



ELSEVIER

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib



Data Article

Dataset of the transcribed 45S ribosomal RNA sequence of the tree crop “yerba mate”



Patricia M. Aguilera^{a,b,*}, Humberto J. Debat^c,
Mauro Grabiele^{a,b}

^a Instituto de Biología Subtropical, Universidad Nacional de Misiones (CONICET), 3300 Posadas, Misiones, Argentina

^b Instituto de Biotecnología de Misiones, Universidad Nacional de Misiones, 3300 Posadas, Misiones, Argentina

^c Instituto de Patología Vegetal, Centro de Investigaciones Agropecuarias (INTA), 5000 Córdoba, Argentina

ARTICLE INFO

Article history:

Received 21 February 2017

Received in revised form

4 April 2017

Accepted 24 April 2017

Available online 6 May 2017

Keywords:

Ilex

Transcriptomics

Assembly

Annotation

RNA unit

ABSTRACT

This contribution contains data related to the research article entitled “The 18S–25S ribosomal RNA unit of yerba mate (*Ilex paraguariensis* A. St.-Hil.)” (Aguilera et al., 2016) [1]. Through a bioinformatic approach involving NGS data, we provide information of the transcribed 45S ribosomal RNA (rRNA) sequence of yerba mate, the first reference for the *Ilex* L. genus. This dataset (Supplementary file 1) comprises information regarding the assembly and annotation of this rRNA unit. The generated data is applicable for comparative analysis and evolutionary studies among *Ilex* and related taxa. The raw sequencing data used here is available at DDBJ/EMBL/GenBank (NCBI Resource Coordinators, 2016) [2] Sequence Read Archive (SRA) under the accession SRP043293 and the consensus 45S ribosomal RNA sequence has been deposited there under the accession GFHV00000000.

© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

* Corresponding author at: Instituto de Biología Subtropical, Universidad Nacional de Misiones (CONICET), 3300 Posadas, Misiones, Argentina.

E-mail address: patriciamaguilera@gmail.com (P.M. Aguilera).

Specifications Table

Subject area	Biology
More specific subject area	Evolutionary biology
Type of data	Transcriptomics, assembly of reads and sequence annotation
How data was acquired	Bioinformatics approach
Data format	Analyzed
Experimental factors	RNA, used for library construction and Illumina sequencing, was isolated from leaf samples at emerging, young, fully expanded, and early and late senescent stages from <i>I. paraguariensis</i> breeding line Pg538
Experimental features	Paired-end 100 nt Raw reads were filtered, <i>de novo</i> assembled and submitted to homology-based searches and annotation
Data source location	Misiones, Argentina
Data accessibility	Data are within this article and at DDBJ/EMBL/GenBank under the accessions SRP043293 and GFHV00000000

Value of the data

- This data provides the first reference sequence of the transcribed 45S rRNA unit in *Ilex* L.
- Data is applicable for comparative analysis and evolutionary studies among *Ilex* and related taxa based on the 18S, 5.8S and 25S rRNA genes and ITS and ETS sequences.
- Accessibility of assembly and annotation data allows researchers to perform further analysis via novel approaches.

1. Data, experimental design, materials and methods

Total RNA extracted of five samples of emerging, young, fully expanded, and early and late senescent stages leaves of *Ilex paraguariensis* breeding line Pg538 were pooled for high throughput sequencing [1].

The attainment of the transcribed 45S rRNA sequence of yerba mate was completed in seven steps:

1. The complete raw sequencing data at SRA under the accession SRP043293 was used to generate a full transcriptome assembly employing the Trinity 2.0.6 platform. All raw sequenced reads were quality filtered and then *de novo* assembled into contigs with optimal parameters of 25 kmer word and group pairs distance of 500.
2. The achieved complete list of 44,907 contigs was subsequently scanned by in-house [3; v.8.1.8] homology searches with BLASTN (word size 11, cut off value of 1e-10) using as baits the conserved 18S, 5.8S and 25S rRNA entire gene regions of *Helianthus annuus* 45S (KF767534). Sunflower was selected, as it is the closest taxon to yerba mate in Euasterids II clade [4] from which complete rRNA sequence information is available yet.
3. Three blast hits were obtained and identified as contigs comp17895_c0_seq. 1 (2936 bp), comp17895_c1_seq. 1 (1053 bp) and comp17901_c0_seq. 2 (2994 bp), which assemble in a 6961 bp sequence (Supplementary Figure 1, BAM file).
4. This sequence was further aligned to the sunflower rRNA reference sequence regions [3] following the methods of Geneious global alignment with free end gaps (93%, gap open penalty 12, gap extension penalty 3) and progressive Mauve algorithm [5] at default values. The alignments were manually checked previous to the transference of homology-based annotations among sunflower and yerba mate.

5. The annotated sequence which in turns spans the complete 45S rRNA unit of yerba mate, shares a consistent pairwise similarity of 54.2%, 97.6%, 54.3%, 96.2%, 66.4%, 95.7% and 49.7% at the 5'ETS, 18S, ITS1, 5.8S, ITS2, 25S and 3'ETS, respectively, with sunflower.
6. This first annotation step of the yerba mate rRNA sequence was further curated by subsequent homology searches [3] in non-redundant NCBI *Ilex* databases by BLASTN (word size 11, cut off value of 1e-10). **Supplementary Figures 2–6 (BAM files)** illustrate this step for each ribosomal RNA region with blast hits, and the accession of hits and coverage are provided.
7. A final yerba mate annotated ribosomal RNA sequence embracing genes (18S, 5.8S, 25S) and spacers (ITS1, ITS2, 5'ETS, 3'ETS) was attained by last integration of sunflower and *Ilex* annotated features. **Supplementary Figure 7 (GFF file)** illustrates this final step.

This Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank [2] under the accession GFHV00000000. The version described in this paper is the first version, GFHV01000000.

Acknowledgements

This study was funded by the Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT-Argentina), Grant no. UNaM PICT 2014-3328 Préstamo BID N° AR-L 1181.

Transparency document. Supporting information

Transparency data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.dib.2017.04.044>.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.dib.2017.04.044>.

References

- [1] P.M. Aguilera, M. Grabile, H.J. Debat, R.E. Bubillo, D.A. Marti, The 18S-25S ribosomal RNA unit of yerba mate (*Ilex paraguariensis* A. St.-Hil.), *Plant Biosyst.* 150 (2016) 1240–1248.
- [2] NCBI Resource Coordinators, Database resources of the National Center for Biotechnology Information, *Nucl. Acids Res.* 44 (2016) D7–D19.
- [3] M. Kearse, R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, et al., Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data, *Bioinformatics* 28 (2012) 1647–1649.
- [4] The Angiosperm Phylogeny Group, An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants, *Bot. J. Linn. Soc.* 161 (2009) 105–121.
- [5] A.C.E. Darling, B. Mau, F.R. Blattner, N.T. Perna, Mauve: multiple alignment of conserved genomic sequence with rearrangements, *Genome Res.* 14 (2004) 1394–1403.