Planta
https://doi.org/10.1007/s00425-018-2841-x

ORIGINAL ARTICLE

The complete plastome of macaw palm [*Acrocomia aculeata* (Jacq.) Lodd. ex Mart.] and extensive molecular analyses of the evolution of plastid genes in Arecaceae

Amanda de Santana Lopes[1] · Túlio Gomes Pacheco[1] · Tabea Nimz[1] · Leila do Nascimento Vieira[2] · Miguel P. Guerra[2] · Rubens O. Nodari[2] · Emanuel Maltempi de Souza[3] · Fábio de Oliveira Pedrosa[3] · Marcelo Rogalski[1]

Received: 1 November 2017 / Accepted: 10 January 2018

Abstract

Main conclusion **The plastome of macaw palm was sequenced allowing analyses of evolution and molecular markers. Additionally, we demonstrated that more than half of plastid protein-coding genes in Arecaceae underwent positive selection.**

Macaw palm is a native species from tropical and subtropical Americas. It shows high production of oil per hectare reaching up to 70% of oil content in fruits and an interesting plasticity to grow in different ecosystems. Its domestication and breeding are still in the beginning, which makes the development of molecular markers essential to assess natural populations and germplasm collections. Therefore, we sequenced and characterized in detail the plastome of macaw palm. A total of 221 SSR loci were identified in the plastome of macaw palm. Additionally, eight polymorphism hotspots were characterized at level of subfamily and tribe. Moreover, several events of gain and loss of RNA editing sites were found within the subfamily Arecoideae. Aiming to uncover evolutionary events in Arecaceae, we also analyzed extensively the evolution of plastid genes. The analyses show that highly divergent genes seem to evolve in a species-specific manner, suggesting that gene degeneration events may be occurring within Arecaceae at the level of genus or species. Unexpectedly, we found that more than half of plastid protein-coding genes are under positive selection, including genes for photosynthesis, gene expression machinery and other essential plastid functions. Furthermore, we performed a phylogenomic analysis using whole plastomes of 40 taxa, representing all subfamilies of Arecaceae, which placed the macaw palm within the tribe Cocoseae. Finally, the data showed here are important for genetic studies in macaw palm and provide new insights into the evolution of plastid genes and environmental adaptation in Arecaceae.

Keywords Palm tree · Plastid genome · Plastid SSRs · Polymorphism hotspots · Gene divergence · Positive selection

Introduction

Macaw palm [*Acrocomia aculeata* (Jacq.) Lodd. Ex Mart.] is a native species distributed in the tropical and subtropical Americas, with center of origin in Brazil (Henderson et al. 1995; Lanes et al. 2015). Its fruits are oil-rich, accumulating up to 70% of oil (dry-weight) and yielding approximately 6200 kg of oil per hectare (Pires et al. 2013). If compared with current oil crops the macaw palm can reach oil production similar to the oil crops with the highest productivity such as oil palm (*Elaeis guineensis* Jacq.) (Motoike and Kuki 2009). In addition to the high oil productivity its oil profile is suitable for biofuel production (Pires et al. 2013; Lanes et al. 2014). The integral use of the fruits still provides

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00425-018-2841-x) contains supplementary material, which is available to authorized users.

✉ Marcelo Rogalski
rogalski@ufv.br

[1] Laboratório de Fisiologia Molecular de Plantas, Departamento de Biologia Vegetal, Universidade Federal de Viçosa, Viçosa, MG, Brazil

[2] Laboratório de Fisiologia do Desenvolvimento e Genética Vegetal, Programa de Pós-Graduação em Recursos Genéticos Vegetais, Universidade Federal de Santa Catarina, Florianópolis, SC, Brazil

[3] Departamento de Bioquímica e Biologia Molecular, Núcleo de Fixação Biológica de Nitrogênio, Universidade Federal do Paraná, Curitiba, PR, Brazil

Published online: 16 January 2018

biomass feedstock for several industrial applications focusing on sustainable energy such as ethanol and charcoal (Vilas-Boas et al. 2010; Gonçalves et al. 2013; Pires et al. 2013). The macaw palm has interesting agronomic and ecological features because it can occupy degraded areas or agroforestry systems given that it has high plasticity to grow in different ecosystems (Henderson et al. 1995; Motoike and Kuki 2009; Pires et al. 2013) avoiding conflict with areas of food production.

Due to its attractive features, the macaw palm domestication, genetic breeding and development of commercial plantations have encouraged financial investments in basic research, biotechnology and industry. Studies in several areas have been carried out with this purpose, including ecophysiology (Pires et al. 2013), mating system (Lanes et al. 2016), vegetative development (Berton et al. 2013; Machado et al. 2016), fruit development (Montoya et al. 2016), phenotypic diversity (Ciconini et al. 2013; Lanes et al. 2015; Conceição et al. 2015; Coser et al. 2016) and biotechnological applications (Moura et al. 2009; Luis and Scherwinski-Pereira 2014; Padilha et al. 2015).

Considering the interest to make macaw palm an alternative platform for the production of renewable energy, the characterization of molecular markers has essential importance to assess the genetic diversity of natural populations aiming the establishment of suitable strategies for conservation, germplasm characterization, domestication and genetic breeding. However, a small number of molecular markers were developed for macaw palm (Mengistu et al. 2016a, b; Nucci et al. 2008) and all of them are based on nuclear microsatellites (SSRs). The nonrecombinant and uniparentally inherited nature of plastid genome (plastome) makes it a great source of molecular markers, particularly intergenic spacers (IGSs) and introns, where the mutation rates are higher in comparison with coding sequences (Rogalski et al. 2015; Vieira et al. 2016a). Plastid SSRs have been used in several genetic studies including phyllogeography, population genetic and gene flow analyses (Provan et al. 2001; Ebert and Peakall 2009; Wheeler et al. 2014). Plastid sequences with high polymorphism are also applied to assess the genetic structure and diversity/divergence of natural populations and germplasm collections (Tsai et al. 2015; Wambulwa et al. 2016; Roy et al. 2016).

Plastome sequences are also of great importance to understand evolutionary events in plants based on the knowledge of gene content, recombination events, loss of genes, gene transfer to the nucleus and genome rearrangements (Vieira et al. 2016b; Bock 2017; Lopes et al. 2017; Park et al. 2017; Ruhlman et al. 2017), as well as for plastid transformation aiming basic research (Rogalski et al. 2006, 2008; Alkatib et al. 2012) and biotechnological applications (Daniell et al. 2016; Zhang et al. 2017) since the complete sequence of plastome is a prerequisite to choose target intergenic regions for insertion of transgenes and development of transplastomic plants (Daniell et al. 2016; Fuentes et al. 2017). The plastid transformation has been used efficiently for metabolic engineering (Daniell et al. 2016; Fuentes et al. 2017) and it is a viable alternative to manipulate plastid fatty acid biosynthesis (Rogalski and Carrer 2011) given that several efficient protocols of plant regeneration based on somatic embryogenesis are available for macaw palm (Moura et al. 2009; Luis and Scherwinski-Pereira 2014; Padilha et al. 2015).

Macaw palm belongs to the family Arecaceae, which contains about 2450 species distributed in five subfamilies, Calamoideae, Nypoideae, Arecoideae, Coryphoideae, and Ceroxyloideae (Dransfield et al. 2005; Asmussen et al. 2006; Barfod et al. 2011). The Arecoideae is the largest subfamily of Arecaceae, including *A. aculeata*, *Cocos nucifera* L., and *E. guineensis*, which are species of great economic importance (Baker et al. 2009; Comer et al. 2015). The genus *Acrocomia* is Neotropical occurring from north Mexico to south Argentina. The number of species is not taxonomically well resolved. The first classification included only two species to the genus, *A. aculeata* and *A. hassleri* (Henderson et al. 1995). The last classification recognizes eight species including *A. aculeata*, *A. crispa*, *A. emensis*, *A. glaucescens*, *A. hassleri*, *A. intumescens*, *A. media,* and *A. totai* (Lorenzi et al. 2010; The Plant List 2013). Recently, a new species was described, *Acrocomia corumbaensis*, showing an arborescent habit similar to *A. aculeata*, *A. crispa*, *A. intumescens,* and *A. totai* (Vianna 2017). The new data have demonstrated an increasing number of new species, which makes the plastid genomics a useful tool to classify correctly them. According to Barrett et al. (2016) the plastomes of Arecaceae have a low rate of variation compared with other commelinids (e.g. grasses). Nevertheless, the diversification rate over time within palm genera seems to increase and a convergent evolution has been reported among palm species adapted to shaded areas (Ma et al. 2015; Faurby et al. 2016). Despite most plastid genes are conserved, several evolutionary events have been described such as new RNA editing sites, loss of introns, high divergence of genes, and positive signatures (Sen et al. 2012; Williams et al. 2015; He et al. 2016; Chen et al. 2017). In grasses, one-third of the plastid genes underwent positive selection, which shows a relationship between gene evolution and environmental conditions regarding the photosynthetic apparatus (Piot et al. 2017). The recent diversification of palm species may be related to some adaptive changes in plastid genes. However, the magnitude of the changes in plastid genes that underwent positive selection and how it could be related to adaptation to different environmental conditions remain unknown in Arecaceae.

Here, we reported the complete plastome of macaw palm, which was molecularly characterized in details. The description and location of all SSR loci and polymorphism hotspots within macaw palm plastome were shown. Among the 221 SSR loci identified, 157 are located in fast-evolving regions (IGSs and introns). In addition, we performed a phylogenomic analysis using whole plastomes of 40 taxa, including 37 species representing all five subfamilies of Arecaceae, which placed the macaw palm within tribe Cocoseae. Moreover, some IGSs at the level of subfamily and tribe were identified as polymorphism hotspots, especially the *trnC*-GCA/*petN* and *psaC*/*ndhE*. Furthermore, 100 putative RNA editing sites were found within Arecoideae, including some possible events of gain and loss of editing sites. Finally, we investigated extensively the molecular evolution of all protein-coding genes from 37 palm plastomes. We identified some highly divergent genes in a species-specific manner suggesting that gene degeneration processes may be occurring within Arecaceae at the level of genus or species. We investigated if the plastid genes of Areacaceae underwent positive selection and the analysis indicated that more than half of the plastid protein-coding genes have one or more positive signatures, which can significantly affect essential plastid functions. Taken together, our data bring new molecular markers useful for genetic studies in macaw palm natural populations and raise questions about the evolution of plastome within Arecaceae and the relationship between positive selection and evolution of plastid genes under different environmental conditions.

Materials and methods

Chloroplast isolation, DNA extraction, sequencing, assembling, annotation, and data archiving statement

Fresh leaves from a young macaw palm plant were collected and kept for 1 week at 4 °C to decrease starch level. The young plant was obtained from seeds of macaw palm plants belonging to the ex situ plant collection, Macaúba Active Germplasm Bank (BAG-Macaúba repository: 084/2013/CGEN/MMA) located in the experimental farm of the Universidade Federal de Viçosa (208400100S, 4283101500W), State of Minas Gerais, Brazil (Coser et al. 2016; Lanes et al. 2015; Lanes et al. 2016; Mengistu et al. 2016a, b; Montoya et al. 2016).

The chloroplast isolation and cp DNA extraction were carried out according to Vieira et al. (2014). Approximately 1 ng of plastid DNA was used to prepare sequencing libraries with Nextera XT DNA Sample Prep Kit (Illumina Inc., San Diego, CA, USA) according to the manufacturer's instructions. The obtained library was sequenced using Illumina MiSeq platform (Illumina Inc., San Diego, CA, USA) at the Federal University of Paraná, State of Paraná, Brazil. The paired-end reads (total of 907,088 reads), with average length of 280.9 bp, were trimmed under the threshold with probability of error < 0.05. The trimmed reads (901,031 reads, average length of 218.4 bp) were de novo assembled in contigs using CLC Genomics Workbench 8.0.2 software (CLC Bio, Aarhus, Denmark). The contigs used for assembling of macaw palm plastome ranged from 1323.96 to 315.12 of average coverage.

The program Dual Organellar GenoMe Annotator (DOGMA) (Wyman et al. 2004) and BLAST were used for preliminarily gene annotation. From this initial annotation, putative start codons, stop codons, and intron positions were determined based on comparisons to homologous genes of other plastomes at the GenBank database. All tRNA genes were further verified by using tRNAscan-SE server (Lowe and Eddy 1997). The physical circular map of the plastome was drawn using Organellar Genome DRAW (OGDRAW) (Lohse et al. 2013). The complete nucleotide sequence of macaw palm plastome sequenced in this study was deposited in the GenBank database under accession number MG020488.

SSR identification, sliding window analysis and RNA editing sites prediction

Simple sequence repeats (SSRs) loci were detected in the macaw palm plastome and in other six Arecoideae plastomes available in the GenBank database (Supplementary Table S1) using the MIcroSAtellite (MISA) Perl script (Thiel et al. 2003). The thresholds were set to eight repeat units for mononucleotide SSRs, four repeat units for di- and trinucleotide SSRs, and three repeat units for tetra-, penta-, and hexanucleotide SSRs.

To identify the hotspots of sequence divergence in the subfamily Arecoideae and tribe Cocoseae, we performed the sliding window analysis. First, complete plastomes were aligned using MAFFT v.7 (Katoh and Standley 2013), and posteriorly, the sliding window analysis was conducted by using the DnaSP v.5 software (Librado and Rozas 2009). The window length and the step size were set as 200 and 50 bp, respectively.

Potential RNA editing sites in plastid protein-coding genes of species belonging to subfamily Arecoideae were predicted by the program predictive RNA editor for plants (PREP) suite (Mower 2009). The program PREP uses 35 reference genes for detecting of possible RNA editing sites in plastomes. The cutoff value was set to 0.8. Additional RNA editing sites were predicted by comparison with Huang et al. (2013) that based on RT-PCR and sequencing data identified several RNA editing sites in transcripts of plastid genes of *C. nucifera*.

Comparative analysis of plastome structure

To characterize the general structure of the subfamily Arecoideae, nucleotide MUMmer (NUCmer) Perl script in MUMmer 3.0 (Kurtz et al. 2004) was used to visualize and compare the plastome structures between *A. aculeata* and other Arecoideae representatives (Supplementary Table S1).

Phylogenomic reconstruction of the family Arecaceae

The phylogenetic reconstruction of the family Arecaceae was carried out using whole plastomes. The GenBank accession number of each taxon used here is shown in the Supplementary Table S1. The species *Hanguana malayana* (Hanguanaceae: Commelinales), *Baxteria australis*, and *Dasypogon bromeliifolius* (Dasypogonaceae: Arecales) were used as outgroups. First, whole plastomes were extracted from GenBank and the IRB was withdrawn to prevent overrepresentation of the IR sequences. The alignment of plastomes was done using MAFFT v.7 (Katoh and Standley 2013) and the best substitution model (GTR + I+ G) was selected by using jModelTest v.2.1.7 (Darriba et al. 2012). Last, Bayesian inference analysis was performed using MrBayes version 3.2 (Ronquist et al. 2012), with one million generations of two runs of four Markov Chains, with three hot and one cold in each run. To check the parameter convergence we used the software Tracer 1.6 (http://tree.bio.ed.ac.uk/software/tracer/). The software FigTree 1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/) was used to visualize the consensus tree.

