Ann. For. Res. 63(1): 3-18, 2020
DOI: 10.15287/afr.2019.1413

ANNALS OF FOREST RESEARCH
www.afrjournal.org

Structural, evolutionary and phylogenomic features of the plastid genome of *Carya illinoinensis* cv. Imperial

Jordana Caroline Nagel[1,2], Lilian de Oliveira Machado[3], Rafael Plá Matielo Lemos[1], Cristiane Barbosa D'Oliveira Matielo[1], Tales Poletto[4], Igor Poletto[1], Valdir Marcos Stefenon[3,1]

Nagel J.C., de Oliveira Machado, L., Lemos R.P.M., Barbosa D'Oliveira Matielo C., Poletto T., Poletto I., Stefenon V.M., 2020. Structural, evolutionary and phylogenomic features of the plastid genome of *Carya illinoinensis* cv. Imperial. Ann. For. Res. 63(1): 3-18.

Abstract. The economically most important nut tree species in the world belong to family Juglandaceae, tribe Jungladeae. Evolutive investigations concerning species from this tribe are important for understanding the molecular basis driving the evolution and systematics of these species. In this study, we release the complete plastid genome of *C. illinoinensis* cv. Imperial. Using an IonTorrent NGS platform we generated 8.5 x 10^8 bp of raw sequences, enabling the assemblage of the complete plastid genome of this species. The plastid genome is 160,818 bp long, having a quadripartite structure with an LSC of 90,041bp, an SSC of 18,791 bp and twoIRs of 25,993 bp. A total of 78 protein-coding, 37 tRNA-coding, and 8 rRNA-coding regions were predicted. Bias in synonymous codon usage was detected in cultivar Imperial and three tRNA-coding regions were identified as hotspots of nucleotide divergence, with high estimations of dN/dS ratio. The high fraction of SSR loci prospected in non-coding regions may provide informative genetic markers, useful to a wide range of genetic researches. Despite the significant structural differences among plastid genomes, the phylogenetic relationships among species is supported by the whole plastid genome analysis, supporting the monophyly of subtribes Caryinae and Juglandinae within family Juglandaceae.
Keywords: Pecan, chloroplast genome, phylogenomics, Juglandaceae

Authors. [1]Federal University of the Pampa, Graduate Program in Biological Sciences, Campus São Gabriel, São Gabriel, RS, Brazil | [2]Universidade Regional Integrada do Alto Uruguai e das Missões, Campus Santo Ângelo, Santo Ângelo, RS, Brazil | [3]Federal University of Santa Catarina, Graduate Program in Plant Genetic Resources, Florianópolis, SC, Brazil | [4]Federal University of Santa Maria, Graduate Program in Forest Engineering; Santa Maria, RS, Brazil.

§ Corresponding author: Valdir Marcos Stefenon (valdirstefenon@gmail.com)
Manuscript received November 3, 2019; revised February 10, 2020; accepted February 18, 2020; online first March 2nd, 2020.

Introduction

In most plant species, the plastid DNA shows maternal inheritance, low recombination frequency, and a relaxed evolution rate. These characteristics make plastid genomes substantially distinctive from the nuclear genomes (Wolfe et al. 1987) and very useful in a wide range of studies about evolutionary relationships in plants. The principal function of the plastid is to carry out photosynthesis, but other major cellular functions also occur in this organelle, including synthesis of starch, fatty acids, pigments and amino acids (Wicke et al. 2011).

The known plastid genomes of green plants typically contain from 100 to 120 genes, of which approximately 70-88 are protein-coding genes, 33-35 are tRNAs and four are rRNAs (Wicke et al. 2011). With the progress of the next generation sequencing (NGS) technologies, complete plastid genomes have been sequenced for several plant species, generating a wide set of genomic resources, which enable the development of molecular markers and more refined taxonomic and evolutionary studies. The development of genetic markers from plastid genome sequences has significantly contributed to studies about gene flow (plastid SSR markers; Perdereau et al. 2014), phylogeography (SNPs and Indels in plastid genes; Stefenon et al. 2019a), and hybridization/introgression (plastid SSR markers; Curtu et al. 2007) in forest tree species. Moreover, identification of genome rearrangements becomes increasingly important, especially as high levels of rearrangement have been observed among both eukaryotes and prokaryotes (Wicke et al. 2011). Such rearrangements may also be very useful to understand phylogenetic and evolutionary trends within and among plant groups. Knowledge about genome rearrangements, gene content, recombination events, loss of genes, and gene transfer to the nucleus are of great importance for understanding evolutionary events in plants (Vieira et al. 2016, Bock 2017, Lopes et al. 2017).

Carya illinoinensis (Wangenh) K. Koch, popularly known as pecan (Figure 1a-c, Supporting Information), is a deciduous tree species of the tribe Junglandeae (Juglandaceae family), native to the temperate zones of North America. Several species of tribe Juglandeae - as *Juglans regia* (walnut), *Cyclocarya palliurus* (wheel wingnut), *Carya sinensis* (Chinese hickory) and *Carya illinoinensis* (hickory or pecan) - are important crop tree species cultivated in several regions of the world aiming at the commercial production of nuts. Pecan is one of the most important nut crop species in the world, cultivated in several countries in North America (USA and Mexico), South America (Peru, Argentina, and Brazil), Africa (South Africa and Egypt), Asia (Israel and China) and Australia (Poletto et al. 2018). The species was introduced in Brazil with commercial interest during the 1870s and farmers needed to select cultivars adapted to the climatic and biological conditions where the orchards were established (Poletto et al. 2015). Currently, more than 40 cultivars are commercially planted in large areas of Southern Brazil.Despite the economic significance of *Carya* as a crop tree species and the importance of plastid genomesfor comparative evolutionary analysis and molecular taxonomy, only the plastid genomes of *Carya sinensis* (Hu et al. 2016)*, Carya kweichowensis* (Yeh et al. 2018), and *Carya cathayensis* (Zhai et al. 2019) were published to date, while one unpublished sequence of the plastid genome of *Carya illinoinensis* is deposited in the NCBI database (Genbank ID MH909599.1). Besides, few molecular studies have been performed concerning *C. illinoinensis* cultivars planted in southern Brazil (e.g Poletto et al. 2019). Considering the scarcity of genomic studies available for *Carya* species and aiming to generate novel genomic resources for *Carya illinoinensis*, we sequenced, assembled and characterized the complete plastid genome of *C. illinoinensis* cv. Imperial using next-generation sequencing

(NGS) technology. Here, we report the main finds obtained from this initiative concerning the genes present and the codon usage bias in the plastid genome of this species. In addition, we identified potentially polymorphic plastid SSR markers and also performed a comparative evolutionary analysis of plastid genomes between *C. illinoinensis* cv. Imperial and other tree species of the tribe Juglandeae using publicly available genomic datasets, thus revealing some features of pecan evolution.

Materials and methods

Plastid genome sequencing and assembling

Healthy leaves of an adult individual of *C. illinoinensis* cv. Imperial were sampled in the municipality of Anta Gorda, Rio Grande do Sul State, southern Brazil. A voucher of this sample was deposited under the number HBEI1624 in the Bruno Edgar Irgang Herbarium of the Federal University of the Pampa, Brazil. Intact chloroplasts were isolated from the leaves as described by Matielo et al. (2019) and the plastid DNA (cpDNA) was isolated using the CTAB method (Doyle and Doyle 1987). The quality of the isolated DNA was checked using a NanoVue™ spectrophotometer (GE Healthcare). The isolated cpDNA was used for library preparation with Ion OneTouch2™ System using the Ion PGM™ Template OT2 400 Kit. The sequencing was performed using Ion PGM™ Sequencing 400 kit on the Ion PGM™ System with an Ion 318™ Chip v2. Raw sequence data were deposited in the NCBI Sequence Read Archive (SRA) database under number SSR10382885, Bioproject PRJNA587009, Biosample SAMN13174479.

