

## RESEARCH ARTICLE

# The plastome sequence of *Bactris gasipaes* and evolutionary analysis in tribe Cocoseae (Arecaceae)

Raquel Santos da Silva<sup>1</sup>, Charles Roland Clement<sup>2</sup>, Eduardo Balsanelli<sup>3,4</sup>, Valter Antonio de Baura<sup>3</sup>, Emanuel Maltempi de Souza<sup>3</sup>, Hugo Pacheco de Freitas Fraga<sup>1</sup>, Leila do Nascimento Vieira<sup>1\*</sup>

**1** Departamento de Botânica, Universidade Federal do Paraná, Curitiba, Paraná, Brazil, **2** Coordenação de Tecnologia e Inovação, Instituto Nacional de Pesquisas da Amazônia, Manaus, AM, Brazil, **3** Departamento de Bioquímica e Biologia Molecular, Núcleo de Fixação Biológica de Nitrogênio, Universidade Federal do Paraná, Curitiba, Paraná, Brazil, **4** Departamento de Bioquímica e Biologia Molecular, GoGenetic, Universidade Federal do Paraná, Curitiba, Paraná, Brazil

\* [leilanvieira@gmail.com](mailto:leilanvieira@gmail.com)



## OPEN ACCESS

**Citation:** Santos da Silva R, Roland Clement C, Balsanelli E, de Baura VA, Maltempi de Souza E, Pacheco de Freitas Fraga H, et al. (2021) The plastome sequence of *Bactris gasipaes* and evolutionary analysis in tribe Cocoseae (Arecaceae). PLoS ONE 16(8): e0256373. <https://doi.org/10.1371/journal.pone.0256373>

**Editor:** Berthold Heinze, Austrian Federal Research Centre for Forests BFW, AUSTRIA

**Received:** April 9, 2021

**Accepted:** August 4, 2021

**Published:** August 24, 2021

**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0256373>

**Copyright:** © 2021 Santos da Silva et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The plastome sequence was deposited in GenBank: MW054718.

## Abstract

The family Arecaceae is distributed throughout tropical and subtropical regions of the world. Among the five subfamilies, Arecoideae is the most species-rich and still contains some ambiguous inter-generic relationships, such as those within subtribes Attaleinae and Bactridinae. The hypervariable regions of plastid genomes (plastomes) are interesting tools to clarify unresolved phylogenetic relationships. We sequenced and characterized the plastome of *Bactris gasipaes* (Bactridinae) and compared it with eight species from the three Cocoseae sub-tribes (Attaleinae, Bactridinae, and Elaeidinae) to perform comparative analysis and to identify hypervariable regions. The *Bactris gasipaes* plastome has 156,646 bp, with 113 unique genes. Among them, four genes have an alternative start codon (*cema*, *rps19*, *rpl2*, and *ndhD*). Plastomes are highly conserved within tribe Cocoseae: 97.3% identity, length variation of ~2 kb, and a single ~4.5 kb inversion in *Astrocaryum* plastomes. The LSC/IR and IR/SSC junctions vary among the subtribes: in Bactridinae and Elaeidinae the *rps19* gene is completely contained in the IR region; in the subtribe Attaleinae the *rps19* gene is only partially contained in the IRs. The hypervariable regions selected according to sequence variation (SV%) and frequency of parsimony informative sites (PIS%) revealed plastome regions with great potential for molecular analysis. The ten regions with greatest SV% showed higher variation than the plastid molecular markers commonly used for phylogenetic analysis in palms. The phylogenetic trees based on the plastomes and the hypervariable regions (SV%) datasets had well-resolved relationships, with consistent topologies within tribe Cocoseae, and confirm the monophyly of the subtribes Bactridinae and Attaleinae.

## Introduction

The family Arecaceae contains 181 genera and about 2,600 species distributed throughout tropical and subtropical regions of the world [1]. The most recent taxonomic review in

**Funding:** This research was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq - <https://www.gov.br/cnpq>) [grant numbers LNV 435200/2018-6, CRC 303477/2018-0]. This research was also partially supported by The Brazilian Program of National Institutes of Science and Technology-INCT/ Brazilian Research Council-CNPq/MCT and Embrapa's Portfólio de Recursos Genéticos (<https://www.embrapa.br/>), which helped to maintain the Peach palm Active Germplasm Bank at INPA. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

Arecaceae, published by Baker and Dransfield [1], recognizes five subfamilies: Arecoideae, Calamoideae, Ceroxyloideae, Coryphoideae, and Nypoideae. Arecoideae is the largest subfamily, with 14 tribes and 108 genera. Several Arecoideae tribes have been extensively studied, but some inter-generic relationships remain ambiguous, such as those within subtribes Attaleinae and Bactridinae [1]. These subtribes are both within tribe Cocoseae, which includes Elaeidinae as its third subtribe.

Plastid genomes (plastomes) are a useful tool for phylogenetic and evolutionary studies [2]. Hypervariable regions in plastomes can provide information to elucidate phylogenetic relationships that are not yet well resolved [3, 4]. However, these highly variable regions vary between clades, and their identification may be necessary for each taxonomic level [3]. Angiosperms show structural rearrangements, loss of genes, introns, and heterogeneous nucleotide substitution rates in protein-coding genes among their plastomes [5, 6]. Also, the fact that palms have a low mutation rate in the plastome [7] makes the identification of clade-specific hypervariable regions singularly relevant for the group.

In addition to use in phylogenetic analysis, these hypervariable regions may also be powerful molecular markers for analysis of population genetic structure, including that among wild and domesticated populations. The family Arecaceae includes species of great ecological importance, either by their interaction with pollinators [8] or with frugivorous animals [9, 10]. Also, palms provide several high-value products for industry (e.g., fibers, construction materials, oil, medicinal compounds, heart-of-palm, fruits) and are especially important for tropical and subtropical indigenous and traditional communities [11, 12]. Among the most useful palms, the coconut (*Cocos nucifera* L.), the date palm (*Phoenix dactylifera* L.), and the oil palm (*Elaeis guineensis* Jacq.) stand out by their global economic importance. The peach palm (*Bactris gasipaes* Kunth) is the only fully domesticated palm in the Neotropics [11]; it was domesticated for its fruit and is becoming important as a source of heart-of-palm [13]. Many Cocoseae species included in our analysis have interesting domestication histories, such as *Bactris gasipaes*, *Elaeis guineensis*, *Cocos nucifera* [11, 14, 15]. Others are from genera with traditional use by indigenous and traditional communities in South America, such as *Butia*, *Astrocaryum*, *Acrocomia*, and *Syagrus* [16–19].

We sequenced and characterized the plastome of *Bactris gasipaes* and compared it with eight species from the three sub-tribes (Attaleinae, Bactridinae, and Elaeidinae) of the tribe Cocoseae to perform comparative structural analysis and phylogenetic inference, and to identify hypervariable regions.

## Material and methods

### Taxon sampling

We collected fresh leaves from a wild individual of *Bactris gasipaes*, i.e., var. *chichagui* type 1 ([20]; accession number F0205/83) in the core collection of Peach palm Active Germplasm Bank [21] at the National Research Institute for Amazonia (INPA, Manaus, AM, Brazil).

Plastomes and nuclear marker sequences from seven genera of tribe Cocoseae were downloaded from GenBank, including individuals from its three subtribes: Attaleinae (3 genera; 3 species), Elaeidinae (1 species), and Bactridinae (3 genera; 4 species). The species names and the GenBank accession numbers for sequences used in the analysis are in [S1 Table](#).

### *Bactris gasipaes* plastome sequencing

The extraction of plastid-enriched DNA was based on the methodology described by Vieira et al. [22] and modified by Sakaguchi et al. [23], proportionally adjusting the buffer volumes for 8 g of fresh leaves. The DNA extraction was performed with CTAB buffer, as described by

Shi et al. [24]. The DNA was purified with the Genomic DNA Clean & Concentrator™-10 Kit (Zymo Research, Irvine, CA, USA). The purified DNA was quantified using Qubit™ dsDNA HS Assay kit (Thermo Fisher Scientific, Carlsbad, CA, USA) in Qubit™ Fluorometer (Thermo Fisher Scientific). Libraries were prepared with Nextera XT DNA Library Preparation Kit (Illumina, San Diego, CA, USA) and sequenced in Illumina MiSeq® (Illumina), obtaining 250 bp paired-end reads.