Molecular evolution analysis of protein-coding genes in Arecaceae plastomes

The 79 protein-coding genes present in Arecaceae plastomes (Supplementary Table S1) were extracted and the codon alignment was done using the software Muscle (Edgar 2004) implemented in Mega 7.0 (Tamura et al. 2013). Phylogenetic reconstruction was performed to assess the gene divergence. First, substitution models for each gene were selected using jModelTest v.2.1.7 (Darriba et al. 2012), which are listed in the Supplementary Table S2. Then, a Bayesian inference analysis was performed using MrBayes version 3.2 (Ronquist et al. 2012), with two million generations of two runs of four Markov Chains, with three hot and one cold in each run and the software FigTree 1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/) was used to visualize the consensus tree. The gene divergence was estimated by the sum of total branch lengths that link the operational taxonomical units to the common ancestor of Arecaceae species sampled here. The convergence parameters of each tree were checked using the software Tracer 1.6 (http://tree.bio.ed.ac.uk/software/tracer/).

Finally, to investigate the presence of positive signatures (positive selection), the genes were aligned as described before. The positive signatures were analyzed using Selecton version 2.4 (http://selecton.tau.ac.il/index.html; Stern et al. 2007), performing the comparison M8 (allows positive selection) against M8a (null model) with one degree of freedom and cutoff (ε) of 0.1. We consider in our analysis only sites where reliable positive selection was inferred (lower bound > 1 and test with probability < 0.01).

Results

General features of macaw palm plastome and comparative analyses within the subfamily Arecoideae

The macaw palm plastome is a circular molecule of 155,829 bp in length with a typical quadripartite structure found in most plastomes of angiosperms (Wicke et al. 2011; Zhu et al. 2016), which include a pair of inverted repeats (IRs) between two regions of single copy, the large single copy (LSC) and the small single copy (SSC) (Fig. 1). The macaw palm plastome contains 113 unique genes, being 79 protein-coding genes, 30 tRNA genes, and four rRNA genes (Table 1). Some of these genes are duplicated in the IRs (Fig. 1); they are eight tRNA genes, all rRNA genes, and nine protein-coding genes (one of them, the *ycf1*, is partially duplicated). Among the 113 unique genes, 16 possess one intron (six tRNA genes and ten protein-coding genes) and two contain two introns (*clpP* and *ycf3* genes).

The size of LSC, SSC, and IRs of macaw palm plastome is compared with other plastomes from species belonging to the subfamily Arecoideae in the Table 2. They have similar dimensions of the quadripartite structure. The only exception, *Areca vestiaria*, contains a very small IR region, which is compensated by a larger LSC region. These general structural features were also explored by dot-plot analyses (Supplementary Fig. S1), which show high similarity and absence of rearrangements between macaw palm and other Arecoideae plastomes. The only exception is *A. vestiaria* that lost most part of the IR region, which corresponds to IRB in the plastome of macaw palm. A more detailed view of the IRA and IRB borders of Arecoideae plastomes (Fig. 2), except for *A. vestiaria* plastome, reveals slight differences within this subfamily. Most plastomes, including the macaw palm, present the *rps19* gene completely duplicated in the IR region, while in *C. nucifera* and *Syagrus coronata* it is partially duplicated in the IR borders and encodes a functional protein only in the LSC-IRA junction. A variability of nucleotide number in the *rps19-rpl22* and *rps19-psbA* intergenic spacers (IGS) is observed in the LSC-IRA and LSC-IRB junctions, respectively. Some variability is also

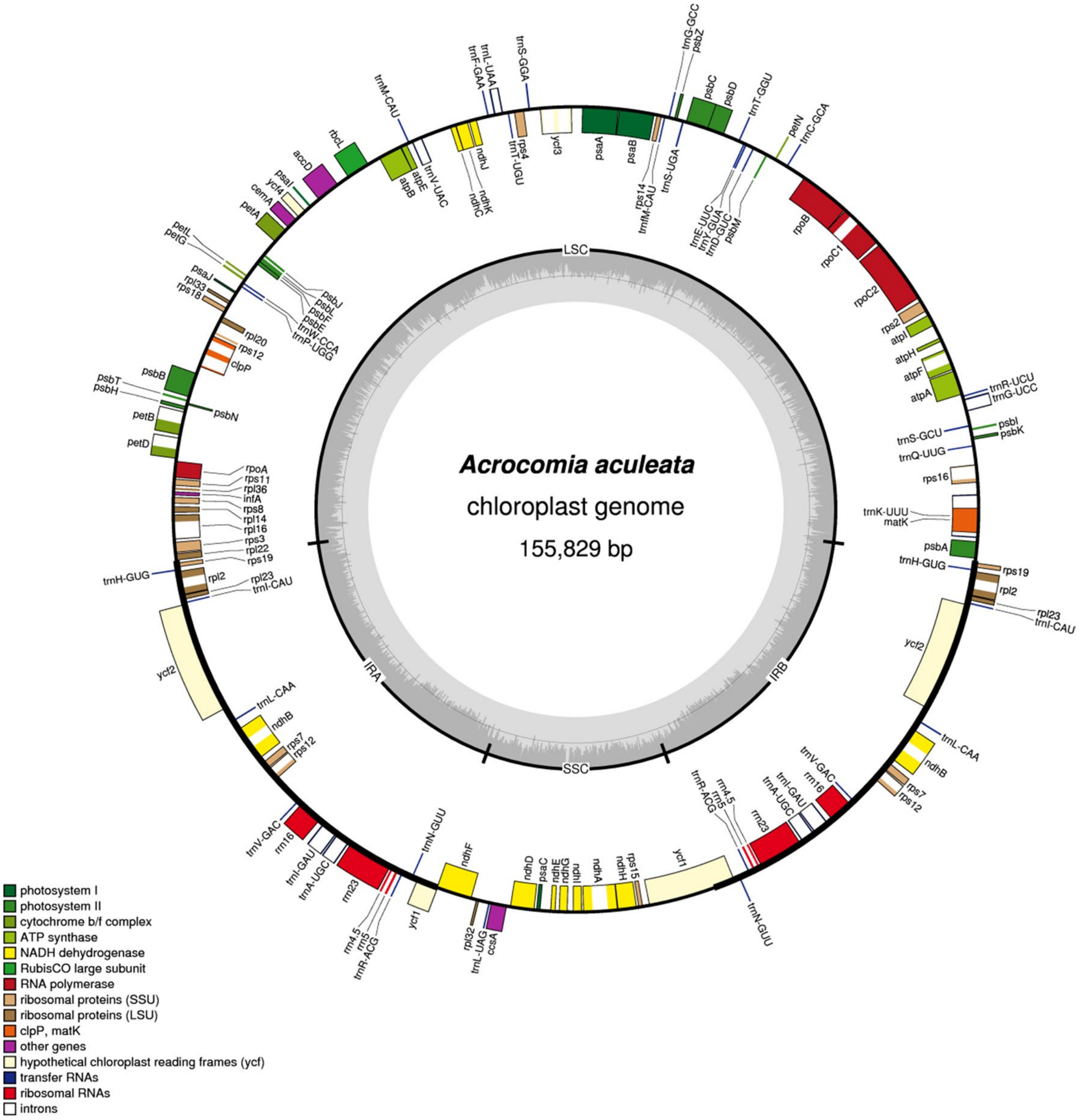


Fig. 1 Gene map of macaw palm (*Acrocomia aculeata*) plastome. Genes drawn inside the circle are transcribed in the clockwise direction, and genes drawn outside are transcribed in the counterclockwise direction. Different functional groups of genes are color-coded. The darker gray in the inner circle corresponds to GC content, and the lighter gray corresponds to AT content. *LSC* large single copy, *SSC* small single copy, *IRA/B* inverted repeat A/B

present in the IR-SSC junctions, mainly in the *ycf1* gene. Part of the *ndhF* gene is located within the IRs (overlapping part of the *ycf1* gene), which represents a stretch of 56 bp highly conserved among all species analyzed here (Fig. 2).

SSR content and nucleotide divergence analysis in macaw palm plastome and other Arecoideae species

The SSR loci number and distribution among Arecoideae plastomes (Fig. 3a), including *A. aculeata*, are very similar, with the majority composed of mononucleotide repeats (Fig. 3b). The mononucleotide repeats in Arecoideae plastomes are basically constituted of A/T sequences (ranging from 94.8 to 96.8% among the species sampled here). The density of SSR loci (SSR number/kilobase) was higher in the SSC (mean of 2.28 ± 0.11), followed by LSC (mean of 1.93 ± 0.07), and lower in the IR (mean of 0.75 ± 0.04). The total mean density, considering only one IR, was 1.75 (± 0.05). In the species *A. vestiaria* we found the same pattern of SSR distribution in the SSC, LSC, and IR regions as delimited in the other Arecoideae species analyzed here.

Table 1 List of genes identified in the plastome of *Acrocomia aculeata*

Group of gene	Name of gene
Gene expression machinery	
Ribosomal RNA genes	*rrn16*[b]; *rrn23*[b]; *rrn5*[b]; *rrn4.5*[b]
Transfer RNA genes	*trnA* –UGC[ab]; *trnC* –GCA; *trnD* –GUC; *trnE* –UUC; *trnF* –GAA; *trnfM* –CAU; *trnG* –UCC[a]; *trnG* –GCC; *trnH* –GUG[b]; *trnI* –CAU[b]; *trnI* –GAU[ab]; *trnK* –UUU[a]; *trnL* –CAA[b]; *trnL* –UAA[a]; *trnL* –UAG; *trnM* –CAU; *trnN* –GUU[b]; *trnP* –UGG; *trnQ* –UUG; *trnR* –ACG[b]; *trnR* –UCU; *trnS* –GCU; *trnS* –UGA; *trnS* –GGA; *trnT* –UGU; *trnT* –GGU; *trnV* –GAC[b]; *trnV* –UAC[a]; *trnW* –CCA; *trnY* –GUA
Small subunit of ribosome	*rps2; rps3; rps4; rps7*[b]; *rps8; rps11; rps12*[ab]; *rps14; rps15*; *rps16*[a]; *rps18; rps19*[b]
Large subunit of ribosome	*rpl2*[ab]; *rpl14; rpl16*[a]; *rpl20; rpl22; rpl23*[b]; *rpl32; rpl33; rpl36*
DNA-dependent RNA polymerase	*rpoA; rpoB; rpoC1*[a]; *rpoC2*
Translational initiation factor	*infA*
Intron maturase	*matK*
Genes for photosynthesis	
Subunits of photosystem I (PSI)	*psaA; psaB; psaC; psaI; psaJ; ycf3*[a]; *ycf4*
Subunits of photosystem II (PSII)	*psbA*; *psbB; psbC; psbD; psbE; psbF; psbH; psbI; psbJ; psbK; psbL; psbM; psbN; psbT; psbZ*
Subunits of cytochrome b_6f	*petA; petB*[a]; *petD*[a]; *petG; petL; petN*
Subunits of ATP synthase	*atpA; atpB; atpE; atpF*[a]; *atpH; atpI*
Subunits of NADH dehydrogenase	*ndhA*[a]; *ndhB*[ab]; *ndhC; ndhD; ndhE; ndhF; ndhG; ndhH*; *ndhI; ndhJ; ndhK*
Large subunit of Rubisco	*rbcL*
Other functions	
Envelope membrane protein	*cemA*
Subunit of acetyl-CoA carboxylase	*accD*
C-type cytochrome synthesis	*ccsA*
Subunit of protease Clp	*clpP*[a]
Component of TIC complex	*ycf1*[c]
Unknown function	*ycf2*[b]

[a] Genes containing introns

[b] Duplicated genes

[c] Partially duplicated genes

Table 2 General features of plastid genomes within the subfamily Arecoideae

	Acrocomia aculeata	*Areca vestiaria*	*Cocos nucifera*	*Elaeis guineensis*	*Podococcus barteri*[a]	*Syagrus coronata*	*Veitchia arecina*[a]
Total cpDNA size	155,829	130,807	154,731	156,973	157,683	155,053	157,194
Length of LSC region	84,265	110,702	84,230	85,192	85,522	84,535	85,647
Length of IR region	27,092	1426	26,555	27,071	27,220	26,522	27,050
Length of SSC	17,380	17,253	17,391	17,639	17,721	17,474	17,447
GC content (%)	37.53	36.58	37.44	37.40	37.66	37.46	37.47

[a] The values are an approximation because the nucleotide sequences of these plastomes are not complete in the GenBank database

A complete description and location of all SSRs identified in the plastome of macaw palm is shown in Supplementary Table S3. Among the 221 SSR loci identified, 123 are located in intergenic spacers (IGSs), 61 in coding sequences (CDSs), 34 in introns, and three in tRNA genes. The CDSs with higher number of SSRs are *ycf1* (18 SSRs), *ycf2* (7 SSRs), and *rpoC2* (7 SSRs) genes. In addition, the *ndhC*/*trnV*-UAC (10 SSRs) and *ndhF*/*rpl32* (6 SSRs) IGSs, and the introns of *clpP* (8 SSRs) and *trnK*-UUU (7 SSRs) genes contain a high number of SSRs. A comparison of SSRs found here with SSRs identified in the plastome of *E. guineensis* (the closest taxon sampled to *A. aculeata*,

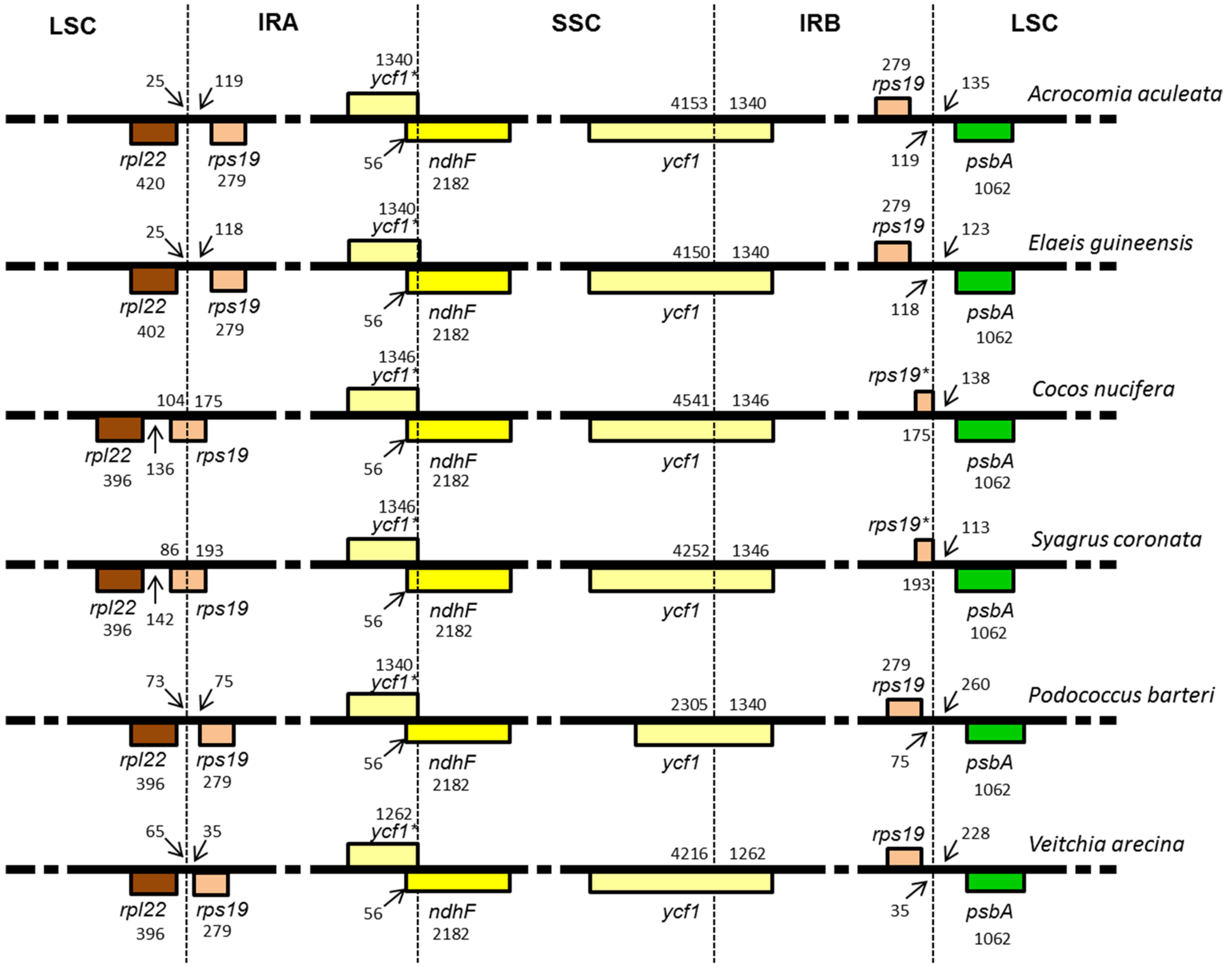


Fig. 2 Comparison of the IRA and IRB borders among Arecoideae species. The numbers indicate the lengths of IGSs, genes, and spacers between IR-LSC and IR-SSC junctions. The *ycf1** and *rps19** genes have incomplete CDSs

according to our phylogenomic tree showed below) showed that 49% of the SSRs located in the IGSs and introns are polymorphic. Only four polymorphic SSRs (all located in the *ycf1* gene) are located in CDS (Supplementary Table S3).