The plastid genome of *C. illinoinensis* cv. Imperial was assembled using a reference-guided approach with the plastid genome of *C. illinoinensis* (NC041449.1) as a reference, in the CLC Genomics Workbench software. The mean coverage of the sequencing was determined as the total size of the sequenced reads divided by the size of the assembled plastid genome.Annotation of the plastid genome was conducted using the GeSeq (Tillich et al 2017) and the cpGAVAS (Liu et al. 2012) platforms. For GeSeq, annotation started from four references chloroplast annotations (*Arabidopsis thaliana, Castanea mollissima, Juglans nigra* and *J. regia*). The software tRNAscan (Chan and Lowe 2019) and Aragorn (Laslett and Camback 2004) were used for searching the tRNAs, while the physical circular map of the plastid genome was built using Organellar Genome DRAW software (Lohse et al. 2013).

Characterization of plastid genome features in *C. illinoinensis* cv. Imperial

Relative synonymous codon usage (RSCU) of all protein-coding genes were determined using MEGA 6.0 software (Tamura et al. 2013). RSCU corresponds to the proportion of the observed occurrence of a codon to its expected occurrence if all the synonymous codons of a particular amino acid are used evenly.
Prospection of simple sequence repeats (SSRs) loci in the plastid genome of *C. illinoinensis* cv. Imperial was performed using the Perl script MISA (Beier et al. 2017) setting minimum thresholds for search at ten for mononucleotide repeats, six for dinucleotide repeats and five for tri-, tetra-, penta- and hexanucleotide repeats. The location of the SSR loci within the plastid genome was determined using the cpGAVAS (Liu et al. 2012) platform.

Evolutive and phylogenetic relationships of *C. illinoinensis* cv. Imperial within Juglandeae

Different comparisons were performed among *Carya illinoinensis* cv. Imperial and 11 species of Juglandeae tribe (*Carya illioinensis* (Wangenh) K.Koch*, Carya kweichowensis* Kuang & A.M. Lu*, Carya sinensis* Dode*, Cyclocarya paliurus* (Batal.) Iljinsk.*, Juglans cinerea* L.*, Juglans regia* L.*, Juglans major* (Torr.)

A. Heller, *Juglans sigillata* Dode, *Juglans hopeiensis* Hu, *Juglans mandshurica* Maxim., and *Juglans cathayensis* Dode), whose plastid genome sequences were downloaded from the Genbank database. Species nomenclature in this study follows the International Plant Names Index (www.ipni.org). An initial pairwise comparison of the gene order of plastid genomes was performed through a dot plot analysis using the software Mafft online service (Katoh et al. 2017). Translocations, inversions or indels occurring in a set of genes are visualized in the pairwise comparison as displacements of the positive and/or negative slopes representing the LSC, SSC (positive slope) and IR (negative slope) regions of the plastid genomes compared. Boundaries of the IRa, IRb, SSC and LSC regions (IR/SSC and IR/LSC boundaries) and sizes of each region were determined using the online platform IRscope (Amiryousefi et al. 2018). Aiming to compare the number and type of SSR loci in all species, these ubiquitous regions were prospected in the plastid genome sequences of all other 11 species, as described above for *C. illinoinensis* cv. Imperial.

A phylogenomic analysis was performed using the complete plastid genome sequences of *C. illinoinensis* cv. Imperial and more 11 species of tribe Juglandeae. The plastid genome sequence of *Castanea mollissima* Blume (Fagaceae) was used as outgroup. The Genbank IDs of all downloaded sequences are given in Figure 3c. The 13 sequences were aligned using the software Mafft online service (Katoh et al. 2017) and the phylogenetic tree was constructed with the Neighbor-Joining algorithm. Support of the analysis was determined through 500 bootstrap replicates using the same software.

Rearrangements and inversions among Juglandeae plastid genomes were visualized using the default parameters of the Multiple Genome Alignment software MAUVE 2.4.0 (Darling 2004). Hotspots of sequence divergence were determined using the sliding windows analysis with the complete plastid genomes of 11 species of tribe Juglandeae with a window length of 400 bp and step size of 100 bp, using the DnaSP v.5 software (Librado and Rozas 2009). *Carya kweichowensis* was excluded from this analysis giving the significative difference in length of the plastid genome in comparison to all other species.

The ratio of the nonsynonymous (*dN*) tosynonymous (*dS*) substitutions - *dN/dS*ratio - of the genomic regions identified as hotspots of divergence was performed to investigate the patterns of selection occurring within Juglandeae. The complete regions containing CDSs, introns and intergenic spacers corresponding to each hotspot were extracted from all species analyzed in this study and aligned individually using the MUSCLE algorithm (Edgar 2004) as implemented in MEGA 6.0 (Tamura et al. 2013), with pairwise deletion set to gaps/missing data treatment. To each alignment, the *dS* and *dN* values were calculated using MEGA under the Kimura 2-parameters model. The pairwise and the overall mean distances were estimated for each hotspot region using the same software.

Results

General features of theplastid genome of *C. illinoinensis* cv. Imperial

The sequencing effort generated a total of 5,639,849 raw reads, representing 852,882,019 assembled nucleotides, which corresponds to a mean plastid genome coverage of 5,303.4×.

The plastid genome of *Carya illinoinensis* cv. Imperial (GenBank ID MN221384) presents 160,818 bp in length and the traditional quadripartite structure of plastid genomes (Figure 1d, Supporting Information). The large single-copy (LSC) region is composed of 90,041 bp with a GC content of 33.73%, and the small single-copy (SSC) region contains18,791 bp with a GC content of 29.89%. The LSC and the SSC regions are separated by two inverted repeat regions (IRs) of 25,993 bp each, with

42.58% of GC content. The overall GC content of this plastid genome was 36.14%. A total of 123 genes, including 78 protein-coding genes, 37 tRNA-coding genes, and eight rRNA-coding genes were predicted.

A total of seven protein-coding genes, seven tRNA-coding genes, and four rRNA-coding genes are located within each IR region of the chloroplast genome of *C. illinoinensis* cv. Imperial (Table 1). The SSC region presents 12 protein-coding genes and one tRNA-coding gene, while the LSC region presents 61 protein-coding genes and 22 tRNA-coding regions (Table 1). Eighteen genes presented introns, while the *rps12* gene is trans-spliced, with the 5′-end located in the LSC and the 3′-end duplicated in the IR regions. The tRNA-coding gene *trnE-UUC* has two copies and the *trnM-CAU* has three copies.

Comparative analysis among chloroplast genomes of Juglandeae species

Cultivar Imperial presents a plastid genome-with one base deletion and 21 single nucleotide polymorphisms (SNPs) in comparison to the plastid genome of *Carya illinoinensis* MH909599.1 deposited in the Genbank database. Concerning the other species of the family Juglandaceae, the size of the plastid genome of cultivar imperial is also very similar (Figure 1a-b), except to *Carya kweichowensis*. This species has a plastid genome 14,495 bases longer than the plastid genome of *C. illinoinensis* cv. Imperial (Table 2). The IR regions of *C. illinoinensis* cv. Imperial were only 65 to 34 bp shorter than the IR regions of *Carya sinensis* and *Cyclocarya* respectively.

