

Publisher: Taylor & Francis

Journal: *Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology*

DOI: <http://dx.doi.org/10.1080/11263504.2015.1018982>

The 18S-25S ribosomal RNA unit of yerba mate (*Ilex paraguariensis* A. St.-Hil.)

P. M. AGUILERA^{1,2}, M. GRABIELE^{1,2}, H. J. DEBAT³, R. E. BUBILLO⁴, & D. A. MARTÍ^{1,2}

¹Instituto de Biología Subtropical, Universidad Nacional de Misiones (IBS-UNaM-CONICET), Félix de Azara 1552, Posadas (3300), Misiones, Argentina. ²Instituto de Biotecnología de Misiones, Universidad Nacional de Misiones (INBIOMIS-FCEQyN-UNaM), Ruta 12 Km 7.5, Posadas (3300), Misiones, Argentina. ³Instituto de Patología Vegetal, Instituto Nacional de Tecnología Agropecuaria (IPAVE-CIAP-INTA), Camino a 60 Cuadras Km 5 ½, Córdoba (X5020ICA), Argentina. ⁴Estación Experimental Agropecuaria Cerro Azul, Instituto Nacional de Tecnología Agropecuaria (EEA Cerro Azul-INTA), Ruta 14 Km 1085, Cerro Azul (3313), Misiones, Argentina.

Running title: 18S-25S rRNA unit of yerba mate

Corresponding author: P. M. Aguilera. Tel/Fax: +54-03764-425414/422186. E-mail: patriciamaguilera@gmail.com

Accepted Manuscript

Abstract

We present the first report of the transcribed 18S-25S rDNA unit of the South-American crop *Ilex paraguariensis*. This tree is consumed as an infusion named “mate”. NGS data was assembled, reference-annotated and further characterized. The assembled consensus sequence of the complete transcribed rDNA unit is 6961 nt, with a coverage of ~18,646X and 99.3% of global sequence similarity. The transcribed 18S and 25S rDNA genes of yerba mate are the first in the genus to be characterized, combined to the 5.8S that is identical to the consensus sequence of *Ilex*. The ITS1 and the ITS2 display species-specific regions, useful to differentiate yerba mate from adulterant taxa of the manufactured product. A replication slippage process seems significant in the evolution of the ITS in *Ilex*. The 5ETS consists in SR1, that encloses a large inverted repeat/hairpin region, and SR2 which is a high homology region in the Asterids clade and embrace six distinctive inverted repeat/hairpin segments. The 3ETS presents a pyrimidine-rich block associated to transcription termination, mirror-like palindromes and repetitive segments, all networking at the secondary structure level. This information will be helpful in leading future studies in the molecular evolution characterization of this crop and related species of *Ilex*.

Key words: Aquifoliaceae, *Ilex*, Next-generation sequencing, Molecular characterization, rDNA

Introduction

Yerba mate (*Ilex paraguariensis*) is an important dioecious crop tree native to the subtropical forest of Argentina, Brazil and Paraguay, where it is widely cultivated on 330,000 ha with a total yield production fairly exceeding the 1,000,000 t (Jerke et al. 2009; INYM 2014). This evergreen holly is mainly consumed as a nutritional and stimulant beverage named “mate”, a hot infusion made from dried milled leaves of *I. paraguariensis*. The “mate” has a profound socio-cultural influence in Latin America, and particularly in Argentina where its regular intake is superior to that of coffee (OPSM 2009). Recently, through next-generation sequencing (NGS) Debat et al. (2014) developed the first genomic resource of yerba mate, characterizing the gene repertoire and generating a vast amount of information for further studies that attempt to reveal the molecular profile of this significant crop.

The ribosomal RNA (rRNA) is essential in protein synthesis and a particularly useful tool in addressing questions ranging from transcription regulation to phylogenetics. In higher eukaryotes, the ribosomal DNA (rDNA) is located at the nucleolar organizer regions (NORs) and composed of tandem arrays of genes (18S, 5.8S, 25S) separated by adjacent spacers (ITS, internal transcribed spacer; IGS, intergenic spacer). Moreover, the external transcribed spacer (ETS) is the transcribed region of the IGS flanking the 18S-25S unit. The rDNA locus comprise several hundreds to thousands copies. This abundance is probably associated to the absence of secondary rounds of amplification via translation, contrasting to protein coding genes (Prokopowich et al. 2003). In addition, the IGS contains different regulatory genetic and epigenetic features that influence the regulation of genes and there is a variable population of active rRNA according to the physiological needs of the cell (Tucker et al. 2010; Grabiele et al. 2012). The rDNA genes, transcribed by RNA polymerase I, are highly conserved throughout eukaryotes, but the spacer sequences are more dynamic. The IGS is variable in both length and organization even among closely related taxa and diverges more rapidly than the ITS, which is sufficient to resolve species relationships within most genera (Matyasek et al. 2012).