Moreover, we performed a nucleotide divergence analysis to identify the regions with higher polymorphism within the subfamily Arecoideae and tribe Cocoseae, in order to select putative fast-evolving plastid sequences in macaw palm plastome (Fig. 4). The *trnC*-GCA/*petN* and *psaC*/*ndhE* IGSs, and the CDS of *ycf1* gene (part included in the SSC region) are shared between the subfamily Arecoideae and the tribe Cocoseae as the regions with the highest nucleotide divergence. Within the subfamily Arecoideae, other three regions of high polymorphism are located in the junctions IRB-LSC [*rps19*/*psbA* (IGS)] and LSC/IRA [*rpl22*/*rps19* (IGS)], and in the IGS between the *accD* and *psaI* genes. Furthermore, aiming to detect more specific polymorphism hotspots, we also performed the sliding window analysis aligning the plastomes of macaw palm and *E. guineensis* (Fig. 4). As the previous analyses revealed for the subfamily Arecoideae and tribe Cocoseae, the plastome regions *trnC*-GCA/*petN* (IGS), *psaC*/*ndhE* (IGS), and the CDS of *ycf1* gene, show again the highest nucleotide divergence. In addition to them, other two regions (the *psbC*/*trnS*-UGA and *trnL*-UAG/*ccsA* IGSs) were found to be exclusive between *A. aculeata* and *E. guineensis*.

Identification of putative RNA editing sites in plastid protein-coding genes of macaw palm and other species within Arecoideae

The prediction of RNA editing sites in the subfamily Arecoideae was carried out based on PREP program and comparison with RNA editing sites validated by Huang et al.

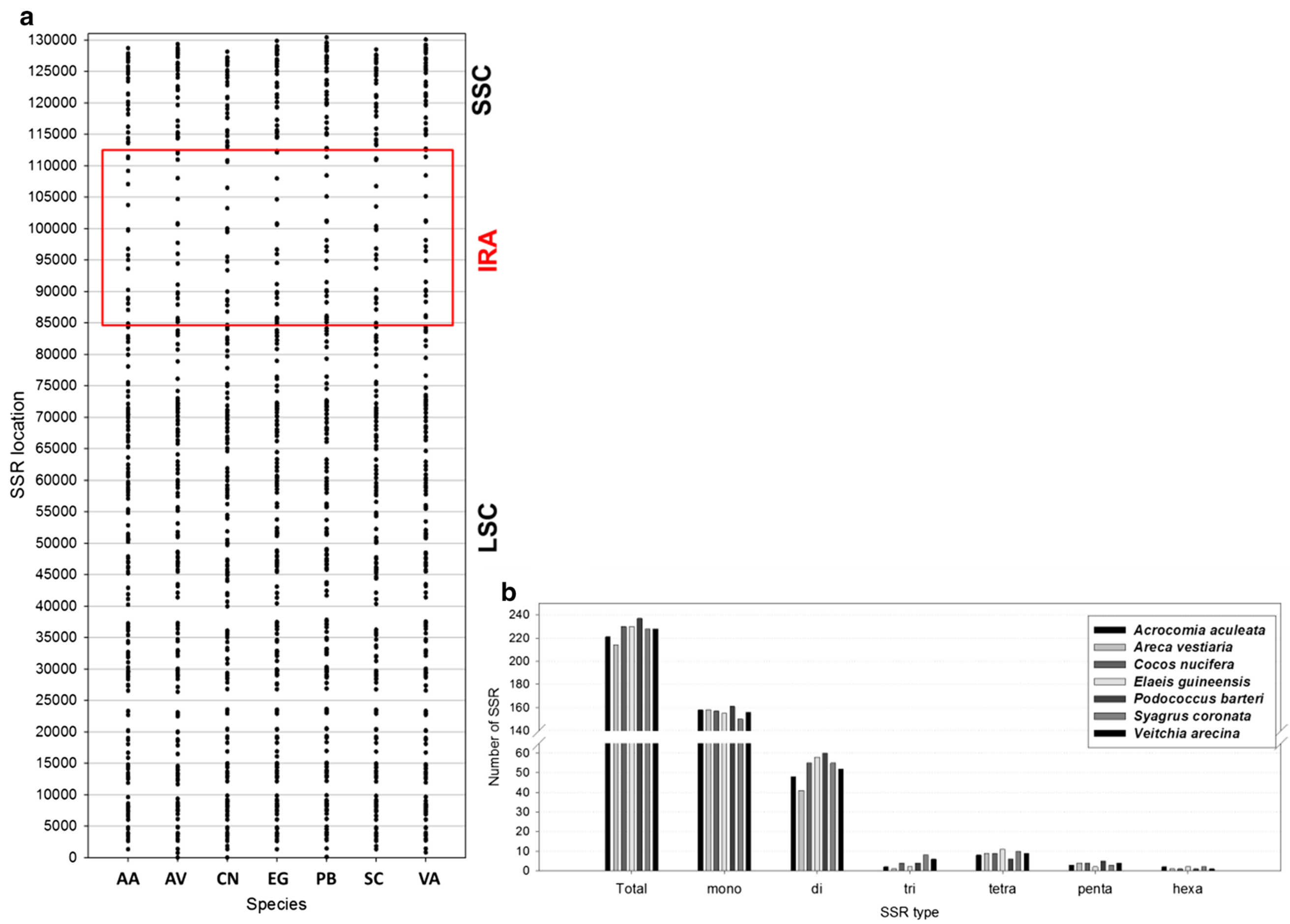


Fig. 3 SSR loci identification in Arecoideae plastomes (IRB omitted). **a** Distribuition of SSR loci. **b** Number of SSR loci. It was set eight repeat units for mononucleotide SSRs, four repeat units for di- and trinucleotide SSRs, and three repeat units for tetra-, penta- and hexanucleotide SSRs. Species sampled: *Acrocomia aculeata* (AA), *Areca vestiaria* (AV), *Cocos nucifera* (CN), *Elaeis guineensis* (EG), *Podococcus barteri* (PB), *Syagrus coronate* (SC), and *Veitchia arecina* (VA)

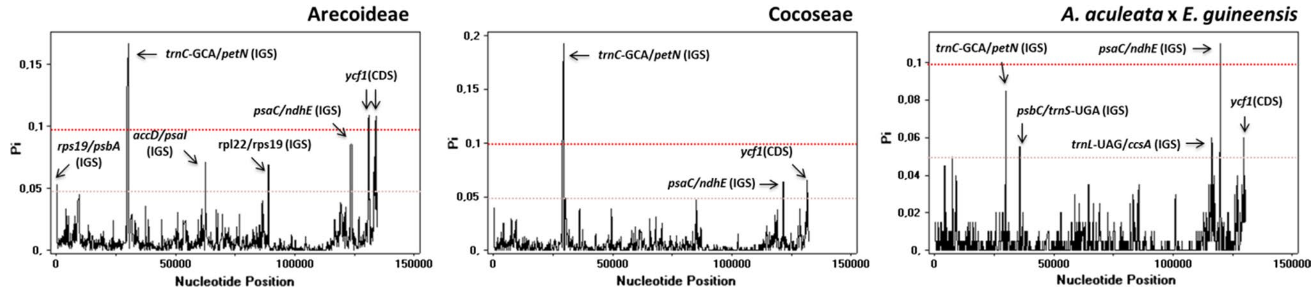


Fig. 4 Sliding window analyses of aligned whole plastomes (IRB omitted) within the subfamily Arecoideae, the tribe Cocoseae, and between *Acrocomia aculeata* and *Elaeis guineesis*. The regions with high nucleotide variability (*P*i > 0.05) are indicated. Pi, nucleotide diversity of each window. Window length 200 pb. Step size 50 pb

(2013) for some plastid protein-coding genes in *C. nucifera* (Supplementary Tables S4 and S5). All RNA editing sites identified are C-to-U conversions, at the first (24%) or second (76%) positions of the codons. A total of 100 RNA editing sites (distributed among 30 protein-coding genes) were predicted here, being 87 of them shared among all species sampled in this study. According to Huang et al. (2013), 74 out of 100 sites identified here are completely or partially edited in *C. nucifera*. Most RNA editing sites predicted here change the encoded amino acid from polar to apolar (62 out of 100). Only one putative RNA editing site changes the amino acid from apolar to polar (proline-serine), while the remaining do not change de polarity (20, polar-polar; 17, apolar-apolar).

Despite the high conservation of RNA editing sites among Arecoideae species, we also identified 13 sites that are not edited in one or more species. In addition, the RNA editing sites shared between this lineage show codon variants in 6 sites in which the RNA editing recovers the same conserved amino acid [e.g. in the site (97) of *ndhF* gene the RNA editing changes different codons TCA (S-L) and CCA (P-L) generating the same conserved amino acid]. All those codons (edited or unedited) that diverge in one or more species are shown in the Table 3. In order to identify if there are correlation between the Arecoideae classification and the evolution of RNA editing sites among the species analyzed here, we performed a Bayesian analysis using the divergent codons concatenated from different genes. The reconstructed tree (Fig. 5a) distinguishes the three tribes represented in the analysis, showing that the tribe Cocoseae is sister to the clade composed by the sister tribes Podococceae and Areceae.

Analyzing the 13 putative RNA editing sites, that are not edited in one or more species of Arecoideae, in the light of the Arecoideae phylogenetic tree (based on our phylogenomic analysis showed below), it is possible to suggest some events of gain and loss of RNA editing sites within the subfamily Arecoideae (Fig. 5b). The *accD* (429) and *ndhF* (131) sites were supposedly gained and lost, respectively, in the tribe Areceae given that these features are not shared with the tribes Podococceae (sister group) and Cocoseae. However, species-specific gains and losses are more frequent. In our analyses we count five gains [the *accD* (241), *accD* (487), *clpP* (118), matK (312), and *ndhF* (607) sites] and four losses [the *matK* (219), *ndhF* (148), *ndhH* (182), and *rpoB* (665) sites]. Lastly, the distribution pattern of editing at *accD* (389) and *ccsA* (274) sites among the species sampled here do not allow hypotheses about gains or losses in the subfamily Arecoideae.

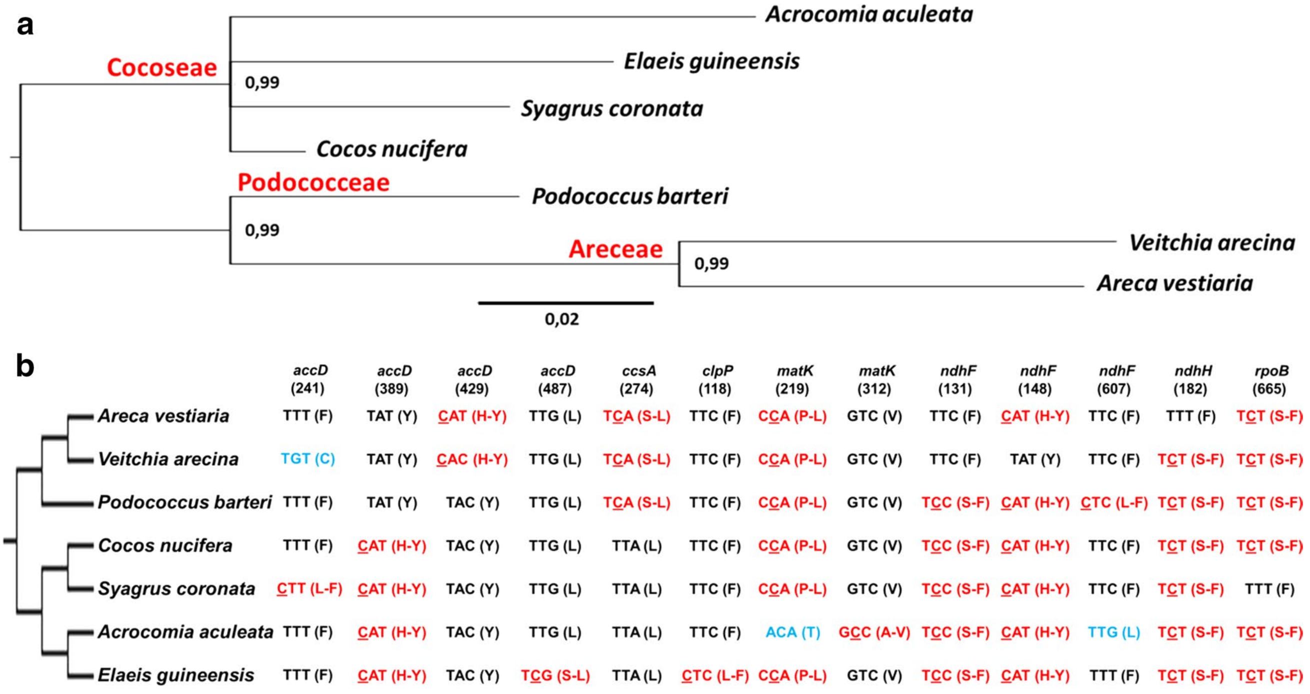
Arecaceae phylogenomic reconstruction based on whole plastomes

The Arecaceae phylogenomic was carried out using whole plastomes of 40 taxa (Supplementary Table S1), including 37 species representing all five subfamilies of Arecaceae: Calamoideae, Nypoideae, Arecoideae, Coryphoideae, and Ceroxyloideae (Dransfield et al. 2005; Asmussen et al. 2006). The three remaining taxa are two species of family Dasypogonaceae (most related to Arecaceae) and *H. malayana* (Hanguanaceae: Commelinales), which were used as outgroup. Bayesian inference (BI) analysis produced a

Table 3 List of RNA editing sites whose codons diverge in one or more species of Arecoideae subfamily