Plastome assembly was performed using CLC Genomics Workbench v.8.0 (Qiagen, Germantown, MD, USA) software with *de novo* strategy. The *Acrocomia aculeata* plastome was used as a reference for the ordering of contigs. Plastome annotation was performed using Geneious Prime® (Biomatters, Auckland, New Zealand). For all genes, manual verification was performed, adjusting the initial and terminal codons. The final plastome sequence was deposited in GenBank: MW054718.

### Plastome structural analysis in tribe Cocoseae

The comparative analysis to identify structural rearrangements in the plastomes of the Cocoseae species was carried out using eight species (S1 Table), excluding one IR from all plastomes, and using the progressive alignment on Mauve software [25]. The IRScope software [26] was used to visualize and compare the plastome junctions (IRb/LSC; IRb/SSC; SSC/IRa; IRa/LSC).

### Identification of hypervariable regions

We estimated the variability of the sequences with the formula proposed by Shaw et al. [27], adapted and used by Zavala-Páez et al. [28]. First, we individually aligned each collinear coding sequence (CDS), intergenic spacers (IGS), and introns of the plastomes (list of species in S1 Table) using MAFFT v.7 software [29]. Then, the alignments were imported into the software DNAsp v6.12.03 [30] to obtain the number of invariable sites (monomorphic), parsimony informative sites (PIS), number of substitutions, and number of InDel events. Sequence variability (SV) was calculated using the formula:  $SV\% = [(number\ of\ substitutions + number\ of\ InDels) / (number\ of\ substitutions + number\ of\ InDels + invariable\ sites)] \times 100$ . The frequency of PIS was calculated using the formula:  $PIS\% = [(number\ of\ parsimony\ informative\ sites / number\ of\ substitutions + number\ of\ InDels + invariable\ sites)] \times 100$ .

The ten regions with the highest SV% and PIS% values were selected to carry out the subsequent analysis. The plastid markers *matK*, *trnQ-rps16*, *rps16* intron, *trnD-trnT*, *trnL-trnF* [31, 32] and the nuclear markers PRK and RPB2 [32, 33], commonly used for phylogenetic analysis in Arecaceae, were used for comparative purposes and subjected to the same procedure to obtain the PIS% and SV% values.

### Phylogenetic inferences

Phylogenetic inferences were made including the following species of the tribe Cocoseae: *Bactris gasipaes*, *Acrocomia aculeata*, *Astrocaryum aculeatum*, *Astrocaryum murumuru*, *Butia eriopatha*, *Cocos nucifera*, *Elaeis guineensis*, *Syagrus coronata*, and two species as outgroup: *Brahea brandegeei* (Purpus) H. E. Moore (subfamily Coryphoideae) and *Archontophoenix alexandrae* (F.Muell.) H.Wendl. & Drude (subfamily Arecoideae). Three data sets were used: i) the plastome alignment (one IR excluded); ii) the ten regions with the greatest SV% value; iii) the ten regions with the greatest PIS% value.

Plastome alignment was performed using progressive alignment on Mauve [25] implemented in Geneious Prime® v.2020.1.2. The Locally Collinear Blocks (LCBs) identified by Mauve were individually extracted and concatenated. The alignment of the ten regions with

**Table 1. Substitution models selected for the phylogenetic inferences using Maximum Likelihood (ML).**

Region	Models for ML
Plastome	K3Pu+F+R2
<i>trnC-petN</i>	HKY+F
<i>psbC-trnS</i>	JC+I
<i>psaC-ndhE</i>	F81+F+G4
<i>ccsA-ndhD</i>	F81+F+I
<i>petN-psbM</i>	TPM2+F+I
<i>accD-psaI</i>	F81+F+I
<i>trnS-trnG</i>	F81+F+I
<i>rps15-ycf1</i>	HKY+F+I
<i>ndhF-rpl32</i>	F81+F+I
<i>rpl16-intron</i>	F81+F+I
<i>petD-rpoA</i>	F81+F+I
<i>petA-psbJ</i>	F81+F+I
<i>trnG-trnfM</i>	F81+F+I
<i>rps8-rpl14</i>	F81+F+I

<https://doi.org/10.1371/journal.pone.0256373.t001>

the highest SV% and PIS% values was carried out using MAFFT v.7.450 [29] implemented in Geneious Prime<sup>®</sup> v.2020.1.2. Phylogenetic inferences were performed by Maximum Likelihood (ML) using W-IQ-tree [34], with 1,000 bootstrap repetitions. The choice of substitution models, including FreeRate heterogeneity model, was made according to Bayesian information criterion (BIC; Table 1). Branch support analysis was performed with 1,000 repetitions of bootstrap and single branch test SH (-aLTR, 1,000 replicates). The resulting trees were represented using Geneious Prime<sup>®</sup> v.2020.1.2.

## Results

### *Bactris gasipaes* plastome

The sequencing of plastid-enriched DNA resulted in 448,600 reads with an average length of 214 bp. Of these, 47,735 were plastome reads (~10%), resulting in an average depth of coverage of 67.64 (SD = 24.32). The assembled plastome has a 21 bp gap in the IGS *trnT-UGU/trnL-UAA* (position 46,800 to 46,820). This gap is in an AT-rich region (sequence 22 bp upstream to 119 bp downstream is only 7.8% of GC-content) and is, therefore, difficult to sequence [35]. We calculated this gap length using other species of tribe Cocoseae as reference.

*Bactris gasipaes* plastome has the quadripartite structure typically found in angiosperms [2], with a pair of inverted repeat (IRs), a large single-copy region (LSC), and a small single-copy region (SSC). The IRs are 27,038 bp in length (each), the LSC is 85,118 bp, and the SSC is 17,452 bp, resulting in a plastome with 156,646 bp.

*Bactris gasipaes* plastome has an average GC-content of 37.5%. When comparing the plastome regions, the SSC has the lowest GC-content, with 31.3%, followed by LSC with 35.5%. The IRs have the highest value, with 42.6% of GC-content. The rRNA and tRNA show high GC-content, with 55.3% and 53.4%, respectively. Protein-coding genes have an average GC-content of 37.9%, intergenic spacers (IGS) of 37.5%, and introns of 37.1%. The plastome GC-content among species from tribe Cocoseae is similar, ranging from 37.40% (*Elaeis guineensis*) to 37.53% (*Acrocomia aculeata*).

In the *Bactris gasipaes* plastome, we annotated 113 unique genes, 79 of which are protein-coding genes, 30 tRNA genes, and 4 rRNA genes (Table 2). Duplicate genes in IRs include 8