Here we report the first analysis of the 18S-25S rRNA unit of yerba mate, assembled from NGS data, annotated and further characterized at the primary and secondary structure levels, in order to aid in the molecular characterization of this important crop and related species of *Ilex* L.

Material and Methods

Illumina reads from the NCBI Short Read Archive (SRA) database under Accession no. SRP043293 (Debat et al. 2014) were relaxedly mapped to a *Helianthus annuus* L. ribosomal sequence (NCBI Accession no. KF767534), employing the map to reference utility of the Geneious 7.0 software (Geneious assembler, medium sensitivity, iterations up to 5 times). The mapped reads were *de novo* assembled into a contig representing the complete 45S rRNA unit. This contig was further corrected by iterative flexible mapping of total Illumina reads to the assembled rRNA unit, and a consensus sequence was generated with the 18S-25S rRNA unit of

yerba mate and aligned to the *H. annuus* ribosomal complete sequence using Geneious global alignment with free end gaps (93%, gap open penalty 12, gap extension penalty 3).

The yerba mate transcribed 18S-25S rRNA unit was annotated integrating the *H. annuus* ribosomal reference sequence to subsequent homology searches by NCBI-BLASTN against Euasterids II and *Ilex* databases, in addition to gene and spacers predictions using *Arabidopsis thaliana* (L.) Heynh (Unfried et al. 1989; Unfried & Gruendler 1990) and *Capsicum pubescens* Ruiz & Pav. (Grabiele et al. 2012) as reference sequences.

Geneious 7.0 software was used in the pairwise alignment, editing and characterization of nucleotide sequences. Diverse 18S-25S rDNA regions of different angiosperms used for comparison were obtained from the NCBI database and their accession numbers are provided in the Supplemental Figures 2-6. MEGA 4.0 software (Tamura et al. 2007) was used to obtain genetic distances in the ITS of *Ilex* (Tamura-Nei method, 500 replicates bootstrap, pairwise deletion option). SplitsTree v.4.10 (Huson & Bryant 2006) at default values was used in the Network construction of the ITS of southern South-American *Ilex*. Secondary structure analyses in the ITS and ETS regions were performed using UNAFold (Markham & Zuker 2008) adjusting the parameters for RNA (2.3) and 24 °C.

Results and Discussion

Ribosomal unit assembly and annotation

To generate an assembly of the transcribed 18S-25S rDNA unit of yerba mate, Illumina reads from the NCBI SRA database under Accession no. SRP043293 (Debat et al. 2014) were relaxedly mapped to a *H. annuus* ribosomal sequence (NCBI Accession no. KF767534), as it constitutes the most closely related taxa within Euasterids II clade, to which *I. paraguariensis* is a part of (APG III 2009), as reference. By iterative mapping of total Illumina reads, an assembled consensus sequence of 6961 ungrouped nucleotids (nt) in length and 55.0% GC content was generated. A summary of the assembly report of the transcribed 18S-25S rDNA unit of yerba mate is shown in Table I. A total of 1,297,998 mapped 100 nt reads that encompass 129,799,800 nt account to a ~18,646X coverage of the yerba mate transcribed rDNA unit which correlates with the reached repetitiveness of this particular element in plant genomes (Prokopowich et al. 2003). An unexpected number of reads derived from the 18S rDNA gene and probably biased the GC richness of the assembly (Table I). In addition, the mapped reads assembly displays a high pairwise similarity of 99.3% (Table I).

The yerba mate transcribed 18S-25S rDNA unit was annotated integrating the *H. annuus* ribosomal reference sequence to subsequent homology searches by NCBI-BLASTN against Euasterids II and *Ilex* databases, in addition to gene and spacers predictions using *A. thaliana* and *C. pubescens* as reference sequences. The primary structure characterization by means of referenced annotation of rDNA genes (18S, 5.8S, 25S) and spacers (ETS, external transcribed spacer; ITS, internal transcribed spacer) is shown in Figure 1A. The assembly reads coverage and pairwise identity are illustrated in Figure 1B and Figure S1, where a nucleotide diversity correlated positively with the coverage density is observed.