Gene	AA position	*Acrocomia aculeata*	*Elaeis guineensis*	*Syagrus coronata*	*Cocos nucifera*	*Podococcus barteri*	*Veitchia arecina*	*Areca vestiaria*
accD	241	UUU (F)	UUU (F)	CUU (L-F)	UUU (F)	UUU (F)	UGU (C)	UUU (F)
accD	389	CAU (H-Y)	CAU (H-Y)	CAU (H-Y)	CAU (H-Y)	UAU (Y)	UAU (Y)	UAU (Y)
accD	429	UAC (Y)	UAC (Y)	UAC (Y)	UAC (Y)	UAC (Y)	CAC (H-Y)	CAU (H-Y)
accD	487	UUG (L)	UCG (S-L)	UUG (L)	UUG (L)	UUG (L)	UUG (L)	UUG (L)
ccsA	274	UUA (L)	UUA (L)	UUA (L)	UUA (L)	UCA (S-L)	UCA (S-L)	UCA (S-L)
clpP	118	UUC (F)	CUC (L-F)	UUC (F)	UUC (F)	UUC (F)	UUC (F)	UUC (F)
matK	219	ACA (U)	CCA (P-L)	CCA (P-L)	CCA (P-L)	CCA (P-L)	CCA (P-L)	CCA (P-L)
matK	310	CAU (H-Y)	CAU (H-Y)	CAU (H-Y)	CAU (H-Y)	CAC (H-Y)	CAC (H-Y)	CAC (H-Y)
matK	312	GCC (A-V)	GUC (V)	GUC (V)	GUC (V)	GUC (V)	GUC (V)	GUC (V)
matK	426	CAC (H-Y)	CAC (H-Y)	CAC (H-Y)	CAC (H-Y)	CAC (H-Y)	CAC (H-Y)	CAU (H-Y)
ndhD	225	UCG (S-L)	UCG (S-L)	UCG (S-L)	UCG (S-L)	UCG (S-L)	UCA (S-L)	UCG (S-L)
ndhD	398	UCA (S-L)	UCA (S-L)	UCA (S-L)	UCA (S-L)	UCA (S-L)	UCG (S-L)	UCG (S-L)
ndhF	97	UCA (S-L)	UCA (S-L)	UCA (S-L)	UCA (S-L)	CCA (P-L)	UCA (S-L)	UCA (S-L)
ndhF	131	UCC (S-F)	UCC (S-F)	UCC (S-F)	UCC (S-F)	UCC (S-F)	UUC (F)	UUC (F)
ndhF	148	CAU (H-Y)	CAU (H-Y)	CAU (H-Y)	CAU (H-Y)	CAU (H-Y)	UAU (Y)	CAU (H-Y)
ndhF	607	UUG (L)	UUU (F)	UUC (F)	UUC (F)	CUC (L-F)	UUC (F)	UUC (F)
ndhF	698	UCU (S-F)	UCC (S-F)	UCC (S-F)	UCC (S-F)	UCC (S-F)	UCC (S-F)	UCC (S-F)
ndhH	182	UCU (S-F)	UCU (S-F)	UCU (S-F)	UCU (S-F)	UCU (S-F)	UCU (S-F)	UUU (F)
rpoB	665	UCU (S-F)	UCU (S-F)	UUU (F)	UCU (S-F)	UCU (S-F)	UCU (S-F)	UCU (S-F)

The cytidines marked are putatively edited to uridines. The letters in parentheses represent the encoded amino acids. The encoded amino acid after RNA editing is also showed



	accD (241)	accD (389)	accD (429)	accD (487)	ccsA (274)	clpP (118)	matK (219)	matK (312)	ndhF (131)	ndhF (148)	ndhF (607)	ndhH (182)	rpoB (665)
Areca vestiaria	TTT (F)	TAT (Y)	CAT (H-Y)	TTG (L)	TCA (S-L)	TTC (F)	CCA (P-L)	GTC (V)	TTC (F)	CAT (H-Y)	TTC (F)	TTT (F)	TCT (S-F)
Veitchia arecina	TGT (C)	TAT (Y)	CAC (H-Y)	TTG (L)	TCA (S-L)	TTC (F)	CCA (P-L)	GTC (V)	TTC (F)	TAT (Y)	TTC (F)	TCT (S-F)	TCT (S-F)
Podococcus barteri	TTT (F)	TAT (Y)	TAC (Y)	TTG (L)	TCA (S-L)	TTC (F)	CCA (P-L)	GTC (V)	TCC (S-F)	CAT (H-Y)	CTC (L-F)	TCT (S-F)	TCT (S-F)
Cocos nucifera	TTT (F)	CAT (H-Y)	TAC (Y)	TTG (L)	TTA (L)	TTC (F)	CCA (P-L)	GTC (V)	TCC (S-F)	CAT (H-Y)	TTC (F)	TCT (S-F)	TCT (S-F)
Syagrus coronata	CTT (L-F)	CAT (H-Y)	TAC (Y)	TTG (L)	TTA (L)	TTC (F)	CCA (P-L)	GTC (V)	TCC (S-F)	CAT (H-Y)	TTC (F)	TCT (S-F)	TTT (F)
Acrocomia aculeata	TTT (F)	CAT (H-Y)	TAC (Y)	TTG (L)	TTA (L)	TTC (F)	ACA (T)	GCC (A-V)	TCC (S-F)	CAT (H-Y)	TTG (L)	TCT (S-F)	TCT (S-F)
Elaeis guineensis	TTT (F)	CAT (H-Y)	TAC (Y)	TCG (S-L)	TTA (L)	CTC (L-F)	CCA (P-L)	GTC (V)	TCC (S-F)	CAT (H-Y)	TTT (F)	TCT (S-F)	TCT (S-F)

Fig. 5 RNA editing analysis. **a** Relationships within the subfamily Arecoideae based on concatenated codons using bayesian inference. The codons used were extracted from putative RNA editing sites that diverge in one or more species of Arecoideae. The tribes are highlighted in red. The bayesian posterior probability of all nodes is 0.99. The branch length is proportional to the inferred divergence level and the scale bar indicates the number of inferred nucleic acids substitutions per site. **b** Distribution of putative RNA editing sites across the Arecoideae phylogeny based on whole plastome. The codons highlighted in red are edited, and the codons in blue codify non-conserved amino acids

phylogenetic tree with a − lnL = 510,541.335 (Fig. 6) and a high branch support (BI posterior probability value of 1 for all nodes). Regarding the relationships among the subfamilies within Arecaceae, the subfamily Calaimoideae is sister to a clade composed by other four subfamilies. This clade contains the subfamily Nypoideae sister to a subclade where the subfamily Coryphoideae forms a sister-group with the subfamilies Arecoideae and Ceroxyloideae. Among the Arecoideae species sampled here, the tribe Cocoseae forms a sister-group with the tribes Podococceae and Areceae. The macaw palm (*A. aculeata*) is placed within the tribe Cocoseae and it is sister to *E. guineensis*.

Gene divergence analysis and identification of positive signatures in plastid protein-coding genes within Arecaceae

Overall, the gene divergence analysis indicates that the plastid protein-coding genes in Arecaceae are well conserved (Fig. 7a). The *ycf1* gene is the most divergent gene, followed by *rpl32* and *rps16* genes. These genes show extensive variation of branch length among the species given that some of them exhibit high divergent branches such as *Podococcus barteri* (*ycf1* and *rpl32*) and *Salacca ramosiana* (*rpl16*) (Supplementary Fig. S2). Other genes, such as *ccsA*, *rpl22*, *psaJ*, and *ndhK*, also occur in one or two species with a highlighted long branch in comparison with the remaining species with small branches (Supplementary Fig. S2).

To investigate whether any plastid gene of Arecaceae underwent positive selection we used the Selecton program. We identified a total of 283 putative positive signatures distributed in more than half of plastid genes (40 out of 79 protein-coding genes) (Fig. 7b). The positive selection shows a tendency to be related to gene divergence rate given that most divergent genes have more positive signatures. Positive signatures were identified in several genes involved in different essential functions such as photosynthesis [including subunits of PSI (*ycf3*, *ycf4* and *psa* genes), PSII (*psb* genes), ATP synthase (*atp* genes), cytochrome b_6f (*petD* gene), and ndh complex (*ndh* genes)], plastid gene expression [including subunits of RNA polymerase (*rpo* genes), RNA maturation (*matK* gene), and ribosomal proteins (*rpl* and *rps* genes)], fatty acid biosynthesis (*accD* gene), cytochrome biosynthesis (*ccsA* gene), import of protein (*ycf1* gene), uptake of inorganic carbon (*cemA* gene), and unknown function (*ycf2* gene).

In order to identify the evolutionary patterns across the palm family, we plotted the sites under positive selection against the phylogeny inferred from whole plastomes (Fig. 8; Supplementary Figs. S3–S9). Three evolutionary types are

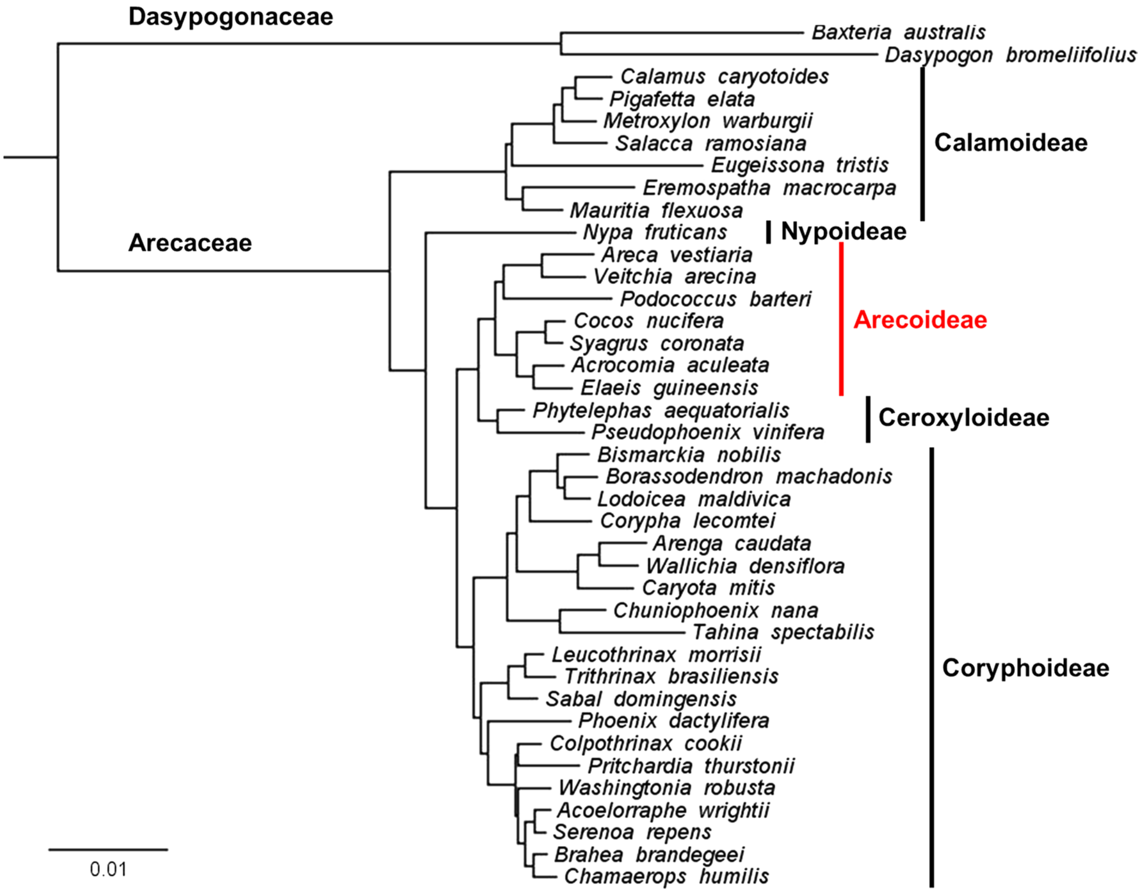


Fig. 6 Arecaceae phylogenomic tree of 40 taxa (37 Arecaceae species and 3 outgroups) based on whole plastomes using bayesian inference. The bayesian posterior probability of all nodes is 1. The branch length is proportional to the inferred divergence level and the scale bar indicates the number of inferred nucleic acids substitutions per site. *Hanguana malayana* (Commelinales) was used to root the tree (omitted from the figure)

noted in the sites of positive selection, including high heterogeneity of amino acid type (e.g. the sites 330, 664, and 686 of *ycf1* gene, Supplementary Fig. S5), unique amino acid type in a particular clade such as subfamily or tribe (e.g. the sites 300, 337, and 338 of *rpoA* gene; Supplementary Fig. S7), and the same amino acid in species or clades that are not closely related (e.g. the sites 89, 142, 219, and 251 of *rbcL* gene; Fig. 8). The latter pattern prevailed in comparison with other types, particularly among photosynthesis-related genes (Fig. 8).

Discussion

The general features of macaw palm plastome are conserved but small expansion and contraction events at the IR boundaries occurred during plastome evolution within Arecoideae

The general structure, the number and type of genes in the plastome of macaw palm are similar to most angiosperms, including other species of Arecaceae (Wicke et al. 2011; Uthaipaisanwong et al. 2012; Huang et al. 2013; Barrett et al. 2016). Nevertheless, at the IR-LSC and IR-SSC junctions we identify sequence variability, indicating the occurrence of small expansion and contraction events at the IR boundaries of Arecoideae plastomes. Similarly, sequence variabilities involving few hundreds of base pairs, 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 (Goulding et al. 1996; Zhu et al. 2016). Despite of these

