




## Characterization of the complete plastid genome of *Butia eriospatha* (Arecaceae)

Jeison Willy de Souza Magnabosco<sup>1</sup>, Hugo Pacheco de Freitas Fraga<sup>1</sup>, Raquel Santos da Silva<sup>1</sup>, Marcelo Rogalski<sup>2</sup>, Emanuel Maltempi de Souza<sup>3</sup>, Miguel Pedro Guerra<sup>4,5</sup>  and Leila do Nascimento Vieira<sup>1</sup> 

<sup>1</sup>Universidade Federal do Paraná, Programa de Pós-graduação em Botânica, Curitiba, Paraná, Brazil.

<sup>2</sup>Universidade Federal de Viçosa, Programa de Pós-graduação em Fisiologia Vegetal, Viçosa, MG, Brazil.

<sup>3</sup>Universidade Federal do Paraná, Programa de Pós-graduação em Bioquímica e Biologia Molecular, Curitiba, PR, Brazil.

<sup>4</sup>Universidade Federal de Santa Catarina, Programa de Pós-graduação em Recursos Genéticos Vegetais, Florianópolis, SC, Brazil.

<sup>5</sup>Universidade Federal de Santa Catarina, Programa de Pós-Graduação em Ecossistemas Agrícolas e Naturais, Curitiba, SC, Brazil.

### Abstract

*Butia eriospatha* is an endemic palm species from the Atlantic Rainforest in Brazil, a biodiversity hotspot. This species is currently listed in the IUCN red list as vulnerable and lacks specific plastid markers for population genetics studies. In addition, the evolutionary relationship within the genus *Butia* is not yet well resolved. Here, we sequenced and characterized the complete plastid genome (plastome) sequence of *B. eriospatha*. The complete plastome sequence is 154,048 bp in length, with the typical quadripartite structure. This plastome length and genes content is consistent with other six species from tribe Cocoseae. However, the Inverted Repeat (IR) borders show some variation among the analyzed species from this tribe. Species from the Bactridinae (*Astrocaryum* and *Acrocomia*) and Elaeidinae (*Elaeis*) subtribes present the *rps19* gene completely duplicated in the IR region. In contrast, all plastomes sequenced from the subtribe Attaleinae (*Butia*, *Cocos*, *Syagrus*) present one complete CDS of *rps19* and one partial copy of *rps19*. The difference in the IR/LSC junctions between Attaleinae and the sister clades Bactridinae + Elaeidinae might be considered an evolutionary signal and the plastome sequence of *B. eriospatha* may be used in future studies of population genetics and phylogeny.

**Keywords:** Palm, atlantic rainforest, plastome, molecular evolution.

Received: February 06, 2020; Accepted: July 10, 2020.

The family Arecaceae, comprises approximately 188 genera and 2,585 species, distributed throughout tropical and subtropical climate of the world, but it is most diverse in tropical forest habitats, being present from Americas to Asia-Pacific region (Palmweb).

*Butia* (Becc.) Becc. is a genus of the family Arecaceae, indigenous from South America. This genus includes 18 species (WFO, 2019), among them *Butia eriospatha* (Mart. ex Drude) Becc.. This species is endemic to southern Brazil and occurs in the Atlantic Rainforest phytogeographic domain: mainly in high altitude grassland, grassland, palm grove (Heiden *et al.*, 2020). *Butia eriospatha* is easily distinguished by its solitary stem with 4-5 meters in height, globose fruit, spathe ferruginous-pubescent on outer surface, and pistillate flowers with 3-7 mm (Deble *et al.*, 2011; Heiden *et al.*,

2020). Among palms, *B. eriospatha* has a high economic importance, being used for ornamentals projects, and with edible fruits. Currently, the fruits are consumed fresh or in pulp, alcoholic drinks, jams, jellies or ice cream (Hoffmann *et al.*, 2014).

Species of the genus *Butia* show the same chromosome number ( $2n = 2x = 32$ ) (Correa *et al.*, 2019) and the genus is well supported as monophyletic (Meerow *et al.*, 2015). However, within *Butia*, few molecular markers were used in phylogenetic studies. The most recent study on *Butia* phylogeny used only one plastid marker (*trnH-psbA*) and two nuclear markers (ITS and WRKY19) and failed to resolve the evolutionary relationship between *Butia* species (Pereira, 2019).