Concerning the plastid genomes of *Juglans* species included in this study, *C. illinoinensis* cv. Imperial revealed a difference in the length of the IR regions ranging from 31 bp to 227 bp (Table 2, Figure 1b). Excluding *C. kweichowensis,* the coefficient of variation of the plastid genome size (total plastid genome, LSC, SSC, and IRs) ranged from 0.25% to 0.88% (Table 2). On the other hand, each IR region of *C. kweichowensis* is 40,943 bp in

Table 1 Distribution of the coding regions within the plastid genome of *C. illinoinensis* cv. Imperial, according to the plastid region

Region	Category	Genes
IRs[1]	Protein-coding	*ndhB*, rpl2*, rpl23, rps12*[3], rps7, ycf1[2],* and *ycf2*
	tRNA-coding	*trnA-UGC*, trnM-CAU, trnI-GAU*, trnL-CAA, trnE-UUC*, trnR-ACG* and *trnV-GAC*
	rRNA-coding	*rrn16, rrn23, rrn4.5* and *rrn5*
SSC	Protein-coding	*ndhF, rps32, ndhD, ndhE, ndhG, ndhI, psaC, ccsA, ndhA*, ndhH, rps15* and *ycf1*[2]
	tRNA-coding	*trnL-UAG*
LSC	Protein-coding	*rps19, rpl22, rps3, rpl16, rpl14, rpsB, infA, rpl36, rps11, rpoA, petA, petD, petB, petG, petL, pbf1, psbH, psbT, psbB, psbE, psbJ, psbL, clpP*, rps12*[3], rpl20, rps18, rpl33, psaJ, psaI, cemA, ycf4, accD, rbcL, atpB, atpE, ndhC, ndhK, ndhJ, rps4, ycf3*, psaA, psaB, rps14, psbZ, psbC, psbD, psbM, petN, rpoB, rpoC1*, rpoC2, rps2, atpI, atpH, atpF*, atpA, psbI, psbK, rps16*, matK, psbA*
	tRNA-coding	*trnP-UGG, trnW-CCA, trnM-CAU[4], trnV-UAC*, trnF-GAA, trnL-UAA*, trnS-GGA, trnT-UGU, trnG-GCC, trnS-UGA, trnT-GGU, trnE-UUC*, trnY-GUA, trnD-GUC, trnC-GCA*, trnR-UCU, trnS-CGA, trnS-GCU, trnQ-UUG, trnK-UUU*, trnH-GUG*

Note. [1]All coding regions within the IR regions are duplicated, [2]A large part of the ycf1 gene is located within the SSC, [3]Trans-spliced gene, [4]This gene has two copies within the LSC region, *Coding regions containing introns.

Table 2 Size of plastid genomes of Juglandaceae species in base pairs (bp)

Species	Total length	LSC	SSC	IR
Carya illinoinensis cv. Imperial	160,818	90,041	18,791	25,993
Carya illinoinensis MH909599.1	160,819	90,042	18,791	25,993
Carya kweichowensis	175,313	89,858	3,569	40,943
Carya sinensis	160,195	89,541	18,538	26,058
Cyclocarya paliurus	160,562	90,007	18,477	26,039
Juglans cathayensis	159,730	89,333	18,351	26,023
Juglans mandshurica	159,729	89,845	18,352	25,766
Juglans hopeiensis	159,714	89,316	18,352	26,023
Juglans sigiallata	160,351	89,871	18,412	26,034
Juglans regia	160,367	89,872	18,423	26,036
Juglans major	160,276	89,829	18,397	26,025
Juglans cinerea	160,288	89,803	18,417	26,034
Coefficient of variation (%)*	0.25	0.29	0.88	0.31

Note. *The coefficient of variation was computed excluding *Carya kweichowensis.*

length, that is, 1.6-fold larger than each IR region of *C. illinoinensis* (Figure 2b), while the SSC region of *C. kweichowensis* is only 3,569 bp in length, i.e., about 6-fold shorter than the SSC of *C. illinoinensis*.

The IRa/LSC border is located between genes *rpl2* (in the IRa region) and *trnH* (in the LSC region) in all species (Figure 2b). In *J. mandshurica,*181 bp of the *rpl2* gene are positioned within the LSC region. The IRa/SSC border is filled by the *ycf1* gene in all species except *Carya kweichowensis*, in which this gene is located internally in the IR regions, distant to the borders with the SSC and LSC regions. The genes *rps19* in the LSC region and *rpl2* in the IRb region define the IRb/LSC border in all species. In *J. mandshurica*, 82 bp of the *rpl2* gene are inside the LSC region. A comparatively long *ndhF* gene is found in the border IRb/SSC in *Carya* and *Cyclocarya.* This gene is much shorter in *Juglans*. A segment of the *ycf1* gene is found in the IRb/SSC border in *C. illinoinensis*, *C. sinensis, Cyclocarya, J. cathayensis,J. mandshurica, J. sinensis, J. paliurus, J. hopeiensis, J. sigillata* and *J. regia* (Figure 2b).

Detailed comparisons of the IR/SSC and IR/LSC junction sites among species of Juglandeae (Figure 2b) show little variation.Variations in IR boundarieswere mainly observed in *C. kweichowensis,* and some substantial differences in the size of the *ndhF*gene in species of *Juglans* in comparison to *C. illinoinensis, C. sinensis,*and *Cyclocarya*.

The synonymous codons in angiosperms genomes retain different usage frequencies, i.e., codon usage biases. In *C. illinoinensis* cv. Imperial, codon usage bias was revealed, with a high proportion of synonymous codons presenting the nucleotides A or U in the third position (Figure 2a). This pattern is consistent with most of the plastid genomes and may be linked firstly to the high A/T content of the plastid genome (A/T = 63.86%), although selective pressure cannot be discarded (see results of the sliding window analysis below). Codons CUG(L), CAC(H), GAC(D), CGC(R) and GGC(G) can be considered under-represented (RSCU < 0.6), while codons UUA(L) and AGA(R) are over-represented (RSCU > 1.6, Barbhuiya et al. 2019).

Twenty-nine codons presented RSCU values higher than 1.0, meaning they are being used more often than the expected. On the other hand, 33 codons are being used less frequently than expected, presenting a RSCU < 1.0. As expected, codons AUG(M) and UGG(W), the unique codons for methionine and tryptophan respectively, presented an RSCU = 1.0 (Figure 2a). The relative synonymous codon usage (RSCU) of the plastid genome of *C. illinoinensis* cv. Imperial revealed a higher proportion