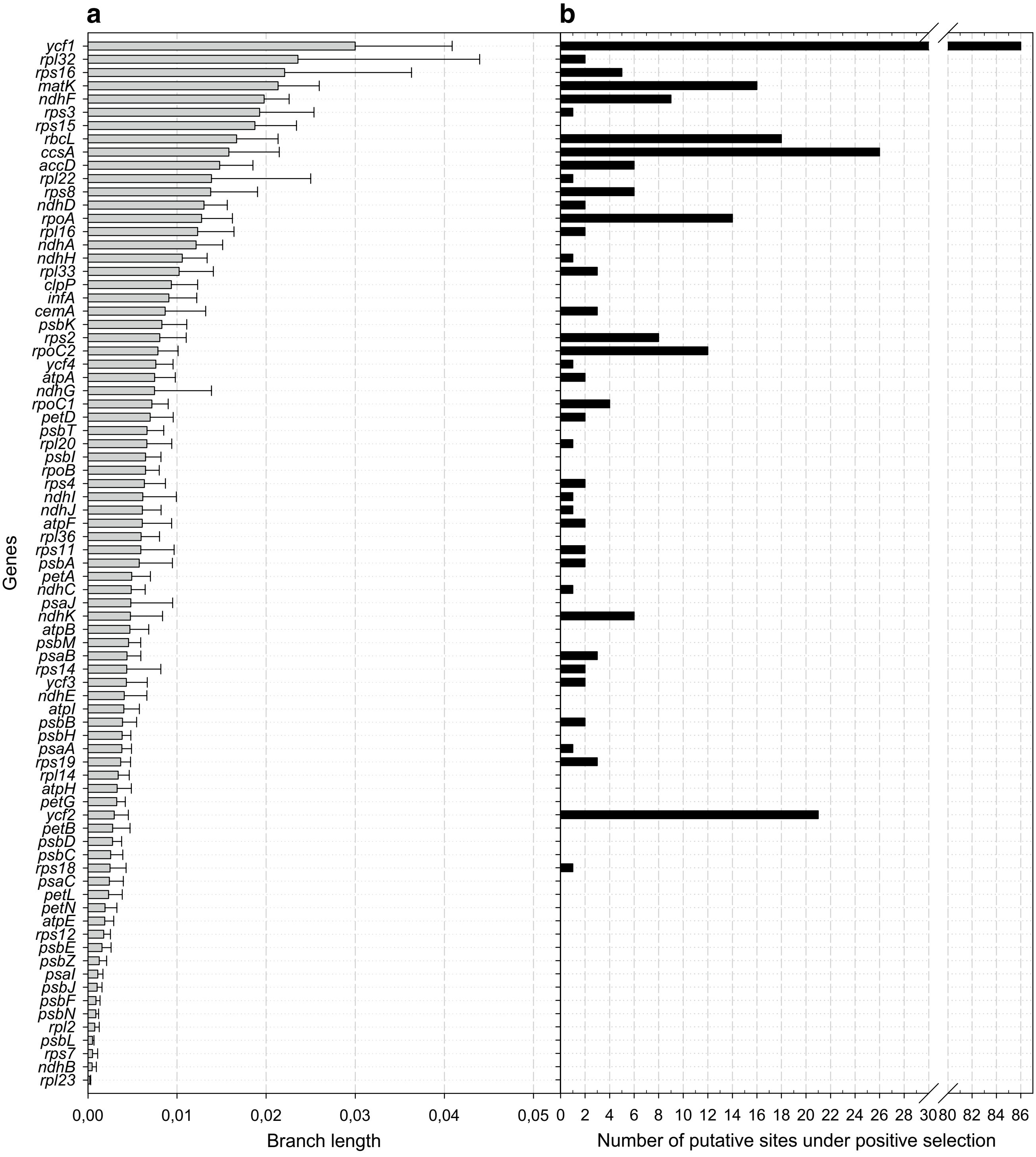
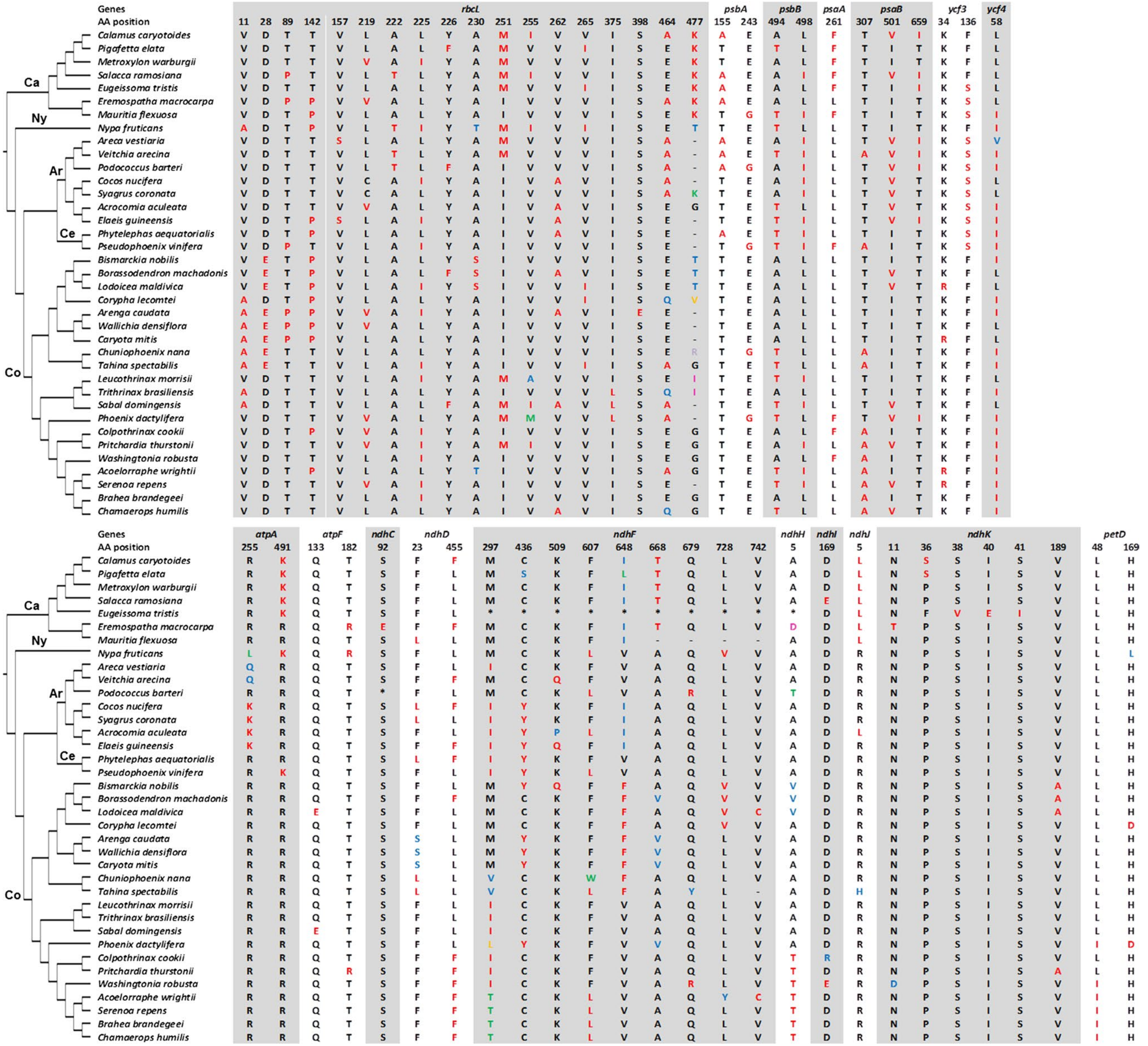


Fig. 7 Molecular evolution analyses of plastid genes within the family Arecaceae. **a** It is shown the divergence of protein-coding genes. The gene divergence was estimated by the sum of total branch lengths in each gene tree inferred (mean ± SD). **b** It is shown the number of putative sites under positive selection

slight differences at IR borders of most Arecoideae plastomes analyzed here, the extremely reduced IRs of *A. vestiaria* and *Tahina spectabilis* plastomes as described by Barrett et al. (2016) are likely the result of large-scale contraction event, which indicates that major changes may have occurred in others species of Arecoideae (Barrett et al. 2016). Large-scale expansion, contraction, and even complete loss of the IR have been reported for some plant lineages belonging to



Genes	rbcL																		psbA		psbB		psaA	psaB			ycf3		ycf4
AA position	11	28	89	142	157	219	222	225	226	230	251	255	262	265	375	398	464	477	155	243	494	498	261	307	501	659	34	136	58
Calamus caryotoides	V	D	T	T	V	L	A	L	Y	A	M	I	V	V	I	S	A	K	A	E	A	L	F	T	V	I	K	F	L
Pigafetta elata	V	D	T	T	V	L	A	L	F	A	M	V	V	I	I	S	E	K	T	E	T	L	F	T	I	T	K	F	L
Metroxylon warburgii	V	D	T	T	V	V	A	I	Y	A	M	V	V	V	I	S	E	K	T	E	A	L	F	T	I	T	K	F	L
Salacca ramosiana	V	D	P	T	V	L	T	L	Y	A	M	I	V	V	I	S	E	K	A	E	A	I	F	T	V	I	K	F	L
Eugeissoma tristis	V	D	T	T	V	L	A	L	Y	A	M	V	V	I	I	S	E	K	A	E	A	L	F	T	I	I	K	S	L
Eremospatha macrocarpa	V	D	P	P	V	V	A	L	Y	A	I	V	V	V	I	S	A	K	A	E	A	L	L	T	I	T	K	S	L
Mauritia flexuosa	V	D	T	P	V	L	A	L	Y	A	I	V	V	V	I	S	E	K	T	G	T	I	F	T	I	T	K	S	I
Nypa fruticans	A	D	T	P	V	L	T	I	Y	T	M	I	V	I	I	S	E	T	T	E	T	L	L	T	I	T	K	F	I
Areca vestiaria	V	D	T	T	S	L	A	L	Y	A	M	V	V	V	I	S	A	-	A	E	A	I	L	T	V	I	K	S	V
Veitchia arecina	V	D	T	T	V	L	T	L	Y	A	M	V	V	V	I	S	A	-	A	E	T	I	L	A	V	I	K	S	I
Podococcus barteri	V	D	T	T	V	L	T	L	F	A	I	V	V	V	I	S	A	-	A	G	A	I	L	T	V	I	K	S	I
Cocos nucifera	V	D	T	T	V	C	A	I	Y	A	I	V	A	V	I	S	A	-	T	E	A	I	L	T	V	T	K	S	L
Syagrus coronata	V	D	T	T	V	C	A	L	Y	A	I	V	V	V	I	S	A	K	T	E	A	I	L	T	V	T	K	S	L
Acrocomia aculeata	V	D	T	T	V	V	A	L	Y	A	I	V	A	V	I	S	E	G	T	E	T	L	L	T	V	T	K	S	I
Elaeis guineensis	V	D	T	P	S	L	A	I	Y	A	I	V	A	V	I	S	E	-	T	E	T	I	L	T	V	I	K	S	I
Phytelephas aequatorialis	V	D	T	P	V	L	A	L	Y	A	I	V	A	V	I	S	E	-	A	E	T	I	L	T	I	T	K	S	I
Pseudophoenix vinifera	V	D	P	T	V	L	A	I	Y	A	I	V	V	V	I	S	E	-	T	G	T	I	F	A	I	T	K	S	I
Bismarckia nobilis	V	E	T	P	V	L	A	L	Y	S	I	V	V	V	I	S	E	T	T	E	A	L	L	T	I	T	K	F	I
Borassodendron machadonis	V	E	T	P	V	L	A	L	F	S	I	V	A	V	I	S	E	T	T	E	A	L	L	T	V	T	K	F	L
Lodoicea maldivica	V	E	T	P	V	L	A	I	Y	S	I	V	V	I	I	S	E	T	T	E	A	L	L	T	V	T	R	F	L
Corypha lecomtei	A	D	T	P	V	L	A	L	Y	A	I	V	V	I	I	S	Q	V	T	E	A	L	L	T	I	T	K	F	I
Arenga caudata	A	E	P	P	V	V	A	I	Y	A	I	V	A	V	I	E	E	-	T	E	A	L	L	T	I	T	K	F	I
Wallichia densiflora	A	E	P	P	V	V	A	L	Y	A	I	V	V	V	I	S	E	-	T	E	A	L	L	T	I	T	K	F	L
Caryota mitis	A	E	P	P	V	L	A	L	Y	A	I	V	V	V	I	S	E	-	T	E	A	L	L	T	I	T	R	F	L
Chuniophoenix nana	A	E	T	T	V	L	A	L	Y	A	I	V	V	V	I	S	E	R	T	G	T	L	L	A	I	T	K	F	I
Tahina spectabilis	A	E	T	T	V	L	A	I	Y	A	I	V	V	I	I	S	A	G	T	E	T	L	L	A	I	T	K	F	I
Leucothrinax morrisii	V	D	T	T	V	L	A	I	Y	A	M	A	V	V	I	S	E	I	T	E	T	I	L	T	I	T	K	F	L
Trithrinax brasiliensis	A	D	T	T	V	L	A	L	Y	A	I	V	V	V	L	S	Q	I	T	E	A	L	L	T	I	T	K	F	I
Sabal domingensis	A	D	T	T	V	L	A	L	F	A	M	I	A	V	L	S	A	-	T	E	T	I	L	T	V	T	K	F	L
Phoenix dactylifera	V	D	T	T	V	V	A	L	Y	A	M	M	V	V	L	S	A	-	T	G	T	L	F	T	V	I	K	F	I
Colpothrinax cookii	V	D	T	P	V	V	A	I	Y	A	I	V	V	V	I	S	E	G	T	E	A	L	F	A	I	T	K	F	I
Pritchardia thurstonii	V	D	T	T	V	V	A	I	Y	A	M	I	V	V	I	S	E	G	T	E	A	I	L	A	V	T	K	F	I
Washingtonia robusta	V	D	T	T	V	L	A	I	Y	A	I	V	V	V	I	S	E	G	T	E	A	L	F	A	I	T	K	F	I
Acoelorraphe wrightii	V	D	T	P	V	L	A	L	Y	T	I	V	V	V	I	S	A	G	T	E	T	I	L	A	I	T	R	F	I
Serenoa repens	V	D	T	T	V	V	A	I	Y	A	I	V	V	V	I	S	E	-	T	E	T	I	L	A	V	T	R	F	I
Brahea brandegeei	V	D	T	T	V	L	A	I	Y	A	I	V	V	V	I	S	E	G	T	E	A	L	L	A	I	T	K	F	I
Chamaerops humilis	V	D	T	T	V	L	A	L	Y	A	I	V	A	V	I	S	Q	G	T	E	T	L	L	A	V	T	K	F	I