Plastome sequences are frequently used to understand evolutionary events and efficiently resolve phylogenetic relationships (Lopes *et al.*, 2018, 2019). Typically, the plastome is represented as a single circular molecule composed by two inverted regions (IR), a large single copy (LSC) and a

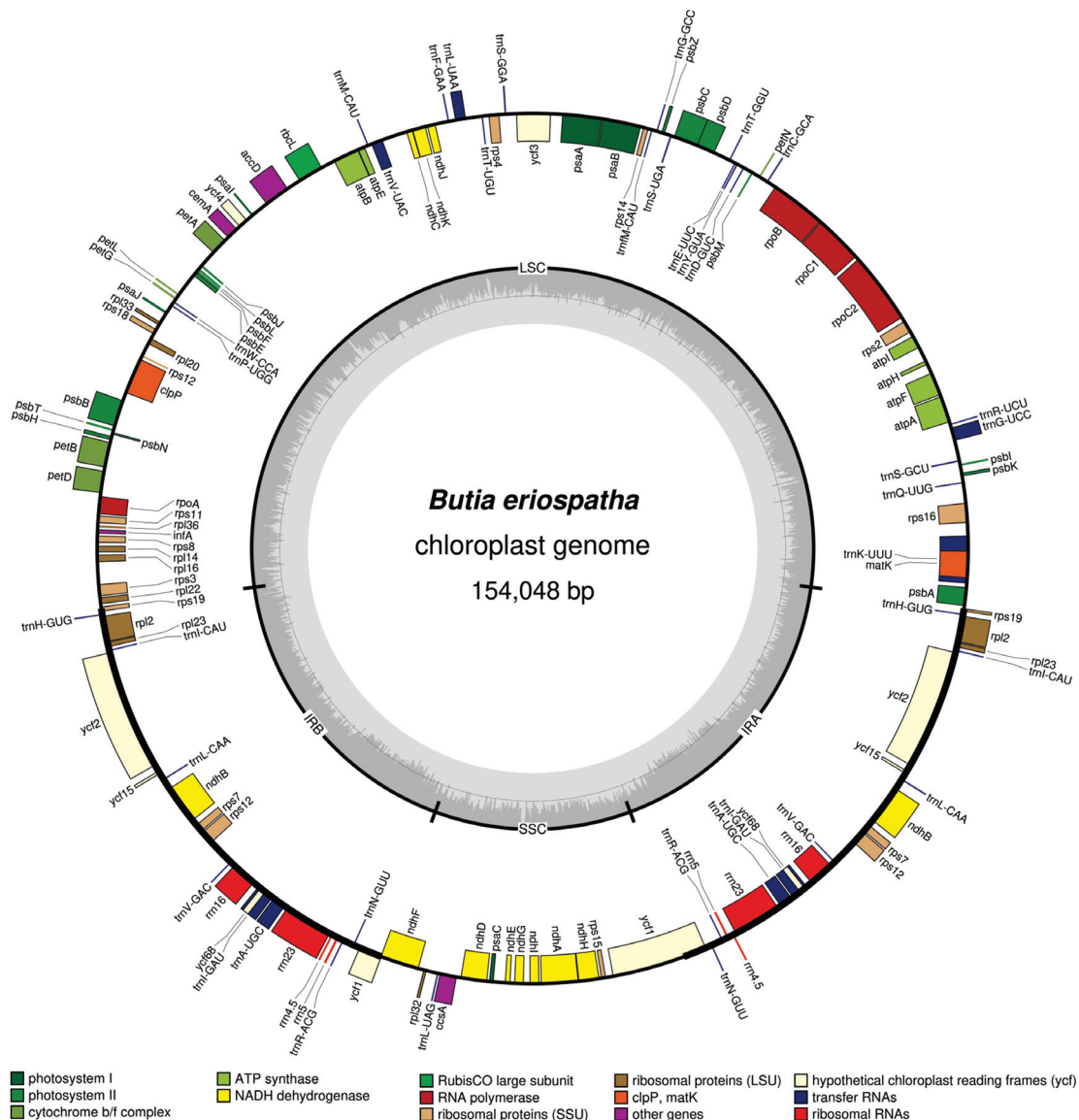
Send correspondence to Leila do Nascimento Vieira. Universidade Federal do Paraná, Programa de Pós-graduação em Botânica, Rua Coronel Francisco H. dos Santos s/n, 81531-980 Curitiba, PR, Brazil. E-mail: leilavieira@ufpr.br.

small single copy (SSC), with 120 to 130 genes, encoding ribosomal RNA (rRNA), transfer RNA (tRNA) and peptides. The plastome size varies from 107 kb to 218 kb, depending on each species (Daniell *et al.*, 2016; Menezes *et al.*, 2018).