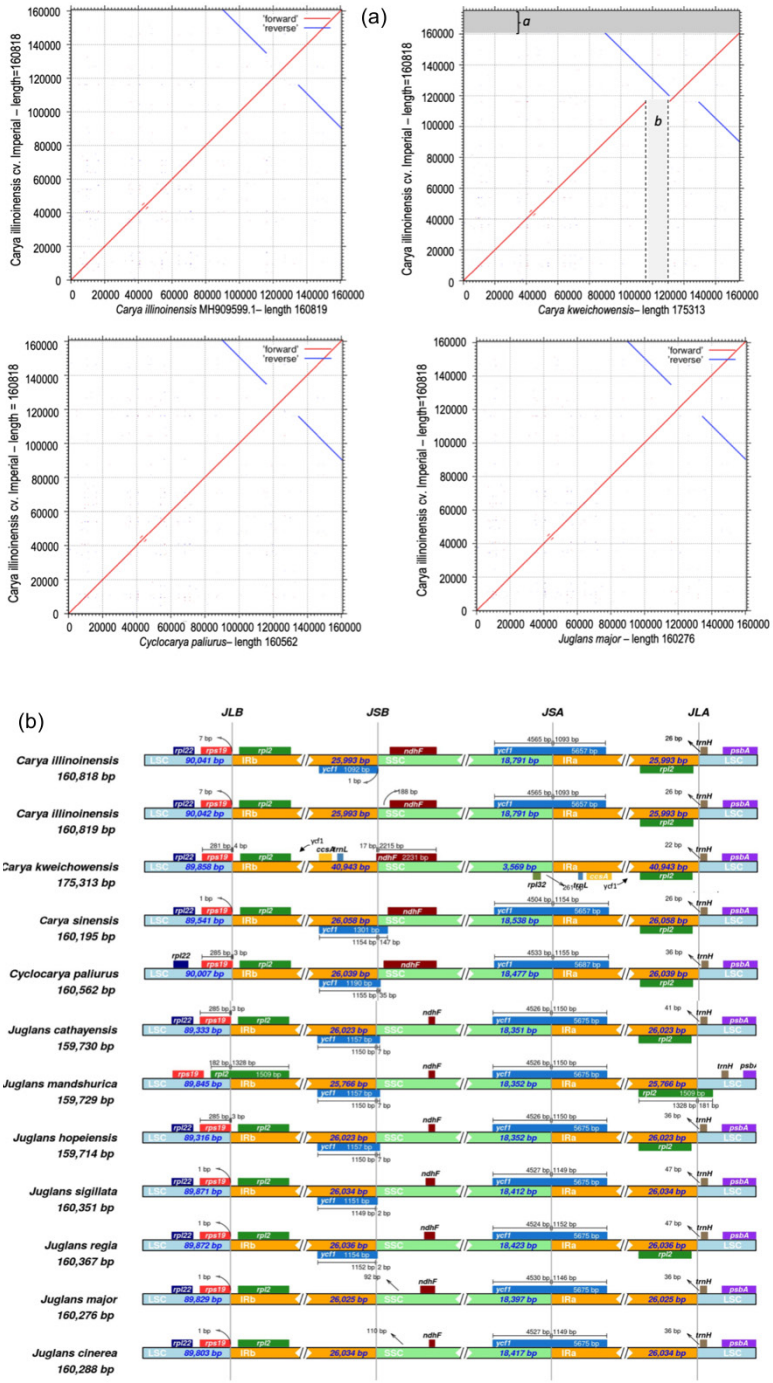


Figure 1

Relationship among plastid genomes of *Carya illinoinensis* cv. Imperial and species of Juglandae. (a) Dot plot analyses of the plastid genome of *C. illinoinensis* cv. Imperial against three Juglandaceae species. The positive red slope means that the pair of sequences compared is in the same orientation. The negative blue slope denotes that the pair of sequences compared can be aligned, but their orientation is opposite. In the dot plot against *C. weichowensis*, the gray region "a" highlights the 14,495 bp without correspondent sequence in *C. illinoinensis* cv. Imperial (as well as in the other species of Juglandaceae). In the same dot plot, region "b" highlights a gap, which means that no similarity was identified between this region, present in the plastid genome of *C. weichowensis*, and the plastid of *C. illinoinensis* cv. Imperial. (b) Boundaries at the junctions of the LSC, IRa, IRb, and SSC regions of the plastid genomes of *Carya, Cyclocarya*, and *Juglans* species.

of AGA(R) codon (Figure 2a), which codifies the amino acid arginine. However, the amino acids with higher observed frequency were leucine (11%) and isoleucine (9%), while the frequency of the arginine amino acid was 6% (Figure 2b).

Prospection of SSR loci in Juglandeae plastid genomes

A total of 77 SSR loci were identified in the plastid genome of *C. illinoinensis* cv. Imperial (Figure 2c), being 66 monomers (A/T), seven dimers (AT/TA) and four trimers (AAT/ATT). Nine SSR loci are located within the SSC region, five within each IR region, and 58 within the LSC region (Figure 2d). Among all 77SSR loci, 13 were located within coding regions, 10 within introns, and 54 in intergenic regions.

In comparison to the plastid genome of *C. illinoinensis* MH909599.1 (retrieved from the GenBank), cultivar Imperial has one more dimer and one more trimer locus. The highest number of SSR loci was identified in *Carya kweichowensis*, which also has the largest plastid genome.

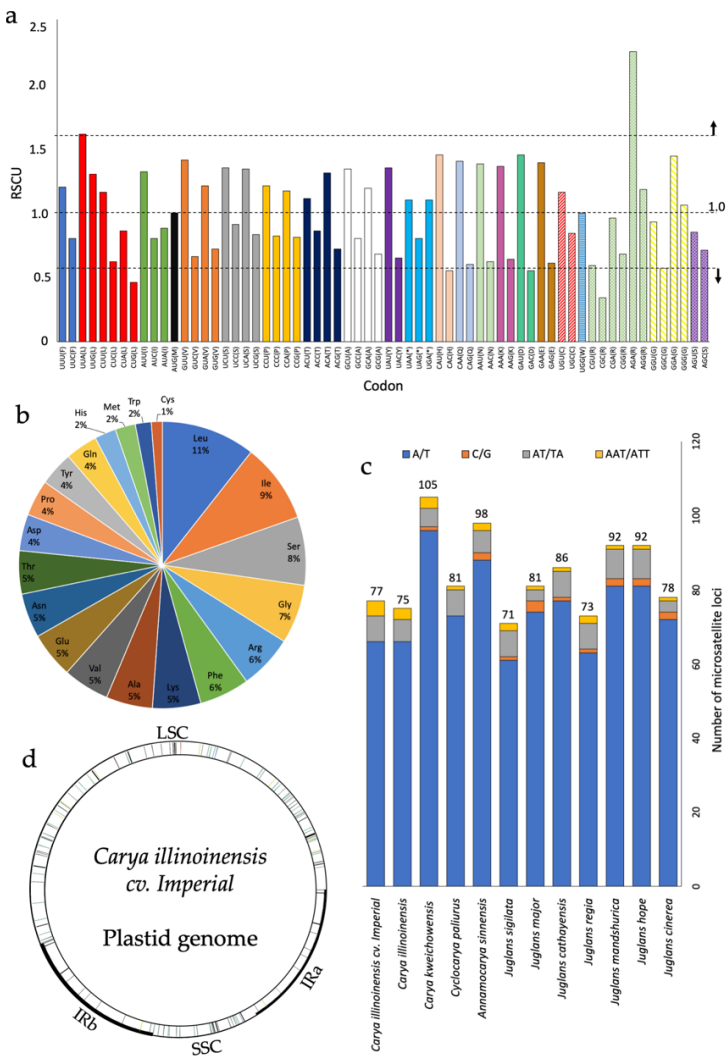


Figure 2 Analyses of codon usage of *C. illinoinensis* cv. Imperial and prospection of SSR loci. (a) Relative codon usage and proportion of amino acids in the CDS regions of the plastid genome of *C. illinoinensis* cv. Imperial. Columns with the same color represent synonymous codons for the same amino acid or the stop codon (*). The upper line ( ) corresponds to RSCU = 1.6 (over-representation limit), while the lower line ( ) corresponds to RSCU = 0.6 (under-representation limit). RSCU = 1.0 means absence of bias in codon usage. (b) Proportion of amino acids in the CDS regions of the plastid genome of *C. illinoinensis* cv. Imperial. (c) Number of monomer, dimer and trimer microsatellite loci prospected with-in the plastid genomes. (d) Distribution of the SSR loci within the plastid genome of *C. illinoinensis* cv. Imperial. Each line in the circle corresponds to one SSR loci.