Genes	atpA		atpF		ndhC	ndhD		ndhF									ndhH	ndhI	ndhJ	ndhK						petD	
AA position	255	491	133	182	92	23	455	297	436	509	607	648	668	679	728	742	5	169	5	11	36	38	40	41	189	48	169
Calamus caryotoides	R	K	Q	T	S	F	F	M	C	K	F	I	T	Q	L	V	A	D	L	N	S	S	I	S	V	L	H
Pigafetta elata	R	K	Q	T	S	F	L	M	S	K	F	L	T	Q	L	V	A	D	L	N	S	S	I	S	V	L	H
Metroxylon warburgii	R	K	Q	T	S	F	L	M	C	K	F	I	T	Q	L	V	A	D	L	N	P	S	I	S	V	L	H
Salacca ramosiana	R	K	Q	T	S	F	L	M	C	K	F	I	T	Q	L	V	A	E	L	N	P	S	I	S	V	L	H
Eugeissoma tristis	R	K	Q	T	S	F	L	*	*	*	*	*	*	*	*	*	*	D	L	N	F	V	E	I	V	L	H
Eremospatha macrocarpa	R	R	Q	R	E	F	F	M	C	K	F	I	T	Q	L	V	D	D	L	T	P	S	I	S	V	L	H
Mauritia flexuosa	R	R	Q	T	S	L	L	M	C	K	F	I	-	-	-	-	A	D	L	N	P	S	I	S	V	L	H
Nypa fruticans	L	K	Q	R	S	F	L	M	C	K	L	V	A	Q	V	V	A	D	R	N	P	S	I	S	V	L	L
Areca vestiaria	Q	R	Q	T	S	F	L	I	C	K	F	V	A	Q	L	V	A	D	R	N	P	S	I	S	V	L	H
Veitchia arecina	Q	R	Q	T	S	F	F	M	C	Q	F	V	A	Q	L	V	A	D	R	N	P	S	I	S	V	L	H
Podococcus barteri	R	R	Q	T	*	F	L	M	C	K	L	V	A	R	L	V	T	D	R	N	P	S	I	S	V	L	H
Cocos nucifera	K	R	Q	T	S	L	F	I	Y	K	F	I	A	Q	L	V	A	D	R	N	P	S	I	S	V	L	H
Syagrus coronata	K	R	Q	T	S	L	L	I	Y	K	F	I	A	Q	L	V	A	D	R	N	P	S	I	S	V	L	H
Acrocomia aculeata	K	R	Q	T	S	F	L	I	Y	P	L	I	A	Q	L	V	A	D	L	N	P	S	I	S	V	L	H
Elaeis guineensis	K	R	Q	T	S	F	F	I	Y	Q	F	I	A	Q	L	V	A	D	R	N	P	S	I	S	V	L	H
Phytelephas aequatorialis	R	R	Q	T	S	L	F	I	Y	K	F	V	A	Q	L	V	A	D	R	N	P	S	I	S	V	L	H
Pseudophoenix vinifera	R	K	Q	T	S	F	L	I	Y	K	L	V	A	Q	L	V	A	D	R	N	P	S	I	S	V	L	H
Bismarckia nobilis	R	R	Q	T	S	F	L	M	Y	Q	F	F	A	Q	V	V	V	D	R	N	P	S	I	S	A	L	H
Borassodendron machadonis	R	R	Q	T	S	F	F	M	C	K	F	F	V	Q	V	V	V	D	R	N	P	S	I	S	A	L	H
Lodoicea maldivica	R	R	E	T	S	F	L	M	C	K	F	F	A	Q	V	C	V	D	R	N	P	S	I	S	A	L	H
Corypha lecomtei	R	R	Q	T	S	F	L	M	C	K	F	F	A	Q	V	V	A	D	R	N	P	S	I	S	V	L	D
Arenga caudata	R	R	Q	T	S	S	L	M	Y	K	F	F	V	Q	L	V	A	D	R	N	P	S	I	S	V	L	H
Wallichia densiflora	R	R	Q	T	S	S	L	M	Y	K	F	F	V	Q	L	V	A	D	R	N	P	S	I	S	V	L	H
Caryota mitis	R	R	Q	T	S	S	L	M	Y	K	F	F	V	Q	L	V	A	D	R	N	P	S	I	S	V	L	H
Chuniophoenix nana	R	R	Q	T	S	L	L	V	C	K	W	F	A	Q	L	V	A	D	R	N	P	S	I	S	V	L	H
Tahina spectabilis	R	R	Q	T	S	L	L	V	C	K	L	F	A	Y	L	-	A	D	H	N	P	S	I	S	V	L	H
Leucothrinax morrisii	R	R	Q	T	S	F	L	I	C	K	F	V	A	Q	L	V	A	D	R	N	P	S	I	S	V	L	H
Trithrinax brasiliensis	R	R	Q	T	S	F	L	I	C	K	F	V	A	Q	L	V	A	D	R	N	P	S	I	S	V	L	H
Sabal domingensis	R	R	E	T	S	F	L	I	C	K	F	V	A	Q	L	V	A	D	R	N	P	S	I	S	V	L	H
Phoenix dactylifera	R	R	Q	T	S	F	L	L	Y	K	F	V	V	Q	L	V	A	D	R	N	P	S	I	S	V	I	D
Colpothrinax cookii	R	R	Q	T	S	F	F	I	C	K	F	V	A	Q	L	V	T	R	R	N	P	S	I	S	V	L	H
Pritchardia thurstonii	R	R	Q	R	S	F	F	I	C	K	F	V	A	Q	L	V	T	D	R	N	P	S	I	S	A	L	H
Washingtonia robusta	R	R	Q	T	S	F	F	I	C	K	F	V	A	R	L	V	T	E	R	D	P	S	I	S	V	I	H
Acoelorraphe wrightii	R	R	Q	T	S	F	F	T	C	K	L	V	A	Q	Y	C	T	D	R	N	P	S	I	S	V	I	H
Serenoa repens	R	R	Q	T	S	F	F	T	C	K	L	V	A	Q	L	V	T	D	R	N	P	S	I	S	V	I	H
Brahea brandegeei	R	R	Q	T	S	F	F	T	C	K	L	V	A	Q	L	V	T	D	R	N	P	S	I	S	V	I	H
Chamaerops humilis	R	R	Q	T	S	F	F	T	C	K	L	V	A	Q	L	V	T	D	R	N	P	S	I	S	V	I	H

Fig. 8 Sites under positive selection in plastid genes involved in the photosynthetic process. The amino acids are plotted across the palm phylogeny based on whole plastomes. Different amino acid types identified at the same position are highlighted in distinct colors. The amino acid positions are relative to macaw palm plastid genes

the families Campanulaceae (Haberle et al. 2008), Fabaceae (Cai et al. 2008; Guisinger et al. 2011), Geraniaceae (Chumley 2006; Weng et al. 2014), Linaceae (Lopes et al. 2017) and Trochodendraceae (Sun et al. 2013).

The identification of SSR loci and polymorphism hotspots in the plastome of macaw palm represent special sources of molecular markers

Plastid SSRs represent potentially useful markers given that they have demonstrated high levels of intraspecific variability in several studies and due to the nonrecombinant and uniparental inheritance of the plastomes. Such SSRs have been applied in a broad range of researches focusing on studies of genetic diversity within and among natural populations, and characterization of evolutionary history of native and agricultural species (Provan et al. 2001; Ebert and Peakall 2009; Wheeler et al. 2014).

The SSRs located in IGS and introns are more interesting to use as molecular markers since these regions evolve faster than CDSs (Rogalski et al. 2015). In our study we found that 49% of the SSRs loci located in the IGSs and introns of macaw palm plastome are polymorphic when compared with *E. guineesis*. The *ndhC*/*trnV*-UAC (ten SSRs) and *ndhF*/*rpl32* (six SSRs) IGSs, and the introns of *clpP* (eight SSRs) and *trnK*-UUU (seven SSRs) genes contain a high number of SSRs and, therefore, configure important source of molecular markers. The SSRs detailed in the macaw palm plastome, mainly those located in the IGS and introns, represent important sequences to be used together with nuclear molecular markers developed in other groups (Mengistu et al. 2016a, b; Nucci et al. 2008) to study the genetic structure and genetic flow in natural populations of macaw palm (Wheeler et al. 2014). In addition, the SSRs found here are useful tools to characterize germplasm collections, promising genotypes and natural population, in order to search suitable strategies for genetic breeding of this wild species. It is also important to note that strategies using species-specific SSR primers have generated an elevated number of polymorphic loci than those employing universal primers (Wheeler et al. 2014), highlighting the importance of plastome sequencing.

The density of SSR loci was lower in the IRs in Arecoideae species. The lower number of SSR loci in the IRs may be correlated with the duplicative nature of the IRs, which enhances the copy-correction activity by gene conversion, resulting in lower evolutionary rate (Yamane et al. 2006; Zhu et al. 2016). Interestingly, the species *A. vestiaria*, that underwent massive reduction of the IRs, maintains a conserved pattern of SSR distribution in the SSC, LSC, and IR regions as found in other Arecoideae species, suggesting that the contraction of the IR region in *A. vestiaria* may be a recent event.

Lastly, we identify putative fast-evolving plastid sequences in macaw palm plastome based on sliding window analyses. Some IGSs at the level of subfamily and tribe were identified as polymorphism hotspots, especially the *trnC*-GCA/*petN* and *psaC*/*ndhE*. The polymorphism hotspots identified here, mainly in IGSs, represent interesting plastid markers to be used in genetic studies aiming the improvement of characteristics of the interest in macaw palm. Fast-evolving plastid IGSs have been used as efficient tools in several studies involving biogeography, genetic structure, genetic diversity, and germplasm characterization of important commercial species (Tsai et al. 2015; Wambulwa et al. 2016; Roy et al. 2016).

The evolution of putative RNA editing sites within the subfamily Arecoideae

The plastid RNA editing is a posttranscriptional modification (C-to-U and U-to-C conversions) identified in all groups of land plants, except in some species of Marchantiales. The mechanism of RNA editing has a monophyletic origin beginning with the evolution of land plants, when hundreds of editing sites were created in basal lineages. Following the evolution of higher plants several RNA editing sites were lost maintaining a relative conserved number of editing sites in angiosperms (Tillich et al. 2006; Takenaka et al. 2013; He et al. 2016). Plastid RNA editing sites have been identified in mRNA, introns, and untranslated regions, being implicated primordially as error correctors, but also acting in regulatory functions and producing variants of proteins to adapt to different physiological needs (Takenaka et al. 2013; Tseng et al. 2013; He et al. 2016; Chen et al. 2017).

In this study, a total of 100 RNA editing sites were predicted to occur within the subfamily Arecoideae. Like most angiosperms, all editing sites identified are C-to-U conversions, at the first or second position of the codons (Takenaka et al. 2013; He et al. 2016). RNA editing sites at the third position of the codon are less common, generally resulting in synonymous substitutions and having low frequency (He et al. 2016; Chen et al. 2017). From the total sites, 87 are shared among all species sampled in this study and 74 are completely or partially edited in *C. nucifera* (Huang et al., 2013), which suggest high conservation of the RNA editing mechanism within Arecoideae. Interestingly, most RNA editing sites predicted here change the encoded amino acid from polar to apolar, increasing the protein hydrophobicity which can affect structural features such as the creation of new transmembrane regions (He et al. 2016; Chen et al. 2017). Thereby, the general editing process is biased towards to increase the hydrophobicity of plastid proteins, which may be involved in protein–protein interactions and transmembrane domains present in the plastid protein complexes.

Despite the high conservation of RNA editing sites among species of Arecoideae, we identified 13 sites that are not edited in one or more species and six editing sites with codon variants. The phylogenetic tree reconstructed based on these divergent codons may suggest the mode of evolution of the RNA editing sites in the subfamily Arecoideae (Fig. 5a). The reconstructed tree distinguishes the three tribes represented in the analysis (Cocoseae, Podococceae, and Areceae) and the tribe relationships are in accordance with our Arecaceae phylogenomic analysis (discussed below) and Baker et al. (2009). However, the tree based on divergent RNA editing sites does not distinguish the relationships among different genera within the tribe Cocoseae. Therefore, the RNA editing evolution within the subfamily Arecoideae accumulated enough differences (gains, losses, and codon variations) to differentiate tribes but they are not enough to differentiate genera.

The loss of RNA editing sites in higher plants usually occurs when a mutation in DNA sequence changes the C to T, which results in a transcript that codifies the conserved amino acid without need for RNA editing. On the other hand, gains of RNA editing sites arise from the need to correct the mutations that change the T to C in DNA restoring the conserved amino acid (Kahlau et al. 2006; Hein et al. 2016). Our data suggest that at least six gains and five losses of RNA editing occurred within the subfamily Arecoideae, being species-specific events or events restricted to a closer taxa. Generally, the edited sites conserve the same amino acid among the species sampled here, except three RNA editing sites [*accD* (241), *matK* (219), and *ndhF* (607)] that showed some variability regarding the encoded amino acid. The editing sites that accept some variability is a common feature of plastid transcripts and it can occur in essential and nonessential genes (Fiebig et al. 2004). Indeed, nonessential *ndh* genes (Horváth et al. 2000; Li et al. 2004) have been reported as the most edited genes (Kahlau et al. 2006; He et al. 2016), including in the analysis performed here (37% of the editing sites; Supplementary Tables S4 and S5).

The number of RNA editing sites predicted here for protein-coding genes in palm species is high in comparison with others monocots, as rice (21 sites), maize (26 sites), and orchids (79 sites) (Corneille et al. 2000; Chen et al. 2017). Therefore, future analyses will be important to validate these putative sites identified in Arecoideae species. In addition, comparisons within Arecaceae and other related families may be able to decipher the origin of this unusual high number of RNA editing sites found here.

Arecaceae phylogenomic reconstruction based on whole plastomes

The placement of subfamilies within Arecaceae in our phylogenomic tree was similar to reported by Barrett et al. (2016) based on maximum likelihood analyses using whole plastomes and plastid protein-coding genes. The placement of the tribes within the subfamilies Calamoideae and Coryphoideae is also in accordance with Barrett et al. (2016). However, some incongruences related to the placement of tribes within the subfamily Arecoideae among phylogenies based on nuclear genes (Baker et al. 2011), plastid genes (Comer et al. 2015), and supertree method (Baker et al. 2009) are observed. In our tree, the relationships within Arecoideae are in accordance with Baker et al. (2009). Among Arecoideae species, the macaw palm (*A. aculeata*) is placed within the tribe Cocoseae closed to the genus *Elaeis*, in accordance with Baker et al. (2011).

Molecular evolution of plastid protein-coding genes within Arecaceae

The general structure and gene content of most plastomes of Arecaceae sequenced to date are conserved and a low substitution rate in DNA was observed within Arecaceae in comparison with other commelinids (Barrett et al. 2016). Nevertheless, some exceptions raise questions concerning the plastome evolution of palms. Barrett et al. (2016) reported massive IR loss and a 1944-bp inversion in the LSC in *T. spectabilis* (Coryphoideae), and unusual loss of gene (*ndhF*) and pseudogenization of several genes (*ndhA*, *ndhH*, and *ndhG*) in *Eugeissoma tristis* (Calamoideae). In addition, as previously mentioned, the species *A. vestiaria* (Arecoideae) also lost most part of the IRs (Supplementary Fig. S1). These structural changes are hypothesized to occur idiosyncratically in species or populations instead as synapomorphies in the palm plastid phylogeny (Barrett et al. 2016). Thus, we performed analyses to assess the gene divergence and to investigate the presence of positive selection in each protein-coding gene of palm plastomes to provide insights into the evolution of palm plastid genes. Our gene divergence analysis shows that some genes present wide variation of branch length (e.g. *ycf1*, *rpl32*, and *rps16* genes) because one or two species have notable long branches while the remaining species appear with small branches. In the case of exceptional genes containing nonconserved structure, the particular high gene divergence of some isolated species could be related to gene degeneration process in a species-specific manner rather than a feature shared by a specific clade, subfamily or tribe. Such hypotheses of species-level idiosyncrasies are in accordance with species-level supertree that suggests recent increase and heterogeneity in the diversification rate within genera of family Arecaceae (Faurby et al. 2016).

Moreover, we investigated whether any plastid gene of Arecaceae underwent positive selection. Unexpectedly, we identified a total of 283 putative positive signatures distributed in more than half of plastid genes (40 out of 79 protein-coding genes), including genes related to essential functions such as photosynthesis, plastid gene expression, and fatty acid biosynthesis. In comparison with other monocots, one-third of the plastid genes were reported to evolve under positive selection among species of grasses (Piot et al. 2017) suggesting that positive selection is strongly acting on species of Arecaceae. The occurrence of positive signatures across the palm phylogeny shows frequently the same amino acid change in species or clades that are not closely related. This evolutionary pattern prevailed in comparison with other types among photosynthesis-related genes, which can indicate convergent evolution associated with environmental conditions being shared among unrelated species. Recently, an emblematic case was reported in different lineages of C_4 grasses that show a convergent evolution of several sites in the *rbcL* gene, which underwent positive selection (Piot et al. 2017). Within Arecaceae, most species grow on shaded lower strata of tropical rain forest and their shade adaptation has been correlated with convergent evolution towards high net carbon gain efficiency among different lineages (Ma et al. 2015). It was also recently reported the adaptation of palm species to dry environmental conditions. Bacon et al. (2017) reported recently the shifting of some palm species from tropical rain forest to dry habitats, which is correlated with morphological adaptations. Here, we showed that several sites in essential genes for photosynthesis (*rbcL*, *psbA*, *psbB*, *psaA*, and *psaB*) are presumably under positive selection. However, the correlation of these sites with the habitats of palm species remains unknown and the existence of specific selective pressures on plastid genes for adaptation to different environmental conditions have to be investigated.