**Table 2. List of genes of *Bactris gasipaes* plastome organized according to their location.**

Plastome region	Name of genes
Large Single Copy (LSC)	<i>psbA</i> , <i>trnK-UUU</i> <sup>*</sup> , <i>matK</i> , <i>rps16</i> <sup>*</sup> , <i>trnQ-UUG</i> , <i>psbK</i> , <i>psbI</i> , <i>trnS-GCU</i> , <i>trnG-UCC</i> <sup>*</sup> , <i>trnR-UCU</i> , <i>atpA</i> , <i>atpF</i> <sup>*</sup> , <i>atpH</i> , <i>atpI</i> , <i>rps2</i> , <i>rpoC2</i> , <i>rpoC1</i> <sup>*</sup> , <i>rpoB</i> , <i>trnC-GCA</i> , <i>petN</i> , <i>psbM</i> , <i>trnD-GUC</i> , <i>trnY-GUA</i> , <i>trnE-UUC</i> , <i>trnT-GGU</i> , <i>psbD</i> , <i>psbC</i> , <i>trnS-UGA</i> , <i>psbZ</i> , <i>trnG-GCC</i> , <i>trnM-CAU</i> , <i>rps14</i> , <i>psaB</i> , <i>psaA</i> , <i>ycf3</i> <sup>*</sup> , <i>trnS-GGA</i> , <i>rps4</i> , <i>trnT-UGU</i> , <i>trnL-UAA</i> <sup>*</sup> , <i>trnF-GAA</i> , <i>ndhJ</i> , <i>ndhK</i> , <i>ndhC</i> , <i>trnV-UAC</i> <sup>*</sup> , <i>trnM-CAU</i> , <i>atpE</i> , <i>atpB</i> , <i>rbcL</i> , <i>accD</i> , <i>psaI</i> , <i>ycf4</i> , <i>cemA</i> , <i>petA</i> , <i>psbJ</i> , <i>psbL</i> , <i>psbF</i> , <i>psbE</i> , <i>petL</i> , <i>petG</i> , <i>trnW-CCA</i> , <i>trnP-UGG</i> , <i>psaJ</i> , <i>rpl33</i> , <i>rps18</i> , <i>rpl20</i> , <i>rps12</i> <sup>*</sup> (exon 1), <i>clpP</i> <sup>*</sup> , <i>psbB</i> , <i>psbT</i> , <i>psbN</i> , <i>psbH</i> , <i>petB</i> <sup>*</sup> , <i>petD</i> <sup>*</sup> , <i>rpoA</i> , <i>rps11</i> , <i>rpl36</i> , <i>infA</i> , <i>rps8</i> , <i>rpl14</i> , <i>rpl16</i> <sup>*</sup> , <i>rps3</i> , <i>rpl22</i>
Inverted Repeat (IR)	<i>rps19</i> , <i>trnH-GUG</i> , <i>rpl2</i> <sup>*</sup> , <i>rpl23</i> , <i>trnI-CAU</i> , <i>ycf2</i> , <i>trnL-CAA</i> , <i>ndhB</i> <sup>*</sup> , <i>rps7</i> , <i>rps12</i> <sup>*</sup> (exons 2 and 3), <i>trnV-GAC</i> , <i>rrn16</i> , <i>trnI-GAU</i> <sup>*</sup> , <i>trnA-UGC</i> <sup>*</sup> , <i>rrn23</i> , <i>rrn4.5</i> , <i>rrn5</i> , <i>trnR-ACG</i> , <i>trnN-GUU</i>
IR / SSC junction	<i>ndhF</i> , <i>ycf1</i>
Small Single Copy (SSC)	<i>rpl32</i> , <i>trnL-UAG</i> , <i>ccsA</i> , <i>ndhD</i> , <i>psaC</i> , <i>ndhE</i> , <i>ndhG</i> , <i>ndhI</i> , <i>ndhA</i> <sup>*</sup> , <i>ndhH</i> , <i>rps15</i>

\* Intron-containing genes.

<https://doi.org/10.1371/journal.pone.0256373.t002>

tRNA genes, 4 rRNA genes, and 7 protein-coding genes (Table 2). Among the 113 genes, 15 genes contain one intron (6 tRNA genes and 9 protein-coding genes) and 3 genes contain two introns (*clpP*, *ycf3*, and *rps12*; Table 2). Among intron-containing genes, 12 are located in LSC (*trnK-UUU*, *rps16*, *trnG-UCC*, *atpF*, *rpoC1*, *ycf3*, *trnL-UAA*, *trnV-UAC*, *clpP*, *petB*, *petD*, *rpl16*), 1 in SSC (*ndhA*), 4 in IRs (*rpl2*, *ndhB*, *trnI-GAU*, *trnA-UGC*), and *rps12* is a trans-splicing gene with the first exon located in the LSC region and the second and third exons in the IRs.

Surprisingly, the *cemA* gene exhibited an alternative start codon, as reported in species of subtribes Attaleinae and Elaeidinae, contrasting with the Bactridinae species sequenced so far (*Astrocaryum aculeatum*, *Astrocaryum murumuru*, and *Acrocomia aculeata*). Three other genes have alternative initiation codons, *rps19* (GTG), *rpl2* (ACG), *ndhD* (ATC).

### Comparative analysis in tribe Cocoseae

Plastomes are highly conserved (97.3% identity) within tribe Cocoseae. Species from subtribes Bactridinae and Elaeidinae have a plastome ~2 kb larger than the species from subtribe Attaleinae. This difference in length between the subtribes is mainly in the IRs and LSC regions (Table 3). The plastomes from Cocoseae species ranged from 154,048 bp (*Butia eriospatha*) to 156,937 bp (*Elaeis guineensis*).

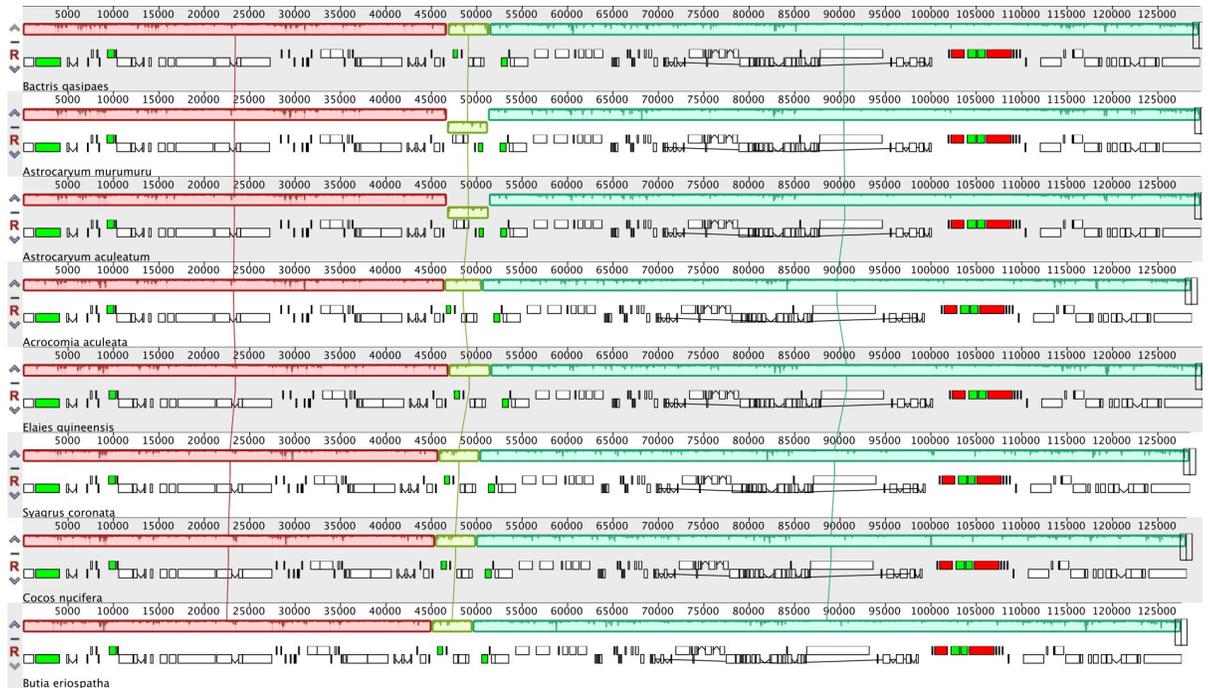
The progressive alignment among species from tribe Cocoseae shows evidence for three Locally Collinear Blocks (LCBs) (Fig 1). These three LCBs are a result of the 4.5 kb inversion

**Table 3. Plastomes of tribe Cocoseae.**

Subtribe	Species	Plastome (bp)	LSC (bp)	IR (bp)	SSC (bp)
Elaeidinae	<i>Elaeis guineensis</i>	156,973	85,192	27,071	17,639
Bactridinae	<i>Astrocaryum aculeatum</i>	156,804	85,037	27,081	17,605
	<i>Astrocaryum murumuru</i>	156,801	85,017	27,081	17,622
	<i>Bactris gasipaes</i>	156,646	85,118	27,038	17,452
	<i>Acrocomia aculeata</i>	156,500	84,936	27,092	17,380
Attaleinae	<i>Syagrus coronata</i>	155,053	84,535	26,522	17,474
	<i>Cocos nucifera</i>	154,731	84,230	26,555	17,391
	<i>Butia eriospatha</i>	154,048	83,805	26,437	17,369

LSC: large single copy region; IR: inverted repeat; SSC: small single copy region.