Characterization of ribosomal genes

The transcribed 18S of yerba mate is 1809 nt in length and display 49.7% in GC content. This is the first 18S gene to be completely characterized in *Ilex* sharing a 99.9% identity with the current largest sequenced DNA fragment (*I. opaca* Aiton; 1745 nt; NCBI Accession no. AF161010). Homology searches against Euasterids II clade in the NCBI database obtained 6697 blast-hits and a pairwise identity of 98.2% comparing the majority consensus sequence (1810 nt and 49.5% GC) to the 18S of yerba mate (Figure S4). The pairwise identity of the 18S assembly of yerba mate reached 99.3% with a low GC content value (29.7%), unlike that of the annotated unit (49.7%), probably caused by the biased number and origin of the mapped reads (Figure S1).

The transcribed 5.8S of yerba mate has 159 nt and is 53.5% GC rich. By homology searches against NCBI *Ilex* clade, 413 blast-hits were mapped reaching a within-group pairwise identity of 96.5% (Figure S5). Meanwhile the 5.8S of yerba mate is 100% identical to the majority consensus sequence of the genus. The pairwise identity of the 5.8S assembly of yerba mate extended to 99.2% and its 50.0% GC content is fairly representative to that of the annotated sequence.

The transcribed 25S of yerba mate is 3354 nt in length and 57.2% GC rich. This gene is the first to be completely characterized at RNA level in *Ilex* since the report of Soltis et al. (2003) on the complete genomic sequence of *I. opaca*, which shares an identity of 98.9% with the transcribed 25S of yerba mate. Homology searches against Euasterids II clade in the NCBI database targeted 2767 blast-hits with a within-group pairwise identity of 98.5% (Figure S6). A pairwise identity of 97.0% is found between the majority consensus sequence of Euasterids II (3349 nt and 57.2% GC) and the 25S of yerba mate. The pairwise identity of the 25S assembly of yerba mate is 98.7% with a GC richness (54.5%) similar to that of the annotated sequence.

Characterization of ribosomal spacers

The ITS1 of yerba mate is 269 nt in length and 62.1% GC rich. Pairwise alignment with the majority consensus sequence of *Ilex*, obtained with 376 blast-hits from NCBI *Ilex* database (overall mean distance of 0.113 ± 0.011), reveals a similarity of 87.7% additionally to specific regions in yerba mate (Figure 2A). Furthermore, the ubiquities of poly nucleotide stretches also pinpoint the significance of the replication slippage process in the evolution of the ITS1 region in the genus (Figure 2A). The pairwise identity of the ITS1 assembly of yerba mate is 96.7% with a lower GC content (51.9%) if compared to that of the annotated sequence. Just eleven *Ilex* species coexist with *I. paraguariensis* in southern South America (Gottlieb et al. 2005), most of them sympatrically and *I. argentina* Lillo in the NW region of Argentina (Giberti 1999). Yet, all of them have been historically recognized as adulterants or substitutes of the manufactured product of the yerba mate, predominantly *I. dumosa* Reissek (Filip et al. 1998; Barchuk et al. 2013). When comparing the nucleotide diversity among the ITS1 of yerba mate and related genomic regions of *I. paraguariensis* and adulterant *Ilex* species at NCBI (39 sequences), regardless their associated-to-functionality features, an overall mean distance of 0.107 ± 0.011 is found (Figure S2), similar to that of *Ilex*. In addition, their pairwise alignment uncovers the presence of structural regions useful to differentiate yerba mate from those adulterant taxa (Figure S2). Further network analysis considering the southern South-American taxa revealed an expected clustering among the transcribed ITS1 of yerba mate and related genomic regions of *I.*

paraguariensis, forming a major cluster with those genomic variants of *I. dumosa* (Figure 2B), this latter a probable illegitimate clade according to Gottlieb et al. (2005) precise criteria and results. In addition, secondary structure folding of the transcribed ITS1 of yerba mate revealed four major stable helices (I-IV) regularly predicted at Gibbs free energy (dG) values ranging from -104.7 to -110.4 (Figure 2C). At positions 169 to 189, embracing the helix III and a part of the contiguous loop, the conserved motif for plants ITS1 described by Liu and Schradl (1994) was pinpointed (Figure 2C).

The ITS2 of yerba mate has 265 nt in length and is 60.0% GC rich. Pairwise alignment with the majority consensus sequence of *Ilex*, obtained with 387 blast-hits from NCBI *Ilex* database (overall mean distance of 0.111 ± 0.010), shows an identity of 86.4% additionally to particular regions in yerba mate (Figure 3A). As in the ITS1, replication slippage is key in the evolution of the ITS2 region in *Ilex* (Figure 3A). The pairwise identity of the ITS2 assembly of yerba mate is 97.7% with a lower GC richness (54.2%) than that of the annotated sequence. By comparing the nucleotide diversity between the ITS2 of yerba mate and related genomic regions of *I. paraguariensis* and adulterant *Ilex* species at NCBI (41 sequences), irrespective to functional features, an overall mean distance of 0.109 ± 0.013 was found (Figure S3), similar to that of the genus. As in the ITS1, the ITS2 pairwise alignment also reveals structural regions suitable to discriminate yerba mate from those adulterant taxa (Figure S3). Network analysis of the ITS2 of southern South-American taxa revealed an anticipated clustering among the transcribed ITS2 of yerba mate and related genomic regions of *I. paraguariensis*, except for a variant, which is strikingly absent from the yerba mate mapped reads pool, that associated to *I. argentina* NCBI Accession no. AJ492655 (Figure 3B). This latter is a functional ITS unit (Gottlieb et al. 2005). RNA folding analysis of the transcribed ITS2 of yerba mate revealed a deep conservation of all structural features, including the four major stable helices (I-IV), compared to the proposed eukaryote ITS2 model of Schultz et al. (2005) (Figure 3C).

The 5-external transcribed spacer (5ETS) of yerba mate is 909 nt in length and 54% GC rich and can be further classified into two distinct structural regions (SRs). The SR1 is a short 5' fragment of 335 nt which shares no apparent similarity to Viridiplantae. However, this particular region encloses a large inverted repeat of 297 nt able to form a potential stem-loop structure (Figure 4A). Related regions of other members of the Asterids clade (APG III 2009), as *H. annuus* (sunflower; NCBI Accession no. KF767534) and the SRVI of the intergenic spacer of *C. pubescens* (locoto; Grabiele et al. 2012) also show comparable secondary structures despite differences at the primary level (Figure 4A). In those taxa, the transcription initiation site (TIS) was characterized at DNA level, ~750 nt in sunflower and ~180 nt in locoto upstream the start of the inverted repeat regions, respectively. The contiguous SR2, encompass 574 nt and constitutes a high homology region (HHR) regarding Asterids clade, with 56.5% of similarity between yerba mate, sunflower and locoto. The HHR of those taxa possesses six distinctive inverted repeat segments each (I-VI), able to form possible stem-loop structures (Figure 4B), as previously described for locoto (Grabiele et al. 2012). Comparable to the transcribed 5ETS of yerba mate, secondary structures have been found in the 5ETS of several eukaryotes, associated to rRNA-processing and probably to the regulation of rRNA genes (Gerbi & Borovjagin 2004; Pendrak & Roberts 2011; Grabiele et al. 2012 and references therein).

The 3-external transcribed spacer (3ETS) of yerba mate is 175 nt in length and 56.0% GC rich and presents no apparent similarity to Viridiplantae. At the 5' end of the 3ETS there is a typical pyrimidine-rich block (Figure 5A), a presumptive transcription termination site (TTS) usually present in higher plants, that shares this status with several distinct motives (Grabiele et al. 2012; Inácio et al. 2014). In addition, the transcribed 3ETS of yerba mate displays several particular structural features, i.e. mirror-like palindromes and repetitive segments (Figure 5A, B). RNA secondary folding demonstrates a potential stem-loop structure at a dG value of -39.9, in which, the interaction between the diverse 3ETS segments, including the palindromic and repetitive ones, is clear (Figure 5C). Though there is no evidence of folding implications, the 3ETS sequence is key in the 3' end formation of the 25S gene during rRNA-processing (Gerbi & Borovjagin 2004).

Conclusions

The transcribed 18S-25S rDNA unit of the crop tree “yerba mate” was analyzed by NGS data assembly, referenced-annotation and further structural characterization. This experimental approach allowed us, for the first time in the literature, to be able to characterize the intra-individual variability of the expressed transcripts of these important genes. The complete transcribed rDNA unit is 6961 nt and 55.0% GC-rich, with an assembly coverage of ~18,646X and 99.3% of similarity. The transcribed 18S and 25S rDNA genes of yerba mate are the first characterized in *Ilex*, and the 5.8S is identical to the *Ilex* consensus. The ITS1 (269 nt) and the ITS2 (265 nt) display specific regions, several useful to differentiate yerba mate from adulterant taxa of the manufactured product. Replication slippage process is significant in the evolution of the ITS in *Ilex*. Secondary structure folding analysis revealed conserved features in both ITS. The 5ETS (909 nt) is further classified in two structural regions (SRs). SR1 (335 nt), with narrow similarity to Viridiplantae, encloses a large inverted repeat/hairpin region. The contiguous SR2 (574 nt) is a high homology region (HHR) in Asterids clade and embraces six distinctive inverted repeat/hairpin segments. The 3ETS (175 nt), with limited similarity to Viridiplantae, presents the typical pyrimidine-rich block associated to transcription termination, added to particular mirror-like palindromes and repetitive segments, all networking at the secondary structure level. The information reported here aids in the molecular characterization of this important crop and related species of *Ilex* and introduces the possibility of exploiting NGS data in deciphering the dynamic nature of these interesting genes.

Acknowledgments

This work was supported by the Instituto Nacional de la Yerba Mate (INYM-Argentina) under Grant PRASY 58/14. M.G. and D.A.M. are career members and P.M.A. is a postdoctoral research fellow of the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET-Argentina). H.J.D. and R.E.B. are researchers at the Instituto Nacional de Tecnología Agropecuaria (INTA-Argentina).

Supplemental data

Supplemental data for this article can be accessed here.

[<https://drive.google.com/folderview?id=0B3H16DDYmVW6Q2d2MHRQdUJFRTQ&usp=sharing>]

Accepted Manuscript

References

APG III. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot J Linn Soc* 161: 105–121.

Barchuk ML, Tiscornia MM, Giorgio EM, Fonseca MI, Zapata PD. 2013. Diseño de un método molecular para la detección de *Ilex dumosa* en yerba mate elaborada utilizando una secuencia específica ubicada en la región ITS2 del DNA ribosómico. *Revista de Ciencia y Tecnología (UNaM)* 19: 28–34.

Debat HJ, Grabiele M, Aguilera PM, Bubillo RE, Otegui MB, Ducasse DA, et al. 2014. Exploring the genes of yerba mate (*Ilex paraguariensis* A. St. Hil.) by NGS and *de novo* transcriptome assembly. *PLoS ONE* 9(10): e109835. doi:10.1371/journal.pone.0109835.

Filip R, Lopez P, Coussio J, Ferraro G. 1998. Mate substitutes or adulterants: study of xanthine content. *Phytother Res* 12: 129–131.

Gerbi SA, Borovjagin AV. 2004. Pre-ribosomal RNA processing in multicellular organisms. In: Olson MOJ, editor. *The Nucleolus*. New York, NY, USA: Kluwer Academic/Plenum Publishers. pp. 170-173.

Giberti GC. 1999. Recursos fitogenéticos relacionados con el cultivo y explotación de la Yerba Mate (*Ilex paraguariensis* St. Hil., Aquifoliáceas) en el cono sur de América. *Acta Horti* 500: 137–144.

Gottlieb AM, Giberti G, 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.

Grabiele M, Debat HJ, Moscone EA, Ducasse DA. 25S-18S rDNA IGS of *Capsicum*: molecular structure and comparison. *Plant Syst Evol* 298(2): 313–321.

Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 23(2): 254–267.

Inácio V, Rocheta M, Morais-Cecílio L. 2014. Molecular organization of the 25S-18S rDNA IGS of *Fagus sylvatica* and *Quercus suber*: a comparative analysis. *PLoS ONE* 9(6): e98678. doi:10.1371/journal.pone.0098678.

INYM 2014. Instituto Nacional de la Yerba Mate (Argentina) website. Available: <http://www.inym.org.ar>. Accessed Jun 2014 4.

Jerke G, Horiński MA, Salvatierra KA. 2009. Evaluación de géneros micotoxigénicos en yerba mate elaborada. *Revista de Ciencia y Tecnología (UNaM)* 12: 41–45.

Liu JS, Schardl CL. 1994. A conserved sequence in internal transcribed spacer 1 of plant nuclear rRNA genes. *Plant Mol Biol* 26: 775–778.

Markham NR, Zuker M. 2008. UNAFold: software for nucleic acid folding and hybridization. In: Keith JM, editor. Bioinformatics, Volume II Structure, Functions and Applications, number 453 in Methods in Molecular Biology, chapter 1. Totowa, NJ: Humana Press. Pp. 3–31.

Matyasek R, Renny-Byfield S, Fulneček J, Macas J, Grandbastien MA, Nichols R, et al. 2012. Next generation sequencing analysis reveals a relationship between rDNA unit diversity and locus number in *Nicotiana* diploids. *BMC Genomics* 13: 722.

OPSM 2009. Yerba mate Argentina website. Available: <http://yerbamateargentina.org.ar/>. Accessed Jun 2014 4.

Pendrak ML, Roberts DD. 2011. Ribosomal RNA processing in *Candida albicans*. *RNA* 17: 2235–2248.

Prokopowich CD, Gregory TR, Crease TJ. 2003 The correlation between rDNA copy number and genome size in eukaryotes. *Genome* 46: 48–50.

Schultz J, Maisel S, Gerlach D, Müller T, Wolf M. 2005. A common core of secondary structure of the internal transcribed spacer 2 (ITS2) throughout the Eukaryota. *RNA* 11: 361–364.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA). Software Version 4.0. *Mol Biol Evol* 24(8): 1596–1599.

Tucker S, Vitins A, Pikaard C. 2010. Nucleolar dominance and ribosomal RNA gene silencing. *Curr Opin Cell Biol* 22: 351–356.

Unfried I, Stocker U, Gruendler P. 1989. Nucleotide sequence of the 18S rRNA gene from *Arabidopsis thaliana* Col0. *Nucleic Acids Res* 17(18): 7513.

Unfried I, Gruendler P. 1990. Nucleotide sequence of the 5.8S and 25S rRNA genes and of the internal transcribed spacers from *Arabidopsis thaliana*. *Nucleic Acids Res* 18(13): 4011.

Table 1. Assembly report of the transcribed 18S-25S rDNA unit of yerba mate.

rDNA	45S	5ETS	18S	ITS1	5.8S	ITS2	25S	3ETS
Mapped reads	1,297,998	113,962	1,015,060	6736	1034	4075	164,095	57
RPKM	27,456.9	17,979.6	76,081.7	3476.1	1040.3	2387.9	7147.1	51.8
Coverage mean	18,646.7	12,537.1	56,111.7	2504.1	650.3	1537.7	4862.1	32.6
Coverage Min.	10	3	1	1	1	1	1	1
Coverage Max.	321,062	37,362	321,062	6526	965	2870	57,328	34
Q30	93.9%	92.8%	95.3%	79.3%	95.1%	90.5%	85.9%	90.4%
Nucleotide length	7816	1014	2211	310	159	273	3673	176
Ungapped nucleotide length	6961	909	1809	269	159	265	3375	175
Pairwise Identity	99.3%	99.0%	99.3%	96.7%	99.2%	97.7%	98.7%	98.0%
Identical Sites	7.7%	0.7%	2.0%	20.6%	27.7%	21.6%	6.5%	83.0%
GC content	33.5%	47.4%	29.7%	51.9%	50.0%	54.2%	54.5%	54.5%
Freq. A	24.1%	16.3%	26.1%	14.6%	26.3%	22.6%	18.0%	17.0%
Freq. T	25.5%	29.3%	26.6%	15.2%	23.7%	21.2%	16.0%	28.2%
Freq. G	27.5%	32.0%	25.5%	32.4%	26.9%	32.4%	36.6%	20.7%
Freq. C	22.9%	22.5%	21.8%	37.8%	23.1%	23.8%	29.5%	34.1%

Figure Legends

Figure 1. Transcribed 18S-25S rDNA unit of yerba mate: A. Primary structure characterization by means of referenced annotation of rDNA genes (18S, 5.8S, 25S) and spacers (ETS, external transcribed spacer; ITS, internal transcribed spacer) and degree of GC content. B. Assembly reads coverage and pairwise identity. Note that nucleotide diversity correlates positively with coverage density.