Among photosynthesis-related genes, the *rbcL* gene (which encodes de large subunit of rubisco enzyme) has the highest number of putative positive signatures (Fig. 8). According to Conserved Domains Database (https://www.ncbi.nlm.nih.gov/cdd), the sites 157, 225, 226, and 230 are related to the protein–protein interaction on the heterodimer interface. In addition, the sites 464 and 477 are part of the C-terminal strand that acts on the closing/opening mechanism of the active site and, therefore, they may be important for the catalytic activity (Burisch et al. 2007). Other essential gene for the photosynthesis is the *psbA* gene (encodes the D1 protein, subunit of the reaction center of the PSII). Two positive signatures were identified in this gene at the amino acid position 155, located in the region of chlorophyll biding site, and at the amino acid position 243, situated in the region involved in the protein–protein interaction on the D1-D2 interface (Conserved Domains Database). Positive selection in the amino acid position 155 was also reported among fern species where the evolution of *psbA* gene was related to the competition between ferns and angiosperms for light (Sen et al. 2012). Additionally, we found positive selection in other genes acting on the photochemical machinery, including the *psbB* (encodes the CP47 subunit of PSII), *psaA* (encodes a reaction center subunit of PSI), *psaB* (encodes a reaction center subunit of PSI), *ycf3* (encodes a PSI assembly factor), *ycf4* (encodes a PSI assembly factor), and *petD* (encodes the subunit IV of Cytochrome b_6f) genes. The selective pressures and the consequences of the changes for the photochemical functions are unknown, but positive selections were also identified in *psbB* and *petD* genes of the species of family Brassicaceae (Hu et al. 2015) and grasses (Piot et al. 2017), respectively. Finally, among the eleven *ndh* genes (encode subunits of the chloroplast Ndh complex) seven of them have positive signatures within Arecaceae. In grasses, nine *ndh* genes were found to evolve under positive selection (Piot et al. 2017). The Ndh complex acts in the minor pathway of PSI cyclic electron transport (Shikanai 2016), which is important for plant fitness under various stress conditions (Horváth et al. 2000; Li et al. 2004). The ndh genes have a specific dynamic of evolution, which includes high number of edition sites, pseudogenization, and gene losses or transfer to the nucleus (Martín and Sabater 2010). The selective pressures driving the conservation, positive selection (Fig. 8), or loss (e.g. *E. tristis*) of the Ndh complex within Arecaceae are a matter to be investigated given that it can be an adaptation to different environmental conditions (e.g. light intensities and hydric stress). Similarly, it is also notorious the amount of genes involved in the gene expression machinery containing positive signatures. They include 3 *rpo* genes (*rpoA*, *rpoC1* and *rpoC2*), 5 *rpl* genes (*rpl16*, *rpl20*, *rpl22*, *rpl32* and *rpl33*), 9 *rps* genes (*rps2*, *rps3*, *rps4*, *rps8*, *rps11*, *rps14*, *rps16*, *rps18* and *rps19*), and the *matK* gene (Supplementary Fig. S3, S7, S8, and S9), which totalizes 18 genes. Several studies have reported accelerated evolutionary rates and positive selection on plastid genes involved in gene expression (Krawczyk and Sawicki 2013; Hu et al. 2015; Xu et al. 2015; Weng et al. 2016; Piot et al. 2017). However, the effect of these signatures on the protein function and the adaptive capacity are poorly understood.

The high number of genes containing positive signatures, the wide distribution of these signatures across palm phylogeny, and several cases of putative convergent evolution may be related to the recent increase of diversification rate of Arecaceae species (Faurby et al. 2016). However, we cannot discard the possibility that some putative positive signatures are RNA editing sites involved in the recovery of conserved amino acids. Comparing the RNA editing sites predicted among species of Arecoideae (Supplementary Tables S4 and S5) with the putative positive signatures of Arecaceae (Fig. 8, Supplementary Figs. S3–S9), we identified five positive signatures that could be RNA editing sites. These sites are located at the position 218 and 309 of *matK*, 273 of *ccsA*, 136 of *ycf3*, and 607 of *ndhF* genes, which correspond to the RNA editing sites 219, 310, 274, 138, and 607 in the respective genes in species of Arecoideae. The use of next-generation sequencing (NGS) to investigate RNA editing sites has expanded the number of new RNA editing sites (He et al. 2016; Hein et al. 2016; Chen et al. 2017), including positive pressures acting on the editing sites of some genes (He et al. 2016). Therefore, it will be interesting to carry out RNA editing studies based on NGS technology in Arecaceae to confirm putative positive signatures and new editing sites expanding our knowledge about the selective pressures acting on plastid genes of Arecaceae species.

Conclusions

Here we reported the complete plastome of macaw palm, which was carefully characterized regarding the gene content, structure and evolution. A total of 221 SSR loci were identified and localized along the plastome of macaw palm. Eight regions of high polymorphism (hotspots) at level of subfamily and tribe were determined, being the most divergent the *trnC*-GCA/*petN* and *psaC*/*ndhE* IGSs. In addition, we performed a phylogenomic analysis using whole plastomes of 40 taxa, including 37 species representing all five subfamilies of Arecaceae, which placed the macaw palm within the tribe Cocoseae closed to *Elaeis* genus. Moreover, the analysis of

RNA editing among Arecoideae species, including *A. aculeata*, indentified 100 putative sites within this subfamily and indicated possible events of gain and loss of editing sites within this subfamily. Furthermore, we analyzed extensively the molecular evolution of plastid genes within the family Arecaceae, covering the full set of plastid protein-coding genes. The data revealed the presence of highly divergent genes in a species-specific manner suggesting that gene degeneration processes may be occurring within Arecaceae at the level of genus and/or species. These analyses also revealed unexpectedly that more than half of all plastid protein-coding genes within Arecaceae are under positive selection, which can presumably affect essential plastid functions. The distribution of these positive signatures across the Arecaceae phylogenomic suggests convergent evolution of most sites, including genes involving in photosynthesis. Finally, the SSRs and polymorphism hotspots identified here increase significantly the genetic information available for boost genetic studies in natural populations and germplasm collections of macaw palm aiming domestication and genetic breeding. The findings showed here via molecular analyses have important implications in the areas of genetic, evolution, conservation, breeding, and biotechnology of macaw palm and other species of Arecaceae.

Author contribution statement ASL, TGP, LNV, MPG, EMS, FOP, and MR conceived and designed the research. ASL, TGP, TN, LNV, EMS, FOP, and MR conducted experiments and analyzed the data. MPG, RON, EMS, FOP, and MR contributed with reagents and materials. ASL and MR wrote the manuscript. All authors read and approved the manuscript.

Acknowledgements This research was supported by the National Council for Scientific and Technological Development, Brazil (CNPq, Grant 459698/2014-1). We are grateful to INCT-FBN and for the scholarships granted by the Brazilian Federal Agency for Support and Evaluation of Graduate Education (CAPES) to TGP and LNV, and those granted by the CNPq to ASL, MPG, RON, EMS and FOP. We would like to thank Dr. Sergio Y. Motoike (Professor—Universidade Federal de Viçosa) and Francisco de Assis Lopes (Technical Assistant—Universidade Federal de Viçosa) for their assistance with the collection of plant material for chloroplast isolation.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Alkatib S, Scharff LB, Rogalski M, Fleischmann TT, Matthes A, Seeger S et al (2012) The conributions of wobbling and superwobbling to the reading of the genetic code. PLoS Genet 8:e1003076. https://doi.org/10.1371/journal.pgen.1003076

Asmussen CB, Dransfield J, Deickmann V, Barfod AS, Pintaud JC, Baker WJ (2006) A new subfamily classification of the palm family (Arecaceae): evidence from plastid DNA phylogeny. Bot J Linn Soc 151:15–38

Bacon CD, Moraes RM, Jaramillo C, Antonelli A (2017) Endemic palm species shed light on habitat shifts and the assembly of the Cerrado and Restinga floras. Mol Phylogenet Evol 110:127–133. https://doi.org/10.1016/j.ympev.2017.03.013

Baker WJ, Savolainen V, Asmussen-Lange CB, Chase MW, Dransfield J, Forest F, Harley MM, Uhl NW, Wilkinson M (2009) Complete generic-level phylogenetic analyses of palms (Arecaceae) with comparisons of supertree and supermatrix approaches. Syst Biol 58:240–256. https://doi.org/10.1093/sysbio/syp021

Baker WJ, Norup MV, Clarkson JJ, Couvreur TLP, Dowe JL, Lewis CE, Pintaud JC, Savolainen V, Wilmot T, Chase MW (2011) Phylogenetic relationships among arecoid palms (Arecaceae: Arecoideae). Ann Bot 108:1417–1432. https://doi.org/10.1093/aob/mcr020

Barfod AS, Hagen M, Borchsenius F (2011) Twenty-five years of progress in understanding pollination mechanisms in palms (Arecaceae). Ann Bot 108:1503–1516. https://doi.org/10.1093/aob/mcr192

Barrett CF, Baker WJ, Comer JR, Conran JG, Lahmeyer SC, Leebens-Mack JH, Li J, Lim GS, Mayfield-Jones DR, Perez L et al (2016) Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots. New Phytol 209:855–870. https://doi.org/10.1111/nph.13617

Berton LHC, de Azevedo Filho JA, Siqueira WJ, Colombo CA (2013) Seed germination and estimates of genetic parameters of promising macaw palm (*Acrocomia aculeata*) progenies for biofuel production. Ind Crops Prod 51:258–266

Bock R (2017) Witnessing genome evolution: experimental reconstruction of endosymbiotic and horizontal gene transfer. Annu Rev Genet. https://doi.org/10.1146/annurev-genet-120215-035329 **(in press)**

Burisch C, Wildner GF, Schlitter J (2007) Bioinformatic tools uncover the C-terminal strand of Rubisco's large subunit as hot-spot for specificity-enhancing mutations. FEBS Lett 581:741–748. https://doi.org/10.1016/j.febslet.2007.01.043

Cai Z, Guisinger M, Kim HG, Ruck E, Blazier JC, McMurtry V, Kuehl JV, Boore J, Jansen RK (2008) Extensive reorganization of the plastid genome of trifolium subterraneum (Fabaceae) is associated with numerous repeated sequences and novel DNA insertions. J Mol Evol 67:696–704. https://doi.org/10.1007/s00239-008-9180-7

Chen TC, Liu YC, Wang X, Wu CH, Huang CH, Chang CC (2017) Whole plastid transcriptomes reveal abundant RNA editing sites and differential editing status in *Phalaenopsis aphrodite* subsp. formosana. Bot Stud 58:38. https://doi.org/10.1186/s40529-017-0193-7

Chumley TW (2006) The complete chloroplast genome sequence of Pelargonium × hortorum: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. Mol Biol Evol 23:2175–2190

Ciconini G, Favaro SP, Roscoe R, Miranda CHB, Tapeti CF, Miyahira MAM, Bearari L, Galvani F, Borsato AV, Colnago LA et al (2013) Biometry and oil contents of *Acrocomia aculeata* fruits from the Cerrados and Pantanal biomes in Mato Grosso do Sul, Brazil. Ind Crops Prod 45:208–214

Comer JR, Zomlefer WB, Barrett CF, Davis JI, Stevenson DW, Heyduk K, Leebens-Mack JH (2015) Resolving relationships within the palm subfamily Arecoideae (Arecaceae) using plastid sequences derived from next-generation sequencing. Am J Bot 102:888–899. https://doi.org/10.3732/ajb.1500057

Conceição LDHCS, Antoniassi R, Junqueira NTV, Braga MF, de Faria-Machado AF, Rogério JB, Duarte ID, Bizzo HR (2015) Genetic diversity of macauba from natural populations of Brazil. BMC Res Notes 8:406. https://doi.org/10.1186/s13104-015-1335-1

Corneille S, Lutz K, Maliga P (2000) Conservation of RNA editing between rice and maize plastids: are most editing events dispensable? Mol Gen Genet MGG 264:419–424

Coser SM, Motoike SY, Corrêa TR, Pires TP, Resende MDV (2016) Breeding of *Acrocomia aculeata* using genetic diversity parameters and correlations to select accessions based on vegetative, phenological, and reproductive characteristics. Genet Mol Res. https://doi.org/10.4238/gmr15048820

Daniell H, Lin CS, Yu M, Chang WJ (2016) Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol 17:134. https://doi.org/10.1186/s13059-016-1004-2

Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. Nat Methods 9:772. https://doi.org/10.1038/nmeth.2109

Dransfield J, Uhl NW, Lange CBA, Baker WJ, Harley MM, Lewis CE (2005) A new phylogenetic classification of the palm family, Arecaceae. Kew Bull 60:559–569

Ebert D, Peakall R (2009) Chloroplast simple sequence repeats (cpSSRs): technical resources and recommendations for expanding cpSSR discovery and applications to a wide array of plant species. Mol Ecol Resour 9:673–690. https://doi.org/10.1111/j.1755-0998.2008.02319.x

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340

Faurby S, Eiserhardt WL, Baker WJ, Svenning JC (2016) An all-evidence species-level supertree for the palms (Arecaceae). Mol Phylogenet Evol 100:57–69. https://doi.org/10.1016/j.ympev.2016.03.002

Fiebig A, Stegemann S, Bock R (2004) Rapid evolution of RNA editing sites in a small non-essential plastid gene. Nucleic Acids Res 32:3615–3622. https://doi.org/10.1093/nar/gkh695

Fuentes P, Armarego-Marriott T, Bock R (2017) Plastid transformation and its application in metabolic engineering. Curr Opin Biotechnol 49:10–15. https://doi.org/10.1016/j.copbio.2017.07.004

Gonçalves DB, Batista AF, Rodrigues MQRB, Nogueira KMV, Santos VL (2013) Ethanol production from macaúba (*Acrocomia aculeata*) presscake hemicellulosic hydrolysate by *Candida boidinii* UFMG14. Bioresour Technol 146:261–266. https://doi.org/10.1016/j.biortech.2013.07.075

Goulding SE, Olmstead RG, Morden CW, Wolfe KH (1996) Ebb and flow of the chloroplast inverted repeat. Mol Gen Genet 252:195–206

Guisinger MM, Kuehl JV, Boore JL, Jansen RK (2011) Extreme reconfiguration of plastid genomes in the angiosperm family geraniaceae: rearrangements, repeats, and codon usage. Mol Biol Evol 28:583–600. https://doi.org/10.1093/molbev/msq229

Haberle RC, Fourcade HM, Boore JL, Jansen RK (2008) Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes. J Mol Evol 66:350–361. https://doi.org/10.1007/s00239-008-9086-4

He P, Huang S, Xiao G, Zhang Y, Yu J (2016) Abundant RNA editing sites of chloroplast protein-coding genes in *Ginkgo biloba* and an evolutionary pattern analysis. BMC Plant Biol 16(1):257. https://doi.org/10.1186/s12870-016-0944-8

Hein A, Polsakiewicz M, Knoop V (2016) Frequent chloroplast RNA editing in early-branching flowering plants: pilot studies on angiosperm-wide coexistence of editing sites and their nuclear specificity factors. BMC Evol Biol 16:23. https://doi.org/10.1186/s12862-016-0589-0

Henderson A, Galeano G, Bernal R (1995) Field guide to the palms of the Americas. Princeton University Press, Princeton

Horváth EM, Peter SO, Joët T, Rumeau D, Cournac L, Horváth GV, Kavanagh TA, Schäfer C, Peltier G, Medgyesy P (2000) Targeted inactivation of the plastid ndhB gene in tobacco results in an enhanced sensitivity of photosynthesis to moderate stomatal closure. Plant Physiol 123:1337–1350

Hu S, Sablok G, Wang B, Qu D, Barbaro E, Viola R, Li M, Varotto C (2015) Plastome organization and evolution of chloroplast genes in Cardamine species adapted to contrasting habitats. BMC Genomics 16:306. https://doi.org/10.1186/s12864-015-1498-0

Huang YY, Matzke AJM, Matzke M (2013) Complete sequence and comparative analysis of the chloroplast genome of coconut palm (*Cocos nucifera*). PLoS One 8:e74736. https://doi.org/10.1371/journal.pone.0074736

Kahlau S, Aspinall S, Gray JC, Bock R (2006) Sequence of the tomato chloroplast DNA and evolutionary comparison of solanaceous plastid genomes. J Mol Evol 63:194–207. https://doi.org/10.1007/s00239-005-0254-5

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010

Krawczyk K, Sawicki J (2013) The uneven rate of the molecular evolution of gene sequences of DNA-dependent RNA polymerase I of the genus *Lamium* L. Int J Mol Sci 14:11376–11391. https://doi.org/10.3390/ijms140611376

Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL (2004) Versatile and open software for comparing large genomes. Genome Biol 5:R12. https://doi.org/10.1186/gb-2004-5-2-r12

Lanes ÉCM, de Almeida Costa PM, Motoike SY (2014) Alternative fuels: Brazil promotes aviation biofuels. Nature 511:31. https://doi.org/10.1038/511031a

Lanes ÉCM, Motoike SY, Kuki KN, Nick C, Freitas RD (2015) Molecular characterization and population structure of the macaw palm, *Acrocomia aculeata* (Arecaceae), ex situ germplasm collection using microsatellites markers. J Hered 106:102–112. https://doi.org/10.1093/jhered/esu073

Lanes ÉCM, Motoike SY, Kuki KN, Resende MDV, Caixeta ET (2016) Mating system and genetic composition of the macaw palm (*Acrocomia aculeata*): implications for breeding and genetic conservation programs. J Hered 107:527–536. https://doi.org/10.1093/jhered/esw038

Li XG, Duan W, Meng QW, Zou Q, Zhao SJ (2004) The function of chloroplastic NAD(P)H dehydrogenase in tobacco during chilling stress under low irradiance. Plant Cell Physiol 45:103–108

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452. https://doi.org/10.1093/bioinformatics/btp187

Lohse M, Drechsel O, Kahlau S, Bock R (2013) OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. Nucleic Acids Res 41:W575–W581. https://doi.org/10.1093/nar/gkt289

Lopes AS, Pacheco TG, Santos KG, Vieira LN, Guerra MP, Nodari RO, Souza EM, Pedrosa FO, 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. https://doi.org/10.1007/s00299-017-2231-z **(in press)**

Lorenzi H, Kahn F, Noblick LR, Ferreira E (2010) Flora Brasileira—Arecaceae (Palmeiras). Instituto Plantarum, Nova Odessa, p 368

Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955–964

Luis ZG, Scherwinski-Pereira EJ (2014) An improved protocol for somatic embryogenesis and plant regeneration in macaw palm (*Acrocomia aculeata*) from mature zygotic embryos. Plant Cell Tissue Organ Cult 118:485–496. https://doi.org/10.1007/s11240-014-0500-x

Ma RY, Zhang JL, Cavaleri MA, Sterck F, Strijk JS, Cao KF (2015) Convergent evolution towards high net carbon gain efficiency contributes to the shade tolerance of palms (Arecaceae). PLoS One 10:e0140384. https://doi.org/10.1371/journal.pone.0140384

Machado W, Figueiredo A, Guimarães MF (2016) Initial development of seedlings of macauba palm (*Acrocomia aculeata*). Ind Crops Prod 87:14–19

Martín M, Sabater B (2010) Plastid ndh genes in plant evolution. Plant Physiol Biochem 48:636–645. https://doi.org/10.1016/j.plaphy.2010.04.009

Mengistu FG, Motoike SY, Caixeta ET, Cruz CD, Kuki KN (2016a) Cross-species amplification and characterization of new microsatellite markers for the macaw palm, *Acrocomia aculeata* (Arecaceae). Plant Genet Resour 14:163–172

Mengistu FG, Motoike SY, Cruz CD (2016b) Molecular characterization and genetic diversity of the macaw palm ex situ germplasm collection revealed by microsatellite markers. Diversity 8:20

Montoya SG, Motoike SY, Kuki KN, Couto AD (2016) Fruit development, growth, and stored reserves in macauba palm (*Acrocomia aculeata*), an alternative bioenergy crop. Planta 244:927–938. https://doi.org/10.1007/s00425-016-2558-7

Motoike SY, Kuki KN (2009) The potential of macaw palm (*Acrocomia aculeata*) as source of biodiesel in Brazil. IRECHE 1:632–635

Moura EF, Motoike SY, Ventrella MC, Sá AQ, Carvalho JM (2009) Somatic embryogenesis in macaw palm (*Acrocomia aculeata*) from zygotic embryos. Sci Hortic 119:447–454. https://doi.org/10.1016/j.scienta.2008.08.033

Mower JP (2009) The PREP suite: predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. Nucleic Acids Res 37:W253–W259. https://doi.org/10.1093/nar/gkp337

Nucci SM, Azevedo-Filho JA, Colombo CA, Priolli RHG, Coelho RM, Mata TL, Zucchi MI (2008) Development and characterization of microsatellites markers from the macaw. Mol Ecol Resour 8:224–226. https://doi.org/10.1111/j.1471-8286.2007.01932.x

Padilha JHD, Ribas LLF, Amano É, Quoirin M (2015) Somatic embryogenesis in *Acrocomia aculeata* Jacq. (Lodd.) ex Mart using the thin cell layer technique. Acta Bot Bras 29:516–523. https://doi.org/10.1590/0102-33062015abb0109

Park S, Ruhlman TA, Weng ML, Hajrah NH, Sabir JSM, Jansen RK (2017) Contrasting patterns of nucleotide substitution rates provide insight into dynamic evolution of plastid and mitochondrial genomes of geranium. Genome Biol Evol 9:1766–1780. https://doi.org/10.1093/gbe/evx124

Piot A, Hackel J, Christin PA, Besnard G (2017) One-third of the plastid genes evolved under positive selection in PACMAD grasses. Planta. https://doi.org/10.1007/s00425-017-2781-x

Pires TP, dos Santos Souza E, Kuki KN, Motoike SY (2013) Ecophysiological traits of the macaw palm: a contribution towards the domestication of a novel oil crop. Ind Crops Prod 44:200–210

Provan J, Powell W, Hollingsworth PM (2001) Chloroplast microsatellites: new tools for studies in plant ecology and evolution. Trends Ecol Evol 16:142–147

Rogalski M, Carrer H (2011) Engineering plastid fatty acid biosynthesis to improve food quality and biofuel production in higher plants: plastid fatty acid biosynthesis. Plant Biotechnol J 9:554–564. https://doi.org/10.1111/j.1467-7652.2011.00621.x

Rogalski M, Ruf S, Bock R (2006) Tobacco plastid ribosomal protein S18 is essential for cell survival. Nucleic Acids Res 34:4537–4545. https://doi.org/10.1093/nar/gkl634

Rogalski M, Schoettler MA, Thiele W, Schulze WX, Bock R (2008) Rpl33, a nonessential plastid encoded ribosomal protein in tobacco, is required under cold stress conditions. Plant Cell 20:2221–2237. https://doi.org/10.1105/tpc.108.060392

Rogalski M, do Nascimento Vieira L, Fraga HP, Guerra MP (2015) Plastid genomics in horticultural species: importance and applications for plant population genetics, evolution, and biotechnology. Front Plant Sci 6:586. https://doi.org/10.3389/fpls.2015.00586

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61:539–542. https://doi.org/10.1093/sysbio/sys029

Roy PS, Rao GJN, Jena S, Samal R, Patnaik A, Patnaik SSC, Jambhulkar NN, Sharma S, Mohapatra T (2016) Nuclear and chloroplast DNA Variation provides insights into population structure and multiple origin of native aromatic rices of Odisha, India. PloS One 11:e0162268. https://doi.org/10.1371/journal.pone.0162268

Ruhlman TA, Zhang J, Blazier JC, Sabir JSM, Jansen RK (2017) Recombination-dependent replication and gene conversion homogenize repeat sequences and diversify plastid genome structure. Am J Bot 104:559–572. https://doi.org/10.3732/ajb.1600453

Sen L, Fares M, Su YJ, Wang T (2012) Molecular evolution of psbA gene in ferns: unraveling selective pressure and co-evolutionary pattern. BMC Evol Biol 12:145. https://doi.org/10.1186/1471-2148-12-145

Shikanai T (2016) Chloroplast NDH: a different enzyme with a structure similar to that of respiratory NADH dehydrogenase. Biochim Biophys Acta 1857:1015–1022. https://doi.org/10.1016/j.bbabio.2015.10.013

Stern A, Doron-Faigenboim A, Erez E, Martz E, Bacharach E, Pupko T (2007) Selecton 2007: advanced models for detecting positive and purifying selection using a Bayesian inference approach. Nucleic Acids Res 35:W506–W511. https://doi.org/10.1093/nar/gkm382

Sun Y, Moore MJ, Meng A, Soltis PS, Soltis DE, Li J, Wang H (2013) Complete plastid genome sequencing of Trochodendraceae reveals a significant expansion of the inverted repeat and suggests a Paleogene divergence between the two extant species. PLoS One 8:e60429. https://doi.org/10.1371/journal.pone.0060429

Takenaka M, Zehrmann A, Verbitskiy D, Härtel B, Brennicke A (2013) RNA editing in plants and its evolution. Annu Rev Genet 47:335–352. https://doi.org/10.1146/annurev-genet-111212-133519

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. https://doi.org/10.1093/molbev/mst197

The Plant List (2013) Version 1.1. Retrieved from http://www.theplantlist.org/

Thiel T, Michalek W, Varshney R, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor Appl Genet 106:411–422. https://doi.org/10.1007/s00122-002-1031-0

Tillich M, Lehwark P, Morton BR, Maier UG (2006) The evolution of chloroplast RNA editing. Mol Biol Evol 23:1912–1921. https://doi.org/10.1093/molbev/msl054

Tsai CC, Chou CH, Wang HV, Ko YZ, Chiang TY, Chiang YC (2015) Biogeography of the *Phalaenopsis amabilis* species complex inferred from nuclear and plastid DNAs. BMC Plant Biol 15:202. https://doi.org/10.1186/s12870-015-0560-z

Tseng CC, Lee CJ, Chung YT, Sung TY, Hsieh MH (2013) Differential regulation of Arabidopsis plastid gene expression and RNA editing in non-photosynthetic tissues. Plant Mol Biol 82(4–5):375–392

Uthaipaisanwong P, Chanprasert J, Shearman JR, Sangsrakru D, Yoocha T, Jomchai N, Jantasuriyarat C, Tragoonrung S, Tangphatsornruang S (2012) Characterization of the chloroplast genome sequence of oil palm (*Elaeis guineensis* Jacq.). Gene 500:172–180. https://doi.org/10.1016/j.gene.2012.03.061

Vianna SA (2017) A new species of Acrocomia (Arecaceae) from Central Brazil. Phytotaxa 314:45–54. https://doi.org/10.11646/phytotaxa.314.1.2

Vieira LN, Faoro H, Fraga HPF, Rogalski M, de Souza EM, de Oliveira Pedrosa F, Nodari RO, Guerra MP (2014) An improved protocol for intact chloroplasts and cpdna isolation in conifers. PLoS One 9:e84792. https://doi.org/10.1371/journal.pone.0084792

Vieira LN, Dos Anjos KG, Faoro H, Fraga HP, Greco TM, Pedrosa FO, de Souza EM, Rogalski M, de Souza RF, Guerra MP (2016a) Phylogenetic inference and SSR characterization of tropical woody bamboos tribe Bambuseae (Poaceae: Bambusoideae) based on complete plastid genome sequences. Curr Genet 62(2):443–453. https://doi.org/10.1007/s00294-015-0549-z

Vieira LN, Rogalski M, Faoro H, Fraga HP, Anjos KG, Picchi GFA, Nodari RO, Pedrosa FO, Souza EM, Guerra MP (2016b) The plastome sequence of the endemic Amazonian conifer, *Retrophyllum piresii* (Silba) C.N.Page, reveals different recombination events and plastome isoforms. Tree Genet Genomes 12:10. https://doi.org/10.1007/s11295-016-0968-0

Vilas-Boas MA, Carneiro ACO, Vital BR, Carvalho AMML, Martins MA (2010) Efeito da temperatura de carbonização e dos resíduos de macaúba na produção de carvão vegetal. Sci For 38:481–490

Wambulwa MC, Meegahakumbura MK, Kamunya S, Muchugi A, Möller M, Liu J, Xu JC, Ranjitkar S, Li DZ, Gao LM (2016) Insights into the genetic relationships and breeding patterns of the African tea germplasm based on nSSR markers and cpDNA sequences. Front Plant Sci 7:1244. https://doi.org/10.3389/fpls.2016.01244

Weng ML, Blazier JC, Govindu M, Jansen RK (2014) Reconstruction of the ancestral plastid genome in geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. Mol Biol Evol 31:645–659. https://doi.org/10.1093/molbev/mst257

Weng ML, Ruhlman TA, Jansen RK (2016) Plastid-nuclear interaction and accelerated coevolution in plastid ribosomal genes in Geraniaceae. Genome Biol Evol 8:1824–1838. https://doi.org/10.1093/gbe/evw115

Wheeler GL, Dorman HE, Buchanan A, Challagundla L, Wallace LE (2014) A review of the prevalence, utility, and caveats of using chloroplast simple sequence repeats for studies of plant biology. Appl Plant Sci. https://doi.org/10.3732/apps.1400059

Wicke S, Schneeweiss GM, dePamphilis CW, Müller KF, Quandt D (2011) The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol Biol 76:273–297. https://doi.org/10.1007/s11103-011-9762-4

Williams AV, Boykin LM, Howell KA, Nevill PG, Small I (2015) The complete sequence of the acacia ligulata chloroplast genome reveals a highly divergent clpP1 gene. PLoS One 10:e0125768. https://doi.org/10.1371/journal.pone.0125768

Wyman SK, Jansen RK, Boore JL (2004) Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20:3252–3255. https://doi.org/10.1093/bioinformatics/bth352

Xu JH, Liu Q, Hu W, Wang T, Xue Q, Messing J (2015) Dynamics of chloroplast genomes in green plants. Genomics 106:221–231. https://doi.org/10.1016/j.ygeno.2015.07.004

Yamane K, Yano K, Kawahara T (2006) Pattern and rate of indel evolution inferred from whole chloroplast intergenic regions in sugarcane, maize and rice. DNA Res Int J Rapid Publ Rep Genes Genomes 13:197–204. https://doi.org/10.1093/dnares/dsl012

Zhang J, Khan SA, Heckel DG, Bock R (2017) Next-generation insect-resistant plants: RNAi-mediated crop protection. Trends Biotechnol 35:871–882. https://doi.org/10.1016/j.tibtech.2017.04.009

Zhu A, Guo W, Gupta S, Fan W, Mower JP (2016) Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. New Phytol 209:1747–1756. https://doi.org/10.1111/nph.13743