<https://doi.org/10.1371/journal.pone.0256373.t003>



**Fig 1. Plastome rearrangement analysis within tribe Cocoseae.** Locally collinear blocks (LCBs) are identified by colors. The 4.5 kb inversion in *Astrocarium* is in green.

<https://doi.org/10.1371/journal.pone.0256373.g001>

present in the plastomes of *Astrocarium murumuru* and *Astrocarium aculeatum* (Fig 1). The set of genes that makes up this structural rearrangement is composed of *ndhC*, *ndhK*, *ndhJ*, *trnF-GAA*, and *trnL-UAA*.

In the LSC/IR and IR/SSC junctions of the plastomes, there are differences among the subtribes (Fig 2). In Bactridinae (*Acrocomia aculeata*, *Astrocarium aculeatum*, *Astrocarium murumuru*, and *Bactris gasipaes*) and Elaeidinae (*Elaeis guineensis*) the *rps19* gene is completely contained in the IR region and, therefore, there are two copies of the complete gene. In contrast, in the subtribe Attaleinae (*Butia eriospatha*, *Cocos nucifera*, and *Syagrus coronata*) the *rps19* gene is only partially contained in the IRs, resulting in a complete *rps19* and a partial *rps19*: the complete *rps19* gene starts at IRb and ends at LSC (LSC/IRb); the partial *rps19* starts at IR, but does not contain the final portion of the gene (Fig 2).

Similarly, the *ycf1* gene is partially contained in IRs, with a complete *ycf1* at IRa/SSC and a partial (pseudo) *ycf1* at IRb. The *ndhF* gene has both position and length conserved at the IRb/SSC junction in tribe Cocoseae, with the portion of the gene contained in the IRb overlapping the *ycf1* gene (56 bp) (Fig 2).

### Hypervariable regions

We carried out the SV% and PIS% estimates to identify the plastome regions with the greatest variation within tribe Cocoseae. All ten regions selected according to the highest SV% showed greater variation than the plastid molecular markers commonly used for phylogenetic analysis in palms (Fig 3A). As expected, they have SV% lower than the nuclear markers PRK and RPB2 (Fig 3A). Among the ten regions selected according to the highest PIS% values, all showed greater values than the plastid molecular markers commonly used for phylogenetic analysis in palms (Fig 3B) and two of them (*trnC-petN* and *psbC-trnS*) were more variable than the

### Inverted Repeats

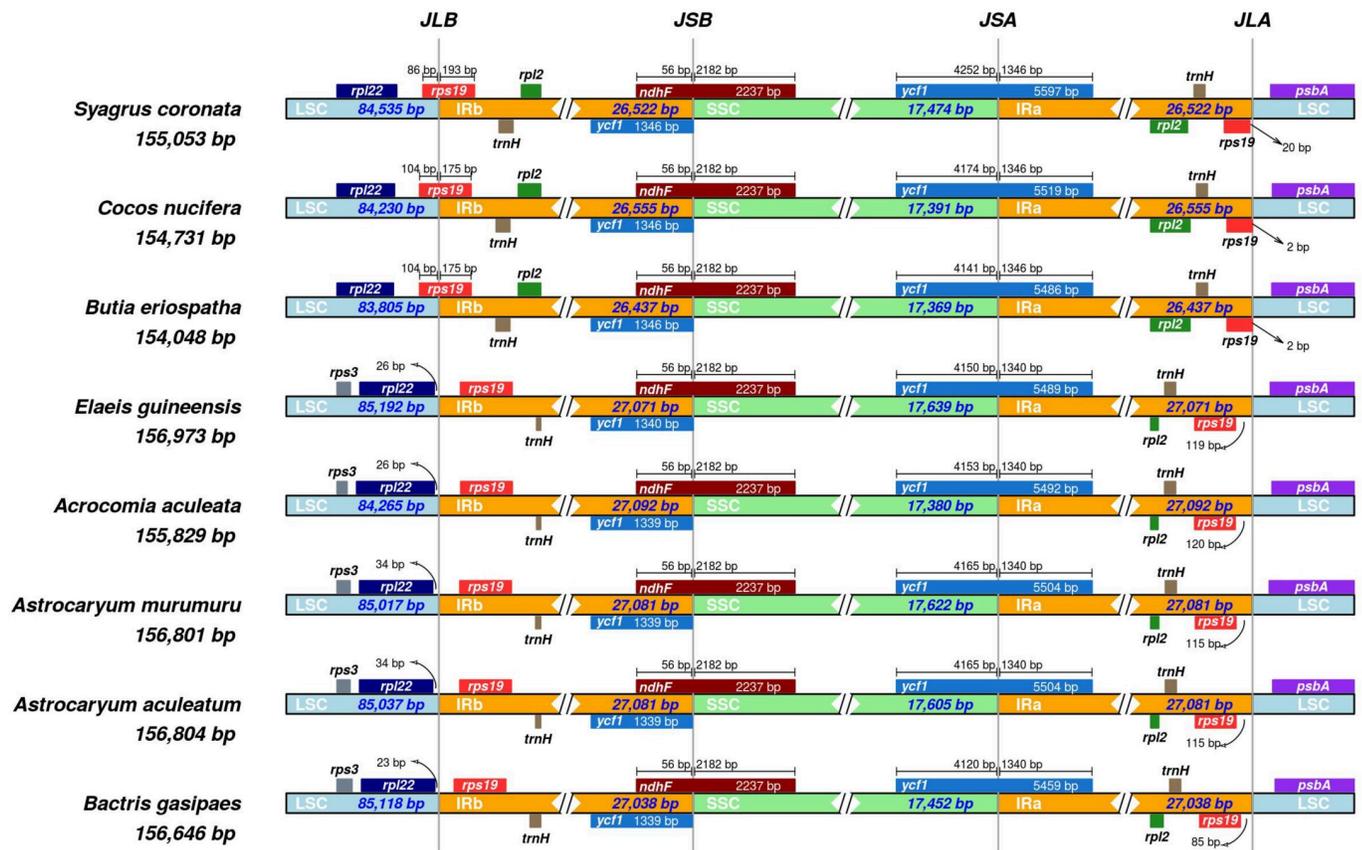


Fig 2. Comparison of plastome junctions (IRb/LSC; IRb/SSC, SSC/IRa; IRa/LSC) among Cocoseae species. The numbers indicate sequence length in base pairs.

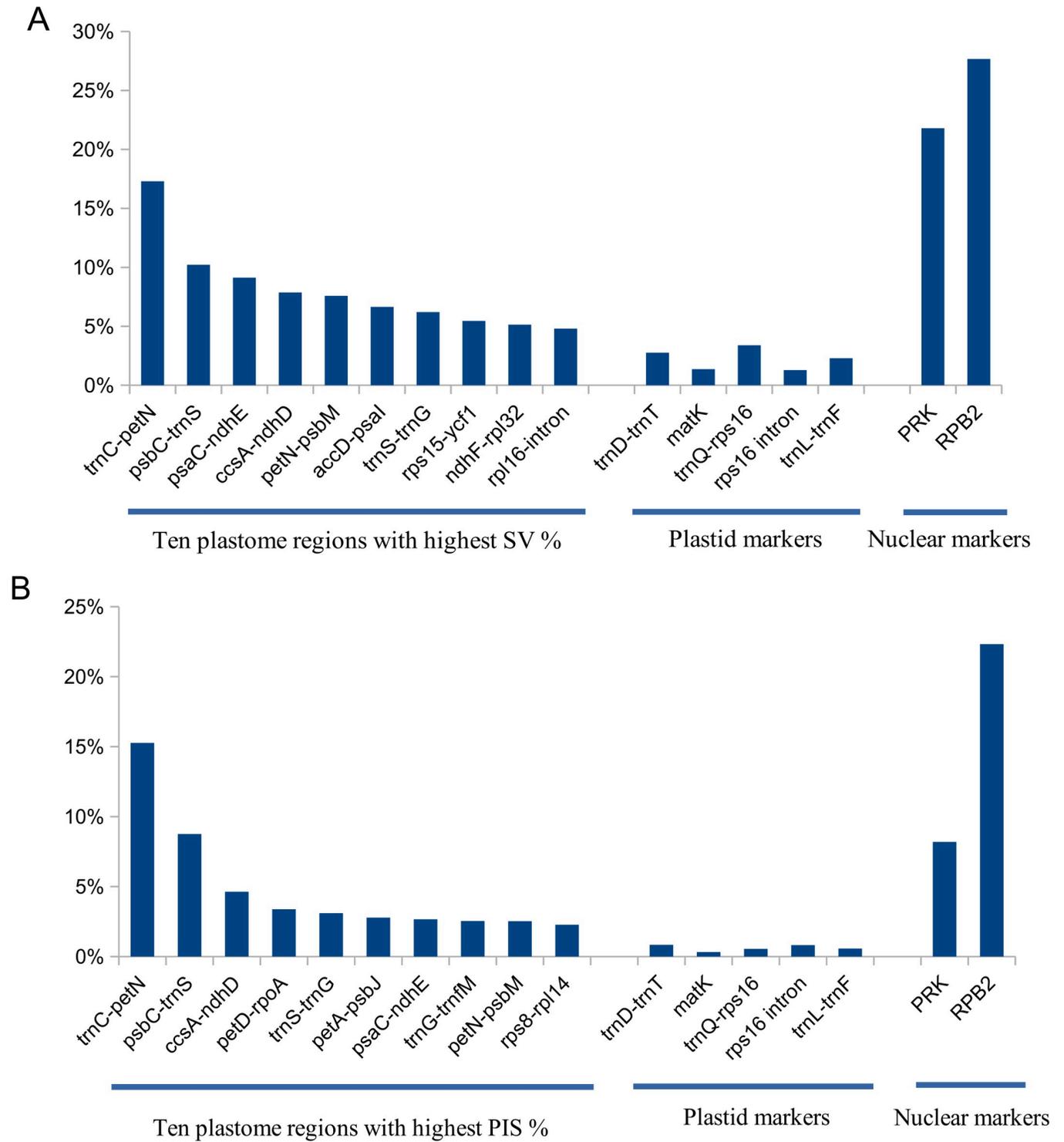
<https://doi.org/10.1371/journal.pone.0256373.g002>

nuclear marker PRK (Fig 3B). The nuclear marker RPB2 showed the highest values for both SV% and PIS% estimates.

We also calculated the frequency of substitutions and frequency of InDel events for each plastome region. The substitution frequency was ~5x higher than InDels in plastomes (Table 4). In the coding sequences (CDSs), substitutions are ~80x more common than InDels. In IGSs and introns, substitutions are ~4x and ~3x more frequent than InDels, respectively (Table 4). In general, these data show that in all regions of Cocoseae plastomes there is a higher frequency of substitutions than InDels.

### Phylogenomic analysis

The complete LCB matrix of the aligned plastomes consisted of 137,452 columns, of which 134,244 are constant and 815 parsimony-informative. The SV matrix contained a total of 9614 columns, with 133 parsimony-informative and 9144 constant sites, and the PIS matrix 7538 columns, with 129 parsimony-informative and 7200 constant sites. The phylogenomic trees inferred using Maximum Likelihood (ML) and based on the plastome and the ten selected regions with greatest SV% datasets showed identical topologies (Fig 4) and high bootstrap



**Fig 3. Hypervariable plastome regions compared with commonly used plastid and nuclear markers.** A—The ten plastome regions with greatest sequence variation (SV%); B—The ten plastome regions with greatest frequency of parsimony informative sites (PIS%).

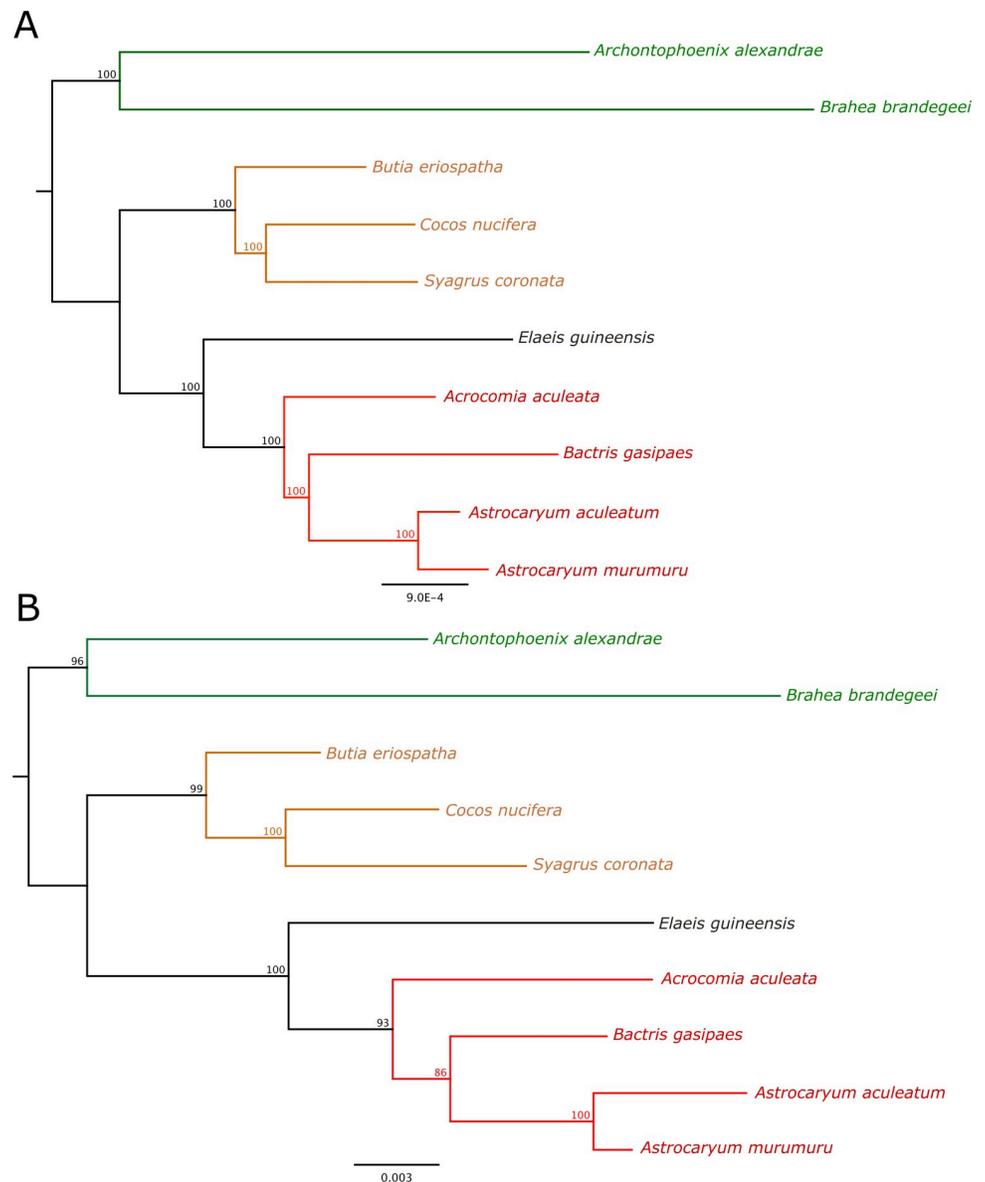
<https://doi.org/10.1371/journal.pone.0256373.g003>

**Table 4. Frequency of substitutions/mutations and insertions/deletions (InDels) among the plastome sequence.**

	Plastome	CDS	IGS	Introns
Mutations (%)	1.53	0.80	2.30	1.12
InDels (%)	0.30	0.01	0.58	0.34
Mutation / InDel ratio	5.02	79.90	3.99	3.26

CDS—coding sequence; IGS—intergenic spacer.