Here we report the complete plastome of *B. eriospatha* and describe its structure and gene composition. For this, *B. eriospatha* leaves were collected from a single individual at the Federal University of Santa Catarina - Brazil. DNA isolation with plastid DNA-enrichment was performed according to Vieira *et al.* (2014). The DNA was quantified using Qubit® fluorometer (Invitrogen, Carlsbad, CA) and 1 ng of DNA was used for preparing sequencing libraries with Nextera XT DNA Sample Prep Kit (Illumina Inc., San Diego, CA). Libraries were sequenced on Illumina MiSeq (Illumina Inc.). The obtained paired-end reads (2 x 300 bp) were used for *de novo* assembly in CLC Genomics Workbench 8.0v.

The plastome annotation was performed using DOGMA software (Wyman *et al.*, 2004) followed by manual curation using Geneious (Kearse *et al.*, 2012). The graphical map of plastomes was generated by OrganellarGenomeDRAW (OGDRAW) (Greiner *et al.*, 2019). The complete nucleotide sequence of *B. eriospatha* (MN329806) plastome was deposited in the GenBank database. The junction site of plastomes (LSC/IR, IR/SSC) from the tribe Cocoseae (Arecaceae: Arecoideae) were visualized and compared using IRScope (Amiryousefi *et al.*, 2018).

*Butia eriospatha* plastome is a circular molecule with 154,048 bp in length and the typical quadripartite structure, including a large single copy with 83,805 bp, two inverted repeat regions (26,437 bp) and a small single copy with 17,369 bp (Figure 1). This plastome assembly was realized with 457,423 plastid reads (all trimmed with 0.05 quality



**Figure 1** - Graphical map of *Butia eriospatha* plastid genome. Genes shown on the outside of the circle are transcribed clockwise; genes on the inside are transcribed counterclockwise. Genes were grouped and represented by different colors according their function group. The GC content is represented by the dashed darker grey area in the inner circle, the lighter grey area represents AT content.

scores in CLC Genomics Workbench), which resulted in approximately 820x genome coverage depth. The minimum genome coverage was 206x, which shows the high coverage sequencing of this plastome.

This plastome encodes 113 unique genes, being 79 protein-code genes, 30 tRNA genes and four rRNA genes (Table 1), the same number as the other Cocoseae species (Lopes *et al.*, 2018, 2019). The 20 duplicated genes in the IRs are: eight protein-coding genes (among them, the *ycf1* and *rps19* are partially duplicated), eight tRNA genes, and

all four rRNA genes. Among the 113 genes, 16 contain one intron (10 protein-codes genes, and six rRNA), and three contain two introns (*clpP*, *ycf3* and *rps12*) (Table 1).

The size of the plastome of *B. eriospatha* was compared with others plastomes from species in the same tribe: Cocoseae (Table 2). They have similar length and structure, varying in the LSC from 83,805 bp to 85,192 bp, the SSC from 17,369 bp to 17,639 bp and the IRs from 26,437 bp to 27,092 bp. This variation can be better visualized in the IRs borders (Figure 2). Species from the subtribes Bactridinae