This pattern is coherent with the large size of the *Carya kweichowensis* plastid genomebecause besides to expansion and contraction of the IR and SSC boundary regions, repetitive genomic elements are also believed to enlarge genome sizes (Stefenon et al. 2019b). With the prospection parameters employed, no C/G SSR was found in the plastid genomes of *C. illinoinensis* and *Cyclocarya paliurus*.

Phylogenetic relationships and evolution of Juglandeae based on the whole plastid genome analysis

The phylogenomic analysis based on the complete plastid sequences of 12 species of the tribe Juglandeae using *Castanea mollissima* (Fagaceae) as outgroup (Figure 3c) supports the phylogenetic studies based on morphological, nuclear and plastid genes. High bootstrap calculations (BP =100%) support all clades.

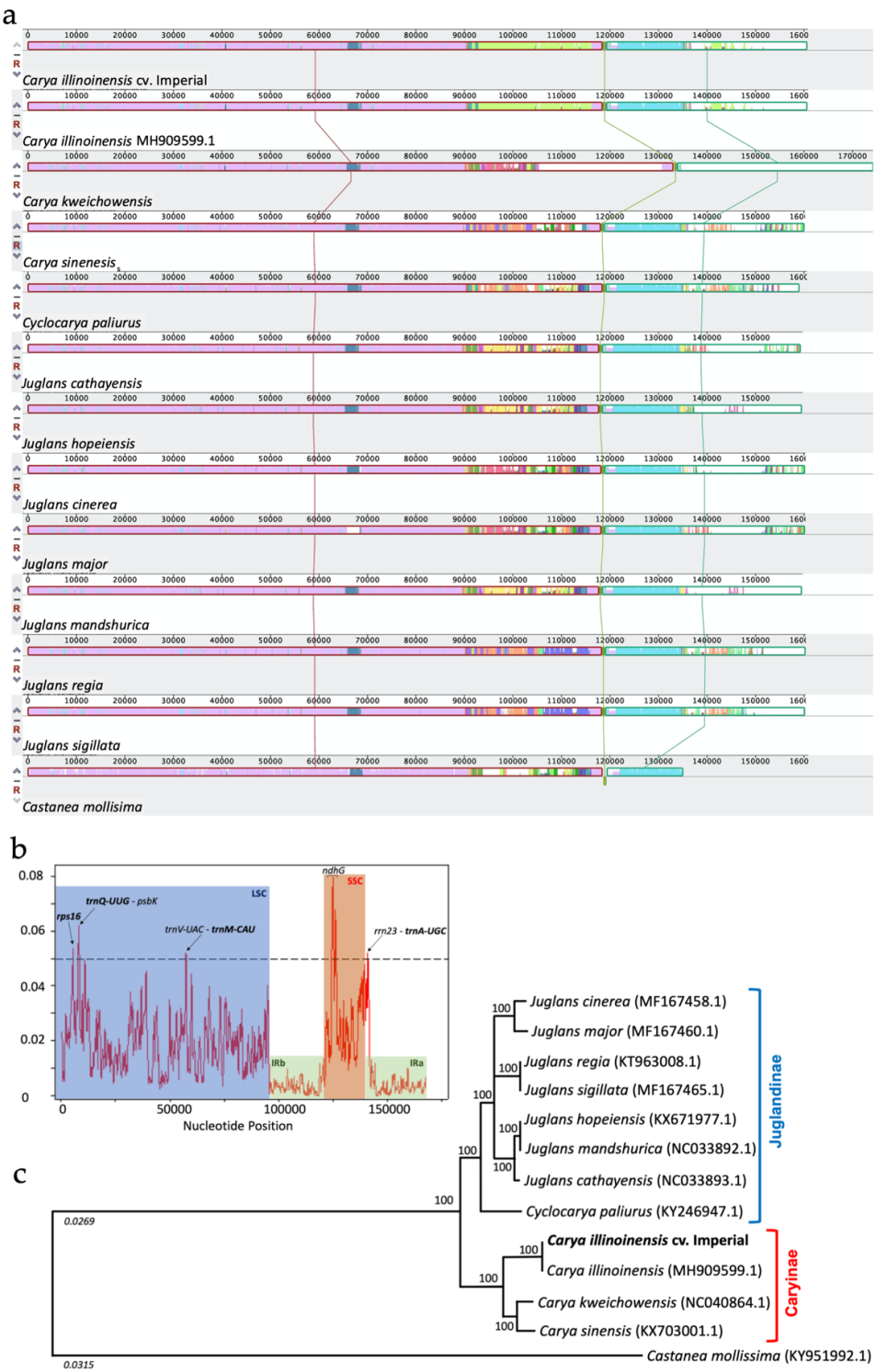
The phylogenomic tree (Figure 3c) clustered *Carya illinoinensis* cv. Imperial with *C. illinoinensis* MH909599.1 with 100% bootstrap support. *Carya kweichowensis* clustered with *Carya sinensis* (BP = 100%) as a sister group of *C. illioniensis* clade (BP = 100%). These species comprise the subtribe Caryinae within tribe Juglandeae. All seven species of *Juglans* formed a monophyletic clade sister to *Cyclocarya paliurus* (BP = 100%).*Juglans* and *Cyclocarya* comprise subtribe Juglandinae.

Genes rearrangements were observed among the evaluated species of Juglandeae and *Castanea mollissima* (Figure 3a). Forty-two homologous DNA regions are not collinear between *C. illinoinensis* cv. Imperial and *C. illinoinensis* MH909599.1 (data not shown). *Carya sinensis* presents 195 and *Carya kweichowensis* presents 236 non-collinear homologous DNA regions in comparison to *C. illinoinensis*, respectively. *Cyclocarya* presented relatively larger non-collinear homologous DNA regions in comparison to the species of *Carya* (Figure 3a).

The non-collinearity of homologous DNA among *Juglans* species is also restricted to minor regions, with a high similarity between *J. regia* and *J. sigillata* and between *J. cinerea* and *J. major* (data not shown), supporting the phylogenetic clustering of these species (Figure 3c). As expected, the larger difference concerning the Local Collinear Blocks (LCBs) is observed with species of Juglandaceae in comparison to *Castanea mollissima* (Figure 3a).

Based on the sliding window analysis, five hotspots of plastid nucleotide divergence were identified in this study (Figure 3b). Three of these hotspots are located within the LSC region (the *rps16* gene, the *trnQ-UUG - psbK* region, and the *trnV-UAC - trnM-CAU* region), one in the SSC region (the *ndhG* gene) and one in the IRb region (the *rrn23 - trnA-UGC* region). In each of the three hotspot sites composed of two coding-regions, the largest portion of the hotspot sequence was related to one tRNA gene (*trnQ-UUG, trnM-CAU*, and *trnA-UGC*, Figure 3b). According to the codon

Figure 3 Evolutionary relationships of species from tribe Juglandeae. (a) Multiple plastid genome alignment with species from tribe Juglandeae and *Castanea mollissima* (Fagaceae). Different colors represent different Local Colinear Blocks (LCBs), indicating sequence rearrangements and/or translocation. (b) Sliding window analysis of aligned whole plastid genomes of the 12 species of tribe Juglandeae. The regions with high nucleotide variability ($\pi > 0.050$) and approximated limits of the SSC, LSC, IRa, and IRb regions are indicated. (c) Phylogenomic relationship among species of Juglandeae, based on the NJ algorithm and using *Castanea mollissima* (Fagaceae) as outgroup. Number at nodes are the bootstrap support after 500 replicates. Numbers below branches are their respective length. Plastid genome sequences of all species were downloaded from Genbank (ID number given after each species name), except for *C. illinoinensis* cv. Imperial, which plastid genome sequence was generated in this study.


a
Carya illinoinensis cv. Imperial
Carya illinoinensis MH909599.1
Carya kweichowensis
Carya sinenesis
Cyclocarya paliurus
Juglans cathayensis
Juglans hopeiensis
Juglans cinerea
Juglans major
Juglans mandshurica
Juglans regia
Juglans sigillata
Castanea mollisima
b
LSC
SSC
IRb
IRa
ndhG
trnQ-UUG - psbK
rps16
trnV-UAC - trnM-CAU
rrn23 - trnA-UGC
Nucleotide Position
c
Juglans cinerea (MF167458.1)
Juglans major (MF167460.1)
Juglans regia (KT963008.1)
Juglans sigillata (MF167465.1)
Juglans hopeiensis (KX671977.1)
Juglans mandshurica (NC033892.1)
Juglans cathayensis (NC033893.1)
Cyclocarya paliurus (KY246947.1)
Carya illinoinensis cv. Imperial
Carya illinoinensis (MH909599.1)
Carya kweichowensis (NC040864.1)
Carya sinensis (KX703001.1)
Castanea mollissima (KY951992.1)
Juglandinae
Caryinae
0.0269
0.0315

usage analysis (Figure 2a), codons UUG(L) and CAU(H) are used more often than expected, while the UGC(C) is used less often than expected without bias in codon usage. This suggests that the bias in codon usage can be an effect of evolutive selection, reflected in the presence of tRNA genes in the hotspots of plastid nucleotide divergence.

The *dN/dS* ratio of the five hotspots of diversity revealed values compatible with a pattern of positive selection (*dN/dS* > 1). The mean values of dN/dS ratio overall codon were (i) *rps16* = 3.9, (ii) *trnQ-UUG* - *psbK* = 4.25, (iii) *trnV-UAC* - *trnM-CAU* = 4.25, (iv) *ndhG* = 4.25, and (v) *rrn23*- *trnA-UGC* = 4.25. The species pairwise distance and the mean overall distance (*d*) based on the mean number of substitutions per site (Tables S1-S5 – Supporting information) and the maximum parsimony tree based on the matrix of species pairwise distances (Figure S2– Supporting information) revealed different patterns for each hotspot. The highest overall distance among populations (*d* = 0.514) was retrieved in the *trnQ-UUG – psbK* region, followed by the *rrn23 - trnA-UGC* region (*d* = 0.418). The mean overall distance for the other hotspot regions of nucleotide diversity were lower than 0.009. The clustering of the species in the maximum parsimony trees was different for each region and presented low bootstrap support for most of the branches, except for the *trnQ-UUG – psbK* region which presented bootstrap values higher than 62%.

Discussion

Carya species are adapted to a broad range of climate types in North America, resulting in the existence of numerous cultivars and hybrid lines of *C. illinoinensis* and worldwide commercial cultivation (Huang et al. 2019). Thus, understanding the genomic structure and evolutive history of this species is crucial for the conservation and exploitation of the species' genetic resources. In reporting the complete plastid genome sequence of *Carya illinoinensis* cv. Imperial and comparing evolutive and structural parameters of the plastid genomes, we add to the scientific literature, novel genomic resources and evolutionary insights about this species.

The general structure, the number, and the category of genes in the plastid genome of *Carya illinoinensis* cv. Imperial (GenBank ID MN221384) is comparable to most angiosperms (Hu et al. 2016, Dong et al. 2017, Hu et al. 2017, Zhai et al. 2019) and very similar to most species of Juglandeae, with small size differences ranging from 1 to 1089 bp (Table 2). An expressive exception is the plastid genome of *C. kweichowensis*, which underwent a very large expansion of the IRs resulting in a meaningful increase in its genome size and a severe reduction in the SSC region (Table 2). It seems that this Chinese endemic species experienced a particular event of IR expansion/SSC contraction, not shared by other species of the tribe Juglandeae, resulting in quite evident differences in the size of the IRa, IRb, and SSC regions. Expansion and contraction of the IR regions and the LSC/IR/SSC boundaries have been recognized as the main cause of the size variation of angiosperm plastid genomes (Dugas et al. 2015). Different from the IR expansion observed in *C. kweichowensis*, the IR/SSC and IR/LSC boundaries are highly conserved among all species of Juglandeae included in this study, with only some subtle expansion/contraction events of the IRs (Figure 1b). Sequence variabilities, especially in the *ycf1* gene at IR-SSC junction and in the *rps19* and *rpl22* genes at IR-LSC junction, are frequently observed as a result of expansion and contractions events by gene conversion (Zhu et al. 2016, Lopes et al. 2018).

Although numerous factors such as mutational pressure, translational selection, compositional constraints, and gene length may cause bias in codon usage patterns, we suggest that the biased usage towards codons with bases A and U in the third position (26 out of 28 preferentially used codons) observed

in *C. illinoinensis* cv. Imperial is an effect of the AT-rich composition of the species' plastid genome. A similar pattern of codon usage bias and amino acid frequency has been reported in plastid genomes of other species as *Angelica polymorpha* (Apiaceae) (Park et al. 2019). However, the existence of three tRNA-coding regions with biased usage and that were also identified as hotspots of nucleotide divergence suggest the occurrence of evolutive selection over these codons. High estimations of *dN/dS* ratio suggest these hotspots of divergence are experiencing positive selection in Junglandeae.

SSR loci are repetitive elements found in nuclear and organellar genomes and are widely employed in genetic studies (Lemos et al. 2018). The plastid SSR loci identified in this study have potential usefulness as molecular markers for studies of species and even cultivars differentiation. The difference in number of SSR loci among species (Figure 2c) means that some of them are species-specific and can be characterized as markers for species differentiation. Moreover, some SSR loci revealed polymorphism when *C. illinoinensis* cv. Imperial and *C. illinoinensis* MH909599.1 are compared (Table 3, Supporting Information) and can be used for the development of haplotypic genetic barcodes for cultivars identification. The high fraction of SSR loci prospected in non-coding regions (intergenic regions and introns) of *C. illinoinensis* cv. Imperial may provide informative genetic markers and their deep characterization *in silico* and the laboratory is an ongoing project of our research group. Plastid markers are uniparentally inherited and lack allele recombination being, therefore, convenient to a wide range of genetic researches including cultivars identification, population genetics, gene flow, hybridization/introgression, and characterization of species evolutionary history.

Approximately 83% of the identified SSR loci are located within introns and intergenic regions. SSR loci located within intergenic regions and introns evolve faster than CDSs (Rogalski et al. 2015) and are, therefore, more useful as molecular markers. Thus, these SSR loci prospected in the plastid genome of *C. illinoinensis* cv. Imperial have the potential for the development of markers for genetic studies at population, species and genus levels. Species-specific SSR markers have generated an elevated number of polymorphic loci than those employing universal primers (Wheeler et al. 2014). The difference in the number of SSR loci observed among species in this study highlights the possibility of selecting species-specific markers for each investigated species.

Moreover, we show that despite the significant structural differences among plastid genomes, as LCBs rearrangements and differences in the IR/SSC and IR/LSC boundaries, the phylogenetic relationship among species is supported when complete plastid genome sequences are analyzed as a single molecule. Overall, the species structure retrieved in our phylogenomic tree resembles the phylogenetic relationships of Juglandeae species based on plastid genes (Stanford et al. 2000; Manos et al. 2007), nuclear ITS (Stanford et al. 2000), and morphological data (Manos et al. 2007). Moreover, our phylogenomic analysis supports the monophyly of subtribes Caryinae and Juglandinae within family Juglandaceae and the monophyly of the genus *Juglans* is corroborated.

Single nucleotide variants, indels, and large structural variants were reported for plastid genomes of *Juglans* species (Hu et al. 2017). In our study, the distribution of the Local Colinear Blocks (Figure 3a) reveals low difference among species (with exception of the quite larger plastid genome of *C. kweichowensis*) and represents the phylogenetic pattern of the species, with phylogenetic closer species sharing the position of the LCBs, without significant gene rearrangements.

Moreover, we investigated whether any plastid gene of the studied species underwent positive selection. The sliding window analysis identified a total of five hotspots of nu-

cleotide divergence, including protein-coding genes (*rps6* and *ndhG*) and intergenic regions (*trnQ*-UUG – *psbK*, *trnV*-UAC – *trnM*-CAU, and *rrn23* – *trnA*-UGC). Six protein-coding genes and three intergenic regions were identified as a hotspot of nucleotide diversity among species of *Juglans* (Hu et al. 2017). However, the hotspots observed among *Juglans* species are all different from the hotspots we identified in our study. Among species of tribe Brasiceae, the *rps16* gene was also found to be a hotspot of nucleotide divergence, presenting a signature of positive selection (Lopes et al. 2018). In our study, the *dN/dS* ratio of the hotspots of nucleotide diversity also revealed a signature of positive selection for all regions. Although the effect of the positive selection on the protein function and the adaptive capacity are still poorly understood, several studies have reported signatures of positive selection on plastid genes (Lopes et al. 2018).

Conclusion

In conclusion, the findings of this study have important implications in the areas of genetics, evolution, conservation, breeding, and biotechnology of *Carya illinoinensis*. Here we reported the complete plastid genome of *C. illinoinensis* cv. Imperial, a pecan variety cultivated in southern Brazil and revealed the existence of minor differences among plastid genomes of species of tribe Juglandeae. On the other hand, the presence of signatures of positive selection in some genes and intergenic regions was also observed. Moreover, this plastid genome rises as an important genomic resource, enabling the improvement of phylogenomic analyses based on whole plastid genome sequences and also the development of plastid SSR markers with wide applications in population genetics, cultivars identification, and genetic improvement.

Acknowledgments

The authors would like to thank Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq/Brazil, Grant n. 302501/2017-7) andCoordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES/Brazil, Finance code 001) for the financial support, scholarshipsand grants awarded. We thank Prof. Dr. Luiz Fernando Würdig Roesch for supporting the NGS facilities.

Data archiving statement

The whole plastid genome sequence of *Carya illinoinensis* cv. Imperial is deposited in the Genbank database under number MN221384. Raw sequence data were deposited in the NCBI Sequence Read Archive (SRA) database under number SSR10382885, Bioproject PRJNA587009, Biosample SAMN13174479.

References

Amiryousefi A., Hyvönen J., Poczai P., 2018. IRscope: an online program to visualize the junction sites of chloroplast genomes. Bioinformatics 34:3030-3031. DOI: 10.1093/bioinformatics/bty220

Beier S., Thiel T., Münch T., Scholz U., Mascher M., 2017. MISA-web: a web server for SSR prediction. Bioinformatics 33:2583-2585. DOI: 10.1093/bioinformatics/btx198

Bock R., 2017. Witnessing genome evolution: experimental reconstruction of endosymbiotic and horizontal gene transfer. Annu Rev Genet. 51:1-22. DOI: 10.1146/annurev-genet-120215-035329

Chan P.P., Lowe T.M., 2019. tRNAscan-SE: Searching for tRNA genes in genomic sequences. Methods Mol Biol. 1962:1-14. DOI: 10.1007/978-1-4939-9173-0_1

Curtu A.L., Gailing O., Finkeldey R., 2007. Evidence for hybridization and introgression within a species-rich oak (*Quercus* spp.) community. BMC Evolutionary Biology 7:218. DOI: 10.1186/1471-2148-7-218

Darling, A.C.E., 2004. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. Genome Res 14:1394-1403. DOI: 10.1101/gr.2289704

Dong W., Xu C., Li W., Xie X., Lu Y., Liu Y., Jin X., Suo Z., 2017. Phylogenetic resolution in Juglans based on complete chloroplast genomes and nuclear DNA

sequences. Front. Plant Sci. 8:1148. DOI: 10.3389/fpls.2017.01148

Doyle J.J., Doyle J.L., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull Bot Soc Am. 19:11-15.

Dugas D.V., Hernandez D., Koenen, et al., 2015. Mimosoid legume plastid genome evolution: IR expansion, tandem repeat expansions, and accelerated rate of evolution in clpP. Sci. Rep. 5:1-13. DOI: 10.1038/srep16958

Edgar R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792-1797. DOI: 10.1093/nar/gkh340

Hu Y., Chen X., Feng X., Woeste K.E., Zhao P., 2016. Characterization of the complete chloroplast genome of the endangered species *Carya sinensis* (Juglandaceae). Conservation Genet Resour 8:467-470. DOI: 10.1007/s12686-016-0601-4

Hu Y., Woeste K.E., Zhao P., 2017. Completion of the chloroplast genomes of five chinese Juglans and their contribution to chloroplast phylogeny. Front. Plant Sci. 7:1955. DOI: 10.3389/fpls.2016.01955

Huang Y., Xiao L., Zhang Z. et al., 2019. The genomes of pecan and Chinese hickory provide insights into *Carya* evolution and nut nutrition. GigaScience 8: 1-17. DOI: 10.1093/gigascience/giz036

Katoh K., Rozewicki J., Yamada K.D., 2017. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Briefings in Bioinformatics bbx108. DOI: 10.1093/bib/bbx108

Laslett D., Canback B., 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res 32:11-16. DOI: 10.1093/nar/gkh152

Lemos R.P.M., Matielo C.B.D'O., Beise D.C., Rosa V.G., Sarzi D.S., Roesch L.F.W., Stefenon V.M., 2018. Characterization of plastidial and nuclear SSR Markers for understanding invasion histories and genetic diversity of *Schinus molle* L. Biology 7:43. DOI: 10.3390/biology7030043

Librado P., Rozas J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451-1452. DOI: 10.1093/bioinformatics/btp187

Liu C., Shi L., Zhu Y., Chen H., Zhang J., Lin X., Guan X., 2012. CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. BMC Genomics 13: 715. DOI: 10.1186/1471-2164-13-715

Lohse M., Drechsel O., Kahlau S., Bock R., 2013. Organellar Genome - DRAW - a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. Nucleic Acids Res 41: W575-W58. DOI: 10.1093/nar/gkt289

Lopes A.S., Pacheco T.G., Santos K.G., Vieira L.N., Guerra M.P., Nodari R.O., Souza E.M., Pedrosa F.O., Rogalski M., 2017. The *Linum usitatissimum* L. plastome reveals atypical structural evolution, new editing sites, and the phylogenetic position of Linaceae within Malpighiales. Plant Cell Rep. 37: 307-328. DOI: 10.1007/s00299-017-2231-z

Lopes A.S., Pacheco T.G., Nimz T., Vieira L.N., Guerra M.P., Nodari R.O., Souza E.M., Pedrosa F.O., Rogalski M., 2018. The complete plastome of macaw palm [*Acrocomia aculeata* (Jacq.) Lodd. ex Mart.] and extensive molecular analyses of the evolution of plastid genes in Arecaceae. Planta 247: 1011-1030. DOI: 10.1007/s00425-018-2841-x

Manos P.S., Soltis P.S., Soltis D.E., et al., 2007.Phylogeny of extant and fossil Juglandaceae inferred from the integration of molecular and morphological data sets. Syst. Biol. 56: 412-430. DOI: 10.1080/10635150701408523

Matielo C.B.D'O., Lemos R.P.M., Sarzi D.S., Machado L.O., Beise D.C., Dobbler P.C.T., Castro R.M., Fett M.S., Roesch L.F.W., Camargo F.O., Stefenon V.M., 2019. Whole plastid genome sequences of two drug-type Cannabis: insights into the use of plastid in forensic analyses. Journal of Forensic Sciences DOI: 10.1111/1556-4029.14155

Park I., Yang S., Kim W.J., et al., 2019. Sequencing and comparative analysis of the chloroplast genome of *Angelica polymorpha* and the development of a novel indel marker for species identification. Molecules 24: 138. DOI: 10.3390/molecules24061038

Perdereau P.C., Kelleher C.T., DouglasG.C., Hodkinson T.R., 2014. High levels of gene flow and genetic diversity in Irish populations of *Salix caprea* L. inferred from chloroplast and nuclear SSR markers. BMC Plant Biology 14:202. DOI: 10.1186/s12870-014-0202-x

Poletto I., Muniz M.F.B., Poletto T., Stefenon V.M., Baggiotto C., Ceconi D.E., 2015. Germination and development of pecan cultivar seedlings by seed stratification. Pesq. Agropec. Bras. 50:1232-1235. DOI: 10.1590/S0100-204X2015001200014

Poletto T., Stefenon V.M., Poletto I., Muniz M.F.B., 2018. Pecan propagation: Seed mass as a reliable tool for seed selection. Horticulturae 4: 26. DOI: 10.3390/horticulturae4030026

Poletto T., Poletto I., Silva L.M.M., Muniz M.F.B., Reiniger L.R.S., Richards N., Stefenon V.M., 2019. Morphological, chemical and genetic analysis of southern Brazilian pecan (*Carya illinoinensis*) accessions. Scientia Horticulturae. DOI: 10.1016/j.scienta.2019.108863

Rogalski M., Vieira L.N., Fraga H.P., Guerra M.P., 2015. Plastid genomics in horticultural species: importance and applications for plant population genetics, evolution, and biotechnology. Front Plant Sci 6: 586. DOI: 10.3389/fpls.2015.00586

Stanford A.M., Harden R., Parks C.R., 2000. Phylogeny and biogeography of Juglans (Juglandaceae) based on matK and its sequence data. Am. J. Bot. 87:872-882. DOI: 10.2307/2656895

Stefenon V.M., Kablunde G., Lemos R.P.M.,Rogalski M., Nodari R.O., 2019a. Phylogeography of plastid DNA sequences suggests post-glacial southward demograph-

ic expansion and the existence of several glacial refugia for *Araucaria angustifolia*. Scientific Reports 9:2752. DOI: 10.1038/s41598-019-39308-w

Stefenon V.M., Sarzi D.S., Roesch L.F.W., 2019b. High throughput sequencing analysis of *Eugenia uniflora*: insights into repetitive DNA, gene content and potential biotechnological applications. 3 Biotech 9:200. DOI: 10.1007/s13205-019-1729-1

Tamura K., Stecher G., Peterson D., Filipski A., Kumar S., 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30:2725-2729. DOI: 10.1093/molbev/mst197

Tillich M., Lehwark P., Pellizzer T., 2017. GeSeq - versatile and accurate annotation of organelle genomes. Nucleic Acids Research 45: W6-W11. DOI: 10.1093/nar/gkx391

Vieira L.N., Rogalski M., Faoro H., Fraga H.P., Anjos K.G., Picchi G.F.A., Nodari R.O., Pedrosa F.O., Souza E.M., Guerra M.P., 2016. The plastome sequence of the endemic Amazonian conifer, *Retro-phyllum piresii* (Silba) C.N.Page, reveals different recombination events and plastome isoforms. Tree Genet Genomes 12: 10. DOI: 10.1007/s11295-016-0968-0

Wheeler G.L., Dorman H.E., Buchanan A., Challagundla L., Wallace L.E., 2014. A review of the prevalence, utility, and caveats of using chloroplast simple sequence repeats for studies of plant biology. Appl Plant Sci. DOI: 10.3732/apps.1400059

Wicke S., Schneeweiss G.M., dePamphilis C.W., Müller K.F., Quandt D., 2011.The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol Biol 76: 273-297. DOI: 10.1007/s11103-011-9762-4

Wolfe K.H., Li W.H., Sharp P.M., 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc. Natl. Acad. Sci. USA 84:9054-9058. DOI: 10.1073/pnas.84.24.9054

Ye L., Fu C., Wang Y., Liu J., Gao L., 2018. Characterization of the complete plastid genome of a Chinese endemic species *Carya kweichowensis*. Mitochondrial DNA Part B: Resources 3:492-493. DOI: 10.1080/23802359.2018.1464414

Zhai D-C., Yao Q., Cao X-F., Hao Q-Q., Ma M-T., Pan J., Bai X-H., 2019.Complete chloroplast genome of the wild-type Hickory *Carya cathayensis*. Mitochondrial DNA Part B: Resources 4:1457-1458. DOI: 10.1080/23802359.2019.1598815

Zhu A., Guo W., Gupta S., Fan W., Mower J.P., 2016. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. New Phytol 209:1747-1756. DOI: 10.1111/nph.13743

Supporting information

Figure S1. *Carya illinoinensis* cv. Imperial. (a) Fruits in intermediate stage of ripening. (b) Ripe nuts and (c) the kernel of the nut. (d) Physical map of the plastid genome of *C. illinoinensis* cv. Imperial. The darker gray area in the inner circle represents the GC content, while light gray is the AT content. Coding regions containing introns are marked with one asterisk.

Tables S1-S5. Mean number of base substitutions per site and the corresponding standard errors of the hotspot sites of nucleotide divergence among 11 species of the tribe Juglandeae.

Figure S2. Maximum Parsimony analysis of 11 species of tribe Juglandeae, using the hotspot regions of nucleotide divergence, determined through a sliding window analysis.

Table S6. Examples of homologous SSR loci in *Carya illinoniensis* cv. Imperial and Carya illinoinensis MH909599.1. Loci 1 and 2 have the same motif, size and position in the plastid genome of both individuals. Loci 3 and 4 differ in the start and end position due to the one base deletion in *C. illinoniensis* cv. Imperial. Loci 5 to 7 have differences concerning the number of repeats of the motif, representing putatively polymorphic markers.