<https://doi.org/10.1371/journal.pone.0256373.t004>



**Fig 4. Phylogenetic trees based on maximum likelihood inference.** A—Phylogenetic inference using plastome sequences (one IR removed); B—Phylogenetic inference using the ten plastome regions with greatest sequence variability (SV%) values. The numbers above the branches are maximum likelihood bootstrap values (1000 replicates).

<https://doi.org/10.1371/journal.pone.0256373.g004>

support (> 86). The phylogenetic tree inferred from the alignment of the ten regions with greatest PIS values showed a similar topology; it differed only by the presence of a polytomy in the Attaleinae clade, generated by the low bootstrap support in intergeneric relationships (S1 Fig). For the topologies generated by ML in the three datasets, the monophyly of the subtribes Bactridinae and Attaleinae was confirmed. Also, the subtribe Elaeidinae appears more closely related to Bactridinae than to Attaleinae. In Bactridinae, *Bactris* and *Astrocaryum* are closely related genera, and in Attaleinae, *Cocos nucifera* as sister to *Syagrus coronata*.

## Discussion

### *Bactris gasipaes* plastome

The plastome of *Bactris gasipaes* has the typical quadripartite structure and gene content of other described Cocoseae species [5, 36–40]. The IRs are identical, which probably occurs due to mechanisms of replication and repair through recombination-dependent replication (RDR), as previously reported in plant plastomes [41]. The GC content in the plastomes of *Bactris gasipaes* and other species of Cocoseae corroborate the mean value described by Kwon et al. [42] for angiosperms (37.71; SD 1.10). The tendency for a higher GC content in the IRs than in the LSC and SSC was previously reported in bryophytes, ferns, lycophytes, and angiosperms [42, 43].

The gene content is conserved among species of Bactridinae (e.g., *Acrocomia aculeata*, *Astrocaryum murumuru*, *Astrocaryum aculeatum*, and *Bactris gasipaes*) and Elaeidinae (*Elaeis guineensis*) [36, 38, 39]. Attaleinae species present one pseudogenized *rps19*, and thus, one less CDS [40].

The gene *cemA* has an unconventional start codon in *Bactris gasipaes*, what was previously described in species of subtribes Attaleinae and Elaeidinae, in *Podococcus barteri* (NC\_027276.1), *Phoenix dactylifera* 'Khalas', and other monocots. However, it is still not clear if this gene, with the unconventional start codon, is translatable to a protein [44]. Although most of the genes encoding proteins have ATG initiation codons [7], some alternative initiation codons are found in plants [45], such as GTG in the *rps19* gene, ACG in *rpl2* and ATC in *ndhD*, which were reported in *Lilium longiflorum*, *Phoenix dactylifera* 'Khalas' and *Amomum compactum*, respectively [44, 46, 47].

### Comparative plastome analysis within Cocoseae

Comparative studies using plastomes of species at different taxonomic levels can bring insights into plastome evolution, phylogenetic relationships, and evolutionary rates [2]. Plastomes of the three subtribes analyzed (Attaleinae, Bactridinae, and Elaeidinae), represented here by eight species, provided information to compare sequence variations in tribe Cocoseae. We identified slight differences in plastome size (~2 kb) and an inversion of 4.5 kb that occurs in the *ndh* complex (LSC region) of the *Astrocaryum* plastome. Similarly, Barrett et al. [48] reported that Arecaceae plastomes are highly conserved structure, describing only one 1.9 kb inversion located between the *rps16* and *trnG-UUC* genes in *Tahina spectabilis*.

Also, the variability among Attaleinae, Bactridinae, and Elaeidinae in the LSC/IR junctions, mainly involving the *rps19* gene, was previously described in *Acrocomia aculeata* [38] and *Butia eriosphata* [40], as well as in *Phoenix dactylifera* [44, 49]. Similarly, the *ndhF* gene overlapping with *yef1* in ~25 bp is commonly observed in palms [38, 40, 44, 49]. Despite the differences in the IR junction, the IR structure and gene content is conserved among palms, corroborating the hypothesis that the IR regions offer an isolation mechanism that stabilizes the structure of the genome [50].

## Hypervariable regions

Plastomes have several non-coding regions, but not all of them have been explored for phylogenetic studies [3, 4, 28]. Among the ten regions with the greatest SV% identified in our study, only five IGSs (e.g., *accD-psaI*, *ndhF-rpl32*, *trnS-trnG*, *psaC-ndhE*, *rps15-ycf1*) and one intron (*rpl16*) were previously used and/or highlighted in angiosperm studies [3, 38, 51]. Among the ten regions with the greatest PIS%, only three IGS (*trnS-trnG*, *petA-psbJ*, and *psaC-ndhE*) were identified in studies carried out by Shaw et al. [3, 27] and Lopes et al. [38]. Thus, in our study, we described four new promising regions based on both SV and PIS values (*trnC-petN*, *psbC-trnS*, *ccsA-ndhD*, *petN-psbM*) and three new regions based on PIS (*petD-rpoA*, *trnG-trnfM*, *rps8-rpl14*). Also, Scarcelli et al. [51] reported 100 primers for phylogeny in monocots. However, four of the hypervariable regions reported in our study were not contemplated (e.g., *psaC-ndhE*, *petN-psbM*, *accD-psaI*, *trnS-trnG*). Among them, *trnS-trnG* and *accD-psaI* primers designed by Scarcelli et al. [51] showed no amplification in Areaceae, and *petN-psbM* and *psaC-ndhE* were not mentioned, probably due to gene rearrangements in monocots plastomes. This reinforces that the highly variable regions vary between clades, and their identification may be necessary for distinct taxonomic levels. As expected, the nuclear genes PRK and RPB2 showed greater variation than most plastidial regions. These nuclear markers are very informative and produce well-resolved topologies [33]. The combined use of the plastidial regions described here and the nuclear markers PRK and RPB2 have great potential for phylogenetic studies in tribe Cocoseae.

## Phylogenomic analysis

The ten regions with the greatest SV% values are suitable for phylogenetic inferences and produce phylogenetic trees with well-resolved and the expected topologies. In the ML analysis, all datasets tested (plastome, ten SV regions, and ten PIS regions) result in subtribe Bactridinae as monophyletic. The monophyly Bactridinae was previously described by Eiserhardt et al. [32], as well as the sister relationship between the subtribes Elaeidinae and Bactridinae [33, 38]. Our results are in contrast with those of Gunn [52], in which the sister relationship between *Astrocaryum* and *Bactris* is weakly supported. In all of our datasets this sister relationship is well-supported. In addition, the monophyly of subtribe Cocoseae was also verified in a plastid DNA analysis [53], in the super-tree method [54] and by the combined analysis with the PRK and RPB2 genes [33]. The sister relationship between *Cocos nucifera* and *Syagrus coronata* was also previously described [55], corroborating our results. Thus, both the plastome and the ten regions with greatest SV values were able to produce well-resolved phylogenetic trees and with consistent topologies within tribe Cocoseae.

## Supporting information

**S1 Table. GenBank accession numbers of the nucleotide sequences used in our analysis.**  
(DOC)

**S1 Fig. Phylogenetic tree based on maximum likelihood inference using the ten plastome regions with greatest PIS% values.** The numbers above the branches are maximum likelihood bootstrap values (1000 replicates).  
(TIF)

## Acknowledgments

We thank Jeison W. de S. Magnabosco and Michelle E. Zavala-Páez for their help with the analyses.

## Author Contributions

**Conceptualization:** Charles Roland Clement, Emanuel Maltempi de Souza, Hugo Pacheco de Freitas Fraga, Leila do Nascimento Vieira.

**Data curation:** Hugo Pacheco de Freitas Fraga, Leila do Nascimento Vieira.

**Formal analysis:** Raquel Santos da Silva, Eduardo Balsanelli, Valter Antonio de Baura, Hugo Pacheco de Freitas Fraga, Leila do Nascimento Vieira.

**Funding acquisition:** Charles Roland Clement, Emanuel Maltempi de Souza, Leila do Nascimento Vieira.

**Investigation:** Raquel Santos da Silva, Charles Roland Clement, Emanuel Maltempi de Souza, Leila do Nascimento Vieira.

**Methodology:** Raquel Santos da Silva, Eduardo Balsanelli, Valter Antonio de Baura, Hugo Pacheco de Freitas Fraga, Leila do Nascimento Vieira.

**Project administration:** Charles Roland Clement, Leila do Nascimento Vieira.

**Resources:** Charles Roland Clement, Eduardo Balsanelli, Leila do Nascimento Vieira.

**Software:** Valter Antonio de Baura.

**Supervision:** Charles Roland Clement, Emanuel Maltempi de Souza, Hugo Pacheco de Freitas Fraga, Leila do Nascimento Vieira.

**Validation:** Charles Roland Clement, Emanuel Maltempi de Souza, Hugo Pacheco de Freitas Fraga, Leila do Nascimento Vieira.

**Visualization:** Raquel Santos da Silva, Eduardo Balsanelli, Valter Antonio de Baura.

**Writing – original draft:** Raquel Santos da Silva.

**Writing – review & editing:** Raquel Santos da Silva, Charles Roland Clement, Hugo Pacheco de Freitas Fraga, Leila do Nascimento Vieira.

## References

1. Baker WJ, Dransfield J. Beyond Genera Palmarum: progress and prospects in palm systematics. *Bot J Linn Soc* 2016; 182:207–233.
2. Rogalski M, Vieira LN, Fraga HP, Guerra MP. Plastid genomics in horticultural species: importance and applications for plant population genetics, evolution and biotechnology. *Front Plant Sci* 2015; 6:1–17. <https://doi.org/10.3389/fpls.2015.00001> PMID: 25653664
3. Shaw J, Lickey EB, Schuilling EE, Small RL. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *Am J Bot* 2007; 94:275–288. <https://doi.org/10.3732/ajb.94.3.275> PMID: 21636401
4. Smidt EC, Zavala-Páez M, Vieira LN, Viruel J, Baura VA, Balsanelli E, et al. Characterization of sequence variability hotspots in Cranichideae plastomes (Orchidaceae, Orchidoideae). *PLoS One* 2020; 15:e0227991. <https://doi.org/10.1371/journal.pone.0227991> PMID: 31990943
5. Wicke S, Schneeweiss GM, Depamphilis CW, Muller KF, Quandt D. The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. *Plant Mol Biol* 2011; 76:273–297. <https://doi.org/10.1007/s11103-011-9762-4> PMID: 21424877
6. Ruhlman TA, Jansen RK. The plastid genomes of flowering plants. In: Maliga P, editor. *Chloroplast biotechnology: methods and protocols*. Totowa: Humana Press; 2014. pp. 3–38.
7. Wilson MA, Gaut B, Clegg MT. Chloroplast DNA evolves slowly in the palm family (Arecaceae). *Mol Biol Evol* 1990; 7:303–314. <https://doi.org/10.1093/oxfordjournals.molbev.a040605> PMID: 1974691
8. Barfod AS, Hagen M, Borchsenius F. Twenty-five years of progress in understanding pollination mechanisms in palms (Arecaceae). *AoB Plants* 2011; 108:503–1516.

9. Onstein RE, Baker WJ, Couvreur TLP, Faurby S, Svenning JC, Kissling WD. Frugivory-related traits promote speciation of tropical palms. *Nat Ecol Evol* 2017; 12:1903–1911. <https://doi.org/10.1038/s41559-017-0348-7> PMID: 29062122
10. Muñoz G, Trøjelsgaard K, Kissling WD. A synthesis of animal-mediated seed dispersal of palms reveals distinct biogeographical differences in species interactions. *J Biogeogr* 2019; 46:466–484.
11. Clement CR. Domesticated palms. *Principes* 1992; 36:70–78.
12. Balslev H, Bernal R, Fay MF. Palms—emblems of tropical forests. *Bot J Linn Soc* 2016; 182:195–200.
13. Clement CR. Peach palm (*Bactris gasipaes*). In: Janick J, Paull RE, editors. *The encyclopedia of fruit and nuts*. Wallingford: CABI Publishing; 2008. pp. 93–101.
14. Clement CR, Cristo-Araújo M, Coppens d'Eeckenbrugge G, Reis VM, Lehnbech R, Picanço-Rodrigues D. Origin and dispersal of domesticated peach palm. *Front Ecol Evol* 2017; 5:1–19.
15. Muñoz-Pérez JM, Cañas GP, López L, Arias T. Genome-wide diversity of northern South America cultivated Coconut (*Cocos nucifera* L.) uncovers diversification times and targets of domestication of coconut globally. *bioRxiv* 2019; 825398.
16. Rivas M, Condón F. Plant domestication and utilization: The case of the Pampa biome. In: Al-Khayri JM, et al., editors. *Advances in plant breeding strategies: breeding, biotechnology and molecular tools*. Cham: Springer; 2015. pp. 3–24.
17. Montoya SG, Motoike SY, Kuki KN, Couto AD. Fruit development, growth, and stored reserves in macauba palm (*Acrocomia aculeata*), an alternative bioenergy crop. *Planta* 2016; 244:927–938. <https://doi.org/10.1007/s00425-016-2558-7> PMID: 27318823
18. Levis C, Flores BM, Moreira PA, Luize BG, Alves RP, Franco-Moraes J, et al. How people domesticated Amazonian forests. *Front Ecol Evol* 2018; 5:171.
19. Pereira Cruz A, Giehl ELH, Levis C, Machado JS, Bueno L, Peroni N. Pre-colonial Amerindian legacies in forest composition of southern Brazil. *PLoS One* 2020; 15:e0235819. <https://doi.org/10.1371/journal.pone.0235819> PMID: 32701950
20. Henderson A. *Bactris* (Palmae). New York: Flora Neotropica; 2000
21. Cristo-Araújo M, Picanço-Rodrigues D, Astolfi-Filho S, Clement CR. Peach palm core collection in Brazilian Amazonia. *Crop Breed Appl Biotech* 2015; 15:18–25.
22. Vieira LN, Fraga HPF, Rogalski M, Souza EM, Pedrosa FO, Guerra MP. An improved protocol for intact chloroplasts and cpDNA isolation in conifers. *PLoS One* 2014; 9:e84792. <https://doi.org/10.1371/journal.pone.0084792> PMID: 24392157
23. Sakaguchi S, Ueno S, Tsumura Y, Setoguchi H, Ito M, Hattori C, et al. Application of a simplified method of chloroplast enrichment to small amounts of tissue for chloroplast genome sequencing. *Appl Plant Sci* 2017; 5:1–8. <https://doi.org/10.3732/apps.1700002> PMID: 28529832
24. Shi C, Hu N, Huang H, Gao J, Zhao YJ, Gao LZ. An improved chloroplast DNA extraction procedure for whole plastid genome sequencing. *PLoS One* 2012; 7:e31468. <https://doi.org/10.1371/journal.pone.0031468> PMID: 22384027
25. Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 2004; 14:1394–1403. <https://doi.org/10.1101/gr.2289704> PMID: 15231754
26. Amiryousefi A, Hyvonen J, Poczai P. IRscope: An online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* 2018; 34:3030–3031. <https://doi.org/10.1093/bioinformatics/bty220> PMID: 29659705
27. Shaw J, Shafer HL, Rayne LO, Kovach MJ, Schorr M, Morris AB. Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: The tortoise and the hare IV. *Am J Bot* 2014; 101:1987–2004. <https://doi.org/10.3732/ajb.1400398> PMID: 25366863
28. Zavala-Páez M, Vieira LN, Baura VA, Balsanelli E, Souza EM, Cerna M, et al. Comparative plastid genomics of neotropical *Bulbophyllum* (Orchidaceae; Epidendroideae). *Front Plant Sci* 2020; 11:799. <https://doi.org/10.3389/fpls.2020.00799> PMID: 32719690
29. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol* 2013; 30:772–780. <https://doi.org/10.1093/molbev/mst010> PMID: 23329690
30. Rozas J, Ferrer-Mata A, Sánchez-Delbarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins S, et al. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Datasets. *Mol Biol Evol* 2017; 34:3299–3302. <https://doi.org/10.1093/molbev/msx248> PMID: 29029172
31. Asmussen CB, Dransfield J, Deickmann V, Barfod AS, Pintaud JC, Baker WJ. A new subfamily classification of the palm family (Arecaceae): evidence from plastid DNA phylogeny. *Bot J Linn Soc* 2006; 151:15–38.

32. Eiserhardt WL, Pintaud JC, Asmussen-Lange C, Hahn WJ, Bernal R, Balslev H, et al. Phylogeny and divergence times of Bactridinae (Arecaceae, Palmae) based on plastid and nuclear DNA sequences. *Taxon* 2011; 60:485–498.
33. Baker WJ, Norup MV, Clarkson JJ, Couvreur TLP, Dowe JL, Lewis CE, et al. Phylogenetic relationships among arecoid palms (Arecaceae: Arecoideae). *AoB Plants* 2011; 108:1417–1432.
34. Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res* 2016; 44:W232–W235. <https://doi.org/10.1093/nar/gkw256> PMID: 27084950
35. Benjamini Y, Speed TP. Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Res* 2012; 40:e72–e72. <https://doi.org/10.1093/nar/gks001> PMID: 22323520
36. Uthaisaisanwong P, Chanprasert J, Shearman JR, Sangsrakru D, Yoo-cha T, Jomchai N, et al. Characterization of the chloroplast genome sequence of oil palm (*Elaeis guineensis* Jacq.). *Gene* 2012; 500:172–180. <https://doi.org/10.1016/j.gene.2012.03.061> PMID: 22487870
37. Áquila R, Silva J, Silva JV, Almeida C. The first complete chloroplast genome sequence of *Syagrus coronata*: comparative analysis in Arecoideae (Arecaceae). *Rev Bras Bot* 2018. 41:393–401.
38. Lopes AS, Pacheco TG, Nimz T, Vieira LN, Guerra MP, Nodari RO, et al. The complete plastome of macaw palm [*Acrocomia aculeata* (Jacq.) Lodd. ex Mart.] and extensive molecular analyses of the evolution of plastid genes in Arecaceae. *Planta* 2018; 247:1011–1030. <https://doi.org/10.1007/s00425-018-2841-x> PMID: 29340796
39. Lopes AS, Pacheco TG, Silva ON, Cruz LM, Balsanelli E, Souza EM, et al. The plastomes of *Astrocaryum aculeatum* G. Mey. and *A. murumuru* Mart. show a flip-flop recombination between two short inverted repeats. *Planta* 2019; 250:1229–1246. <https://doi.org/10.1007/s00425-019-03217-z> PMID: 31222493
40. Magnabosco JWS, Fraga HPF, Silva RS, Rogalski M, Souza EM, Guerra MP, et al. Characterization of the complete chloroplast genome of *Butia eriospatha* (Arecaceae). *Genet Mol Biol* 2020; 43:e20200023. <https://doi.org/10.1590/1678-4685-GMB-2020-0023> PMID: 32926069
41. Gonçalves D.JP, et al. Under the rug: abandoning persistent misconceptions that obfuscate organelle evolution. *Mol Phylogenet Evol* 2020; 151:106903. <https://doi.org/10.1016/j.ympev.2020.106903> PMID: 32628998
42. Kwon EC, Kim JH, Kim NS. Comprehensive genomic analyses with 115 plastomes from algae to seed plants: structure, gene contents, GC contents, and introns. *Genes Genomics* 2020; 42:553–570. <https://doi.org/10.1007/s13258-020-00923-x> PMID: 32200544
43. Li FW, Kuo LY, Pryer KM, Rothfels CJ. Genes translocated into the plastid inverted repeat show decelerated substitution rates and elevated GC content. *Genome Biol Evol* 2016; 8:2452–2458. <https://doi.org/10.1093/gbe/evw167> PMID: 27401175
44. Yang M, Zhang X, Liu G, Yin Y, Chen K, Yun Q, et al. The complete chloroplast genome sequence of date palm (*Phoenix dactylifera* L.). *PLoS One* 2010; 5:e12762. <https://doi.org/10.1371/journal.pone.0012762> PMID: 20856810
45. Huang YY, Matzke AJM, Matzke M. Complete sequence and comparative analysis of the chloroplast genome of coconut palm (*Cocos nucifera*). *PLoS One* 2013; 8:74736.
46. Kim JS, Kim JH. Comparative genome analysis and phylogenetic relationship of order Liliales insight from the complete plastid genome sequences of two Lilies (*Lilium longiflorum* and *Alstroemeria aurea*). *PLoS One* 2013; 8:e68180. <https://doi.org/10.1371/journal.pone.0068180> PMID: 23950788
47. Wu M-L, et al. Complete chloroplast genome of the medicinal plant *Amomum compactum*: gene organization, comparative analysis, and phylogenetic relationships within Zingiberales. *Chin Med* 2018; 13:1–12. <https://doi.org/10.1186/s13020-017-0157-6> PMID: 29321807
48. Barrett CF, Baker WJ, Comer JR, Conran JG, Lahmeyer SC, Leebens-Mack JH, et al. Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots. *New Phytol* 2016; 209:855–870. <https://doi.org/10.1111/nph.13617> PMID: 26350789
49. Khan A, Khan IA, Heinze B, Azim MK. The chloroplast genome sequence of date palm (*Phoenix dactylifera* L. cv. 'Aseel'). *Plant Mol Biol Report* 2012; 30:666–678.
50. Yue F, Cui L, dePamphilis CW, Moret BME, Tang J. Gene rearrangement analysis and ancestral order inference from chloroplast genomes with inverted repeat. *BMC Genomics* 2008; 9:S25. <https://doi.org/10.1186/1471-2164-9-S1-S25> PMID: 18366615
51. Scarcelli N, Barnaud A, Eiserhardt W, Treier UA, Seveno M, d'Anfray A, et al. A set of 100 chloroplast DNA primer pairs to study population genetics and phylogeny in monocotyledons. *PLoS One* 2011; 6:e19954. <https://doi.org/10.1371/journal.pone.0019954> PMID: 21637837

52. Gunn BF. The phylogeny of the Cocoeae (Arecaceae) with emphasis on *Cocos nucifera*. *Ann Mo Bot Gard* 2004; 91:505–522.
53. Hahn WJ. A phylogenetic analysis of the Arecoideae line of palms based on plastid DNA sequence data. *Mol Phylogenet Evol* 2002; 23:189–204. [https://doi.org/10.1016/S1055-7903\(02\)00022-2](https://doi.org/10.1016/S1055-7903(02)00022-2) PMID: 12069550
54. Baker WJ, Savolainen V, Asmussen-Lange CB, Chase MW, Dransfield J, Forest F, et al. Complete generic-level phylogenetic analyses of palms (Arecaceae) with comparisons of supertree and supermatrix approaches. *Syst Biol* 2009; 58:240–256. <https://doi.org/10.1093/sysbio/syp021> PMID: 20525581
55. Meerow AW, et al. Phylogeny and historical biogeography of the cocosoid palms (Arecaceae, Arecoideae, Cocoseae) inferred from sequences of six WRKY gene family loci. *Cladistics* 2015; 31: 509–534.