (bases)		35
Repetitive elements characterization	Number of elements	Total length (bases)
Retroelements	115	69,319
LINE elements		
L1-CIN4	20	5240
LTR retrotransposons		
Ty1- <i>Copia</i>	61	48,132
Gypsy-DIRS1	28	13,407
Others	6	2540
DNA transposons		
hobo-Activator	1	153
Others	8	624
Unclassified elements	1	105
Small RNA	5	596
Simple repeats	328	11,305
Low complexity	70	2943
Total interspersed repeats	—	70,201
Gene prediction		
Number of predicted genes	1856	
Total length of the predicted genes (bases)	2,258,812	
Average length of the predicted genes (bases)	1068.34	
Number of exons	2354	

^aAssuming a genome size of 244.99 Mb (Costa et al. 2008)

corresponding to a coverage of about 0.03×, on average, of the whole genome estimated through flow cytometry for this species (Costa et al. 2008).

The de novo assembling performed using SPAdes resulted in 3,152,134 bases of sequence in 2601 contigs larger than 1000 bases. The largest contig had 4477 bases and the N50 for the assembly was 1168 (Table 1).

Repetitive elements characterized in *E. uniflora*

About 2.60% (81,855 bases) of the assembled WGSAS was characterized as repetitive sequences, including retroelements, DNA-transposons, simple repeats, small RNAs, low complexity elements and one unclassified element (Table 1).

Simple repeats were the most abundant category of repetitive elements, followed by retroelements. As expected, LTR retrotransposons were the most frequent retroelements (Macas et al. 2015).

Gene prediction and functional annotation

The BLAST hits of the 2601 contigs discovered 964 sequences presenting high similarity to sequences of *Eucalyptus grandis* available in the NCBI databank (Supplementary Fig. 1), suggesting the retrieval of genes conserved within Myrtaceae.

Gene prediction (Table 1) based on the homology searching approach yielded 1856 putative genes,

including 1492 complete and 364 partial gene sequences, ranging from 201 to 4411 bases in length. The total length of the predicted genes reached 2,258,812 bases, i.e., 71.66% of the assembled WGSAS with a mean of one gene at every 1697 bases. Predicted genes are distributed into biological process, cellular components and molecular function terms (Fig. 1, Supplementary Fig. 2).

Scanning into the InterPro database, 1248 sequences (contigs) matched to deposited protein sequences, 534 from which matching terms from the Gene Ontology (GO) database (Supplementary Fig. 3A). After merging Blastn and InterPro searches, a total of 1798 sequences received GO annotation, 937 from which had confirmed InterPro GOs and 85 were classified as too general (Supplementary Fig. 3B). Concerning the enzymatic functions, most of the predicted genes are classified as transferases and hydrolases (Supplementary Fig. 3C). The map distribution of GOs across contigs revealed most sequences with two GO terms, while some sequences received more than a dozen GO terms (Supplementary Fig. 3D). The search within the KEGG database returned 147 metabolic pathways related to enzymes characterized in the *E. uniflora* WGSAS, revealing a wide range of metabolic functional represented, despite the low coverage of the genome.

Discussion

In this study, we assume that the analyzed sequences are random segments across the genome. However, we cannot exclude the possibility that certain regions of the genome were more or less likely to be represented due to features such as GC content. In fact, the gene density observed in this WGSAS (one gene every 1697 bases) is likely biased, in comparison to the gene density of *Eucalyptus grandis* (one gene every 16,669) and *Metrosideros polymorpha* (one gene every 8836 bases), although the GC content is quite similar among these species (Izuno et al. 2016). Nonetheless, we provide a useful starting point for understanding the dynamic of repetitive elements in *E. uniflora* and further information for closer characterization of genes of biotechnological relevance in this species.

The differential accumulation/exclusion of repetitive DNA is a key driver of genome size variation in flowering plants, as demonstrated for the genome size differences across 23 species of Fabaceae (Macas et al. 2015). Based on the length of the microRNAs identified in *E. uniflora*, Guzman et al. (2012) suggested a possible genome complexity of this species, with a high content of repetitive sequences. Conversely, our WGSAS is composed by around 2.60% of repetitive elements, a proportion coherent with the relatively small genome of *E. uniflora*. Repetitive sequences are speculated to influence cytoplasmic, cellular and developmental processes not only by changing genome size, but also by

Fig. 1 Summary of the main 20 gene ontology categories predicted to the WGSAS of *E. uniflora*, characterized into biological process [BP], molecular function [MF], or cellular component [CC] terms

affecting chromosomal recombination. Simple repeats, the most frequent repetitive element in this WGSAS, represent, in turn, recombination “hotspots” of genome reorganization (Mehrotra and Goyal 2014). Given the high frequency and usual hypervariability of tandem simple sequence repeats (SSRs) in plant genomes, high-throughput sequencing has been largely used as a tool for the characterization of SSR genetic markers of non-model species (e.g., Staton et al. 2015; Lemos et al. 2018). In a former study, a total of 111 di- and tri-nucleotides SSR regions were identified in *E. uniflora* using the same WGSAS of this study (Sarzi et al. 2019).

LTR retrotransposons, the second more frequent repetitive element, presented a 10-fold higher frequency than DNA transposons in *E. uniflora*. This result probably is not artefact of the low genome coverage since independent of the genome size; retrotransposons have been more frequently noticed than DNA transposons (Won et al. 2018). In addition, this higher frequency of retrotransposons is coherent with their mode of transposition: copy-and-paste for retrotransposons and cut-and-paste for DNA-transposons.

Lineage-specific accumulation/exclusion of transposons is a critical element in genome evolution. *Gypsy* and *Copia*, the two major superfamilies of LTR retrotransposons, are known to be differentially disseminated across many species and serve as a driving force of plant genome evolution. Genome sequencing of plant species has demonstrated that usually *Gypsy* superfamily is more highly amplified than *Copia* (Barghini et al. 2014). In *E. uniflora*, an opposite trend was observed with a ratio between *Copia* and *Gypsy* frequencies of 2.28, revealing a particular dynamic of repetitive elements distribution in this species. The larger number of *Copia* retrotransposons in comparison to *Gypsy* suggests that some families of the former element have experienced larger amplification during the evolution of the *E. uniflora* genome. A larger number of *Copia* elements in relation to *Gypsy* were also observed in *Conyza canadensis*, *Cynara cardunculus* (Won et al. 2018) and *Vitis vinifera* (Jaillon et al. 2007), among other species, without clear reason found for such pattern.

Plastid genome may also be represented as high-copy DNA when sequencing total DNA from leaves. However, the amount of plastid DNA depends on the developmental stage of the leaf and on species-specific factors (Scott and Possingham 1983). No meaningful amount of plastid sequences was retrieved in our WGSAS, likely because young leaves (the sample material used for DNA isolation) contain less plastid DNA (Rauwolf et al. 2010).

Opposite to the high probability of finding repetitive genetic elements in whole-genome sequencing, mining low-copy nuclear genes is a much more challenging task. Using a transcriptomic approach, Guzman et al. (2014) reported 72,742 unigenes, with 8625 genes assigned to 141 metabolic pathways. Our WGSAS retrieved 1856 putative

genes, assigned to 147 metabolic pathways. Although most of these genes are identified as constitutively expressed and related to essentially basal functions of the cell, our annotation revealed the existence of coding sequences associated with processes of antioxidant activity, plant defense, and response to stress. This occurred because our genomic approach is able to identify all genes, including that expressed only under particular environmental conditions, while the transcriptome only identified expressed genes on *E. uniflora* leaves.

The analysis of the antioxidative potential of *E. uniflora* ethanolic extract has pointed out phenylpropanoids, phenols, and flavonoids as the main active compounds (Cunha et al. 2016b). Enzymes of flavonoid and phenylpropanoid biosynthesis pathways were identified among the genes predicted in this study. Since the active compounds may have common synthetic pathways, these enzymes may be appealing for a closer understanding of the synthesis of the compounds with pharmaceutical potential.

In addition, proteins related to antifungal activity (ginkbilobin-2-like), defense response to bacterium and fungi (PR-4-like and EMSY-LIKE 3 proteins) were identified. Plant pathogens have adapted numerous ways to overcome host immune systems while the hosts respond with new defenses. This constant conflict against pathogens has led to a basal response to pathogen-associated molecular patterns, and to a gene-for-gene response specific to a pathogen (Grennan 2006). Consequently, the response of *E. uniflora* to bacterial and fungal attacks may have such specificity and transferring it to other species through genetic transformation is an appealing biotechnological advance for cultivated species. Considering that genetic engineering has proven a promising tool for the development of stress tolerance in plants, this partial genome sequencing of *E. uniflora* reveals a still unexplored biotechnological potential of this species directly related to agriculture. The genes of *E. uniflora* with antibacterial/antifungal function can be of high importance for the gene transfer of such resistance into cultivated plants of agro-nomical/horticultural interest.

Finally, transcriptions factors such as the *MYB98* and the *JUB1* were also identified. These transcription factors presented 72% and 88% similarity to protein sequences of *Eucalyptus grandis*, respectively. The *JUB1* transcription factor plays an important role in abiotic stress responses and may be related to the high potential of *E. uniflora* for responding to environmental pressures, contributing to the wide distribution of the species in different ecosystems. Closer examining this transcription factor in *E. uniflora* may expand our understanding of the species-environment interaction at the gene expression level. This information also contributes to the set of microRNAs reported by Guzman et al. (2012), who identified 1,852,722 sequences from 45

conserved microRNA families, including novel microRNAs whose targets are transcription factors.

Conclusions

In this study, the repeat elements and the basic composition of gene families in *E. uniflora* were characterized. However, two main limitations have to be considered. First, the sequences obtained do not fully cover the entire structure of each retroelement and transposon. Second, the low coverage of the genome may provide an incomplete overview of the number of the major gene families in *E. uniflora*. Therefore, a broader genome sequencing of the species is recommended in order to overpass these constraints.

Nonetheless, the partial genome sequencing of *E. uniflora* is an important tool for identifying genetic control of key biotechnological and horticultural traits. Despite the very high number of genes involved in metabolic pathways, a close characterization of the *E. uniflora* genes may allow their use in genetic transformation of plants, towards the cultivation of species of agronomical importance. In addition, the results presented here may support a wide application of marker-assisted and genomic selection, promoting the establishment of commercial orchards with improved cultivars of *E. uniflora*. These cultivars may have differential production of compounds of biotechnological interest into the pharmaceutical and agronomical sectors.

Acknowledgements We would like to thank CAPES for scholarship to D.S.S., CNPq/Brazil by financial support (Grant number 442995/2014-8) and Research Fellowship provided to V.M.S. (Process 302501/2017-7) and Universidade Federal do Pampa—UNIPAMPA (Edital PROPESQ/AGP/2016). We specially thank to MSc. Rafael Plá Matielo Lemos, Dr. Filipe de Carvalho Victoria and Dr. Victor Pyrlo by technical support in data analysis.

Compliance with ethical standards

Conflict of interest All authors hereby declare that there is no conflict of interest.

Ethical approval This article does not include any studies with human participants or animals performed by any of the authors.

Informed consent This article does not involve any informed consent.

References

- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477
- Barghini E, Natali L, Cossu RM, Giordani T, Pindo M, Cattonaro F, Scalabrin S, Velasco R, Morgante M, Cavallini A (2014) The Peculiar Landscape of Repetitive Sequences in the Olive (*Olea europaea* L.) Genome. *Genome Biol Evol* 6:776–791
- Chikhi R, Medvedev P (2014) Informed and automated K-Mer size selection for genome assembly. *Bioinformatics* 30:31–37
- Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Obles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21:3674–3676
- Coradin L, Siminski A, Reis A (2011) Espécies nativas da flora brasileira de valor econômico atual ou potencial: plantas para o futuro – Região Sul. MMA, Brasília
- Costa IR, Dornelas MC, Forni-Martins EL (2008) Nuclear genome size variation in fleshy-fruited Neotropical Myrtaceae. *Plant Syst Evol* 276:209–217
- Costa DP, Filho EGA, Silva LMA, Santos SC, Passos XS, Silva MRR, Seraphin JC, Ferri PH (2010) Influence of fruits biotypes on the chemical composition and antifungal activity of the essential oils of *Eugenia uniflora* leaves. *J Braz Chem Soc* 21:851–858
- Cunha FAB, Wallau GL, Pinho AI, Nunes MEM, Leite NF, Tintino SR, Da Costa GM, Athayde ML, Boligon AA, Coutinho HDM, Pereira AB, Posser T, Franco JL (2015) *Eugenia uniflora* leaves essential oil induces toxicity in *Drosophila melanogaster*: involvement of oxidative stress mechanisms. *Toxicol Res* 4:634–644
- Cunha FAB, Pinho AI, Santos JFS, Sobral-Souza CE, Albuquerque RS, Matias EFF, Leite NF, Tintino SR, Costa JGM, Boligon AA, Waczuk EP, Rocha JBT, Posser T, Coutinho HDM, Quintans-Júnior LJ, Franco JL (2016a) Cytoprotective effect of *Eugenia uniflora* L. against the waste contaminant mercury chloride. *Arab J Chem X*:1–10
- Cunha FAB, Waczuk EP, Duarte AE, Barros LM, Elekofehinti OO, Matias EFF, da Costa JGM, Sanmi AA, Boligon AA, Da Rocha JBT, Souza DO, Posser T, Coutinho HDM, Franco JL, Kamdem JP (2016b) Cytotoxic and antioxidative potentials of ethanolic extract of *Eugenia uniflora* L. (Myrtaceae) leaves on human blood cells. *Biomed Pharmacot* 84:614–621
- Ferreira-Ramos R, Laborda PR, de Oliveira Santos M, Mayor MS, Mestriner MA, de Souza AP, Alzate-Marin AL (2008) Genetic analysis of forest species *Eugenia uniflora* L. through of newly developed SSR markers. *Conserv Genet* 9:1281
- Ferreira-Ramos R, Accoroni KAG, Rossi A, Guidugli MC, Mestriner MA, Martinez CA, Alzate-Marin AL (2014) Genetic diversity assessment for *Eugenia uniflora* L., *E. pyriformis* Cambess., *E. brasiliensis* Lam. and *E. francavilleana* O. Berg neotropical tree species (Myrtaceae) with heterologous SSR markers. *Genet Resour Crop Evol* 61:267
- Grennan AK (2006) Plant response to bacterial pathogens. Overlap between Innate and gene-for-gene defense response. *Plant Physiol* 142:809–811
- Guzman F, Almerão MP, Körbes AP, Loss-Morais G, Margis R (2012) Identification of MicroRNAs from *Eugenia uniflora* by high-throughput sequencing and bioinformatics analysis. *PLoS One* 7:e49811
- Guzman F, Kulcheski FR, Turchetto-Zolet AC, Margis R (2014) De novo assembly of *Eugenia uniflora* L. transcriptome and identification of genes from the terpenoid biosynthesis pathway. *Plant Sci* 229:238–246
- Izuno A, Hatakeyama M, Nishiyama T, Tamaki I, Shimizu-Inatsugi R, Sasaki R, Shimizu KK, Isagi Y (2016) Genome sequencing of *Metrosideros polymorpha* (Myrtaceae), a dominant species in various habitats in the Hawaiian Islands with remarkable phenotypic variations. *J Plant Res* 129:727–736
- Jaillon O et al (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449:463–467

- Lemos RPM, Matielo CBDO, Beise DC, Rosa VG, Sarzi DS, Roesch LFW, Stefenon VM (2018) Characterization of ptSSR, nSSR and EST-SSR markers for understanding invasion histories and genetic diversity of *Schinus molle* L. *Biology* 7:43
- Macas J, Novák P, Pellicer J, Čížková J, Koblížková A, Neumann P, Fuková I, Doležel J, Kelly LJ, Leitch IJ (2015) In depth characterization of repetitive DNA in 23 plant genomes reveals sources of genome size variation in the legume tribe fabaeae. *PLoS One* 10:e0143424
- Margis R, Felix D, Caldas J, Salgueiro F, de Araujo DSD, Breyne P, van Montagu M, de Oliveira D, Margis-Pinheiro M (2002) Genetic differentiation among three neighboring Brazil-cherry (*Eugenia uniflora* L.) populations within the Brazilian Atlantic rain forest. *Biodivers Conserv* 11:149–163
- Mehrotra S, Goyal V (2014) Repetitive Sequences in Plant Nuclear DNA: types, distribution, evolution and function. *Genom Proteom Bioinf* 12:164–171
- Rauwolf U, Golczyk H, Greiner S, Herrmann RG (2010) Variable amounts of DNA related to the size of chloroplasts III. Biochemical determinations of DNA amounts per organelle. *Mol Genet Genom* 283:35–47
- Salgueiro F, Felix D, Caldas JF, Margis-Pinheiro M, Margis R (2004) Even population differentiation for maternal and biparental gene markers in *Eugenia uniflora*, a widely distributed species from the Brazilian coastal Atlantic rain forest. *Biodivers Distrib* 10:201–2010
- Sarzi DS, Justolin B, Silva C, Lemos RPM, Stefenon VM (2019) Discovery and characterization of SSR markers in *Eugenia uniflora* L. (Myrtaceae) using low coverage genome sequencing. *An Acad Bras Cienc* 91:e20180420
- Scott NS, Possingham JV (1983) Changes in chloroplast DNA levels during growth of spinach leaves. *J Exp Bot* 34:1756–1767
- Stanke M, Morgenstern B (2005) AUGUSTUS: a web server for gene prediction in eukaryotes that allows user defined constraints. *Nucleic Acid Res* 33:W465–W467
- Staton M, Best T, Khodwekar S, Owusu S, Xu T, Xu Y, Jennings T, Cronn R, Arumuganathan AK, Coggeshall M et al (2015) Preliminary genomic characterization of ten hardwood tree species from multiplexed low coverage whole genome sequencing. *PLoS One* 10:e0145031
- Won SY, Hwang Y-J, Jung J-A, Kin JS, Kang S-H, Sohn S-H (2018) Identification of repetitive DNA sequences in the *Chrysanthemum boreale* genome. *Sci Hortic* 236:238–243