Figure 2. Characterization of the ITS1 of yerba mate: A. Pairwise alignment with majority consensus sequence of *Ilex* reveals a similarity of 87.7% additionally to specific regions in yerba mate. Note the importance of replication slippage in the evolution of the ITS1 region in *Ilex*. B. Network analysis considering yerba mate and related regions of *I. paraguariensis* (blue) and adulterant *Ilex* species; only the cluster that includes yerba mate is captioned. C. Secondary structure folding of the ITS1 of yerba mate with 269 nucleotides in length and a dG of -107.9. Four major stable helices (I-IV) are regularly recognized at dG values ranging from -104.7 and 110.4. Liu & Schradl (1994) conserved motif for plants ITS1 is marked in blue.

Figure 3. Characterization of the ITS2 of yerba mate: A. Pairwise alignment with majority consensus sequence of *Ilex* reveals a similarity of 86.4% and several specific regions in yerba mate; particularly, region comprising positions 216 to 234 is a useful marker to distinguish yerba mate from those adulterant *Ilex* species. Note the importance of replication slippage in the evolution of the ITS2 region in *Ilex*. B. Network analysis considering yerba mate and related regions of *I. paraguariensis* (blue) and adulterant *Ilex* species; only the cluster that includes yerba mate is captioned. C. Secondary structure folding of the ITS2 of yerba mate with 265 nucleotides in length and a dG of -106.4. All structural features, including the four major stable helices (I-IV), are in agreement with the proposed eukaryote ITS2 model (Schultz et al. 2005).

Figure 4. Characterization of the 5ETS of yerba mate: A. A short 5' fragment of 335 nt presents no apparent similarity to Viridiplantae. This region encloses a large inverted repeat able to form a potential secondary structure as in related coordinates of others members of the Asterids clade (sunflower and locoto) used for comparison. B. Downstream that zone there is a 574 nt high homology region (HHR) in Asterids, with 56.5% similarity between yerba mate, sunflower and locoto. The HHR of those taxa encompass six distinct inverted repeat regions each (I-VI) able to form possible stem-loops structures. Negative numbers refers to dG values of the folded structure.

Figure 5. Characterization of the 3ETS of yerba mate: A. Identification of structural features, i.e. the universal pyrimidine-rich block, mirror-like palindromes and repetitive segments (Rep). The fragment of 175 nt presents no apparent similarity to Viridiplantae. B. Pairwise alignment of the distinct Rep segments identified. C. Potential stem-loop structure at a dG of -39.9. Note the interactions between the diverse segments.

Supplemental Figure Legends

Figure S1. Assembly reads coverage and pairwise identity in the transcribed 18S-25S rDNA unit of yerba mate. Primary structure was characterized by means of referenced annotation of rDNA genes (18S, 5.8S, 25S) and spacers (ETS, external transcribed spacer; ITS, internal transcribed spacer).

Figure S2. Blast hits alignment of the transcribed ITS1 of yerba mate vs. related genomic regions of *I. paraguariensis* and adulterants *Ilex* species from NCBI. The overall mean distance is 0.107 ± 0.011 . Note that the poly-C and poly-G blocks at positions 133 to 148 in the pairwise alignment are useful markers to distinguish yerba mate from the other taxa.

Figure S3. Blast hits alignment of the transcribed ITS2 of yerba mate vs. related genomic regions of *I. paraguariensis* and adulterants *Ilex* species from NCBI. The overall mean distance is 0.109 ± 0.013 . Note that distinct segments including that at positions 216 to 225 in the pairwise alignment, that enclose the poly-purine block, are useful markers to distinguish yerba mate from the other taxa.

Figure S4. Blast hits alignment of the transcribed 18S of yerba mate vs. 6697 related regions of Euasterids II clade from NCBI. The within-group pairwise identity is 98.2%.

Figure S5. Blast hits alignment of the transcribed 5.8S of yerba mate vs. 413 related regions of *Ilex* clade from NCBI. The within-group pairwise identity is 96.5%.

Figure S6. Blast hits alignment of the transcribed 25S of yerba mate vs. 2767 related regions of Euasterids II clade from NCBI. The within-group pairwise identity is 98.5%.