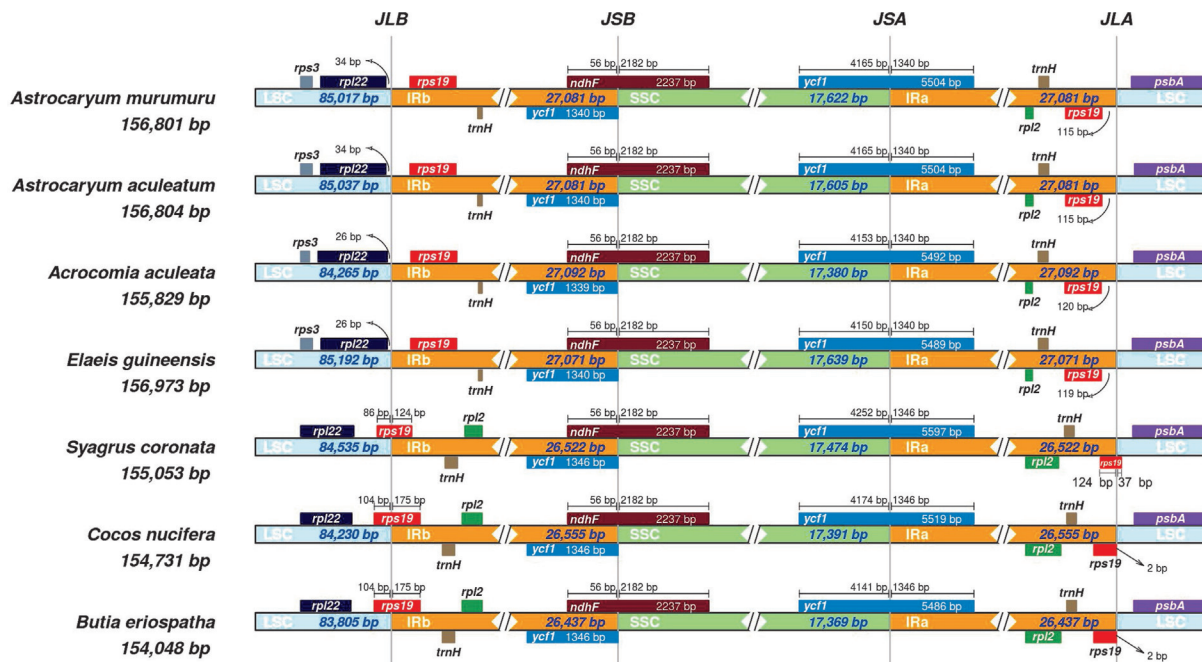
**Table 1** - List of genes identified in the plastome of *Butia eriospatha*.

Group of genes	Name of gene
<i>Gene expression machinery</i>	
Ribosomal RNA genes	<i>rrn4.5<sup>a</sup>; rrn5<sup>a</sup>; rrn16<sup>a</sup>; rrn23<sup>a</sup></i>
Transfer RNA genes	<i>trnA-UGC<sup>ab</sup>; trnC-GCA; trnD-GUC; trnE-UUC; trnF-GAA; trnM-CAU; trnG-UCC<sup>b</sup>; trnG-GCC; trnH-GUG<sup>a</sup>; trnI-CAU<sup>a</sup>; trnI-GAU<sup>ab</sup>; trnK-UUU<sup>b</sup>; trnL-CAA<sup>a</sup>; trnL-UAA<sup>b</sup>; trnL-UAG; trnM-CAU; trnN-GUU<sup>a</sup>; trnP-UGG; trnQ-UUG; trnR-ACG<sup>a</sup>; trnR-UCU; trnS-GCU; trnS-UGA; trnS-GGA; trnT-UGU; trnT-GGU; trnV-GAC<sup>a</sup>; trnV-UAC<sup>b</sup>; trnW-CCA; trnY-GUA</i>
Small subunit of Ribosome	<i>rps2; rps3; rps4; rps7<sup>a</sup>; rps8; rps11; rps12<sup>ab</sup>; rps14; rps15; rps16<sup>b</sup>; rps18; rps19<sup>a</sup></i>
Large subunit of Ribosome	<i>rpl2<sup>ab</sup>; rpl4; rpl16<sup>b</sup>; rpl20; rpl22; rpl23<sup>a</sup>; rpl32; rpl33; rpl36</i>
DNA-dependent RNA Polymerase	<i>rpoA; rpoB; rpoC1<sup>b</sup>; rpoC2</i>
Translational initiation factor	<i>infA</i>
Maturase	<i>matK</i>
<i>Genes for photosynthesis</i>	
Subunits of photosystem I (PSI)	<i>psaA; psaB; psaC; psaI; psaJ; ycf3<sup>b</sup>; ycf4</i>
Subunits of photosystem II (PSII)	<i>psbA; psbB; psbC; psbD; psbE; psbF; psbH; psbI; psbJ; psbK; psbL; psbM; psbN; psbT; psbZ</i>
Subunits of cytochrome b6f	<i>petA; petB<sup>b</sup>; petD<sup>b</sup>; petG; petL; petN</i>
Subunits of ATP synthase	<i>atpA; atpB; atpE; atpF<sup>b</sup>; atpH; atpI</i>
Subunits of NADH dehydrogenase	<i>ndhA<sup>b</sup>; ndhB<sup>ab</sup>; ndhC; ndhD; ndhE; ndhF; ndhG; ndhH; ndhI; ndhJ; ndhK</i>
Large subunit of Rubisco	<i>rbcL</i>
<i>Other functions</i>	
Envelope membrane protein	<i>cemA</i>
Subunit of acetyl-CoA carboxylase	<i>accD</i>
C-type cytochrome synthesis	<i>ccsA</i>
Subunit of protease Clp	<i>clpP<sup>b</sup></i>
Component of TIC complex	<i>ycf1</i>
Unknown function	<i>ycf2<sup>a</sup></i>

<sup>a</sup> Duplicated genes; <sup>b</sup> Genes containing introns

**Table 2** - General features of plastid genomes within the tribe Cocoseae.

Subtribe	Species	Total (bp)	LSC (bp)	IR (bp)	SSC (bp)	Accession number
Attaleinae	<i>Butia eriospatha</i>	154,048	83,805	26,437	17,369	MN_329806
	<i>Syagrus coronata</i>	155,053	84,535	26,522	17,474	NC_029241
	<i>Cocos nucifera</i>	154,731	84,230	26,555	17,391	NC_022417
Elaeidinae	<i>Elaeis guineensis</i>	156,973	85,192	27,071	17,639	NC_017602
Bactridinae	<i>Astrocaryum aculeatum</i>	156,804	85,037	27,081	17,605	MH_537788
	<i>Astrocaryum murumuru</i>	156,801	85,017	27,081	17,622	MH_537787
	<i>Acrocomia aculeata</i>	156,500	84,936	27,092	17,380	NC_037084



**Figure 2** - Comparison of plastid genome LSC, IR, and SSC junction positions within the tribe Cocoseae (Areaceae: Arcoideae). Genes shown below are transcribed counterclockwise and those shown above the lines are transcribed clockwise.

(*Acrocomia* and *Astrocaryum*) and Elaeidinae (*Elaeis*) present the *rps19* gene completely duplicated in the IR region. In contrast, all plastomes sequenced from subtribe Attaleinae (*Butia*, *Cocos*, *Syagrus*) present one complete CDS of *rps19* and one partial copy of *rps19*, thus presenting only one functional copy of the gene (Figure 2). The intergenic spacers (IGS) between *rpl22-rps19* genes and *rps19-psbA* genes, located in the LSC/IR<sub>B</sub> and IR<sub>A</sub>/LSC junctions, respectively, show variations in size within Cocoseae species (Figure 2). The SSC/IR<sub>A</sub> junction varies mainly in the length of the *ycf1* gene portion contained within the SSC. The *ndhF* gene is identical in length among all species studied and part of it is located within the IR<sub>B</sub>, overlapping in part the *ycf1* gene (56 bp). The differences in the IR/LSC junctions between Attaleinae and the sister clades Bactridinae + Elaeidinae might be considered an evolutionary signal and the plastome sequence of *B. eriospatha* may be used in future studies of population genetics and for phylogenetic studies.

## Acknowledgments

This work was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina (FAPESC 14848/2011-2, 3770/2012, and 2780/2012-4). The authors thank to the Brazilian Program of National Institutes of Science and Technology-INCT/Brazilian Research Council-CNPq/MCT.

## Conflict of Interest

The authors declare that they do not have any conflict of interest.

## Author Contributions

LNV, MPG, EMS, MR, and HPFF conceptualized the experiment; JWMS, LNV, HPFF and RSS performed the experiments and data analysis; EMS, MR, and MPG provided the reagents and materials; JWMS, LNV, and HPFF wrote the manuscript. All authors revised the final version of the manuscript.

## References

- Amiryousefi A, Hyvönen J and Poczai P (2018) IRscope: An online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* 34:3030-3031.
- Correa LB, Barbieri RL, Rossato M, Büttow MV and Heiden G (2009) Karyological characterization of *Butia* (Areaceae) palm trees. *Rev Bras Frutic* 31:1111-1116.
- Daniell H, Lin CS, Yu M and Chang WJ (2016) Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. *Genome Biol* 17:1-29.
- Deble LP, Marchiori JNC, Alves FS and Oliveira-Deble AS (2011) Survey on *Butia* (Becc.) Becc. (Areaceae) from Rio Grande do Sul state (Brazil). *Balduinia* 30:3-24.
- Greiner S, Lehwark P and Bock R (2019) OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res* 47:59-64
- Hoffmann JF, Barbieri RL, Rombaldi CV and Chaves FC (2014) *Butia* spp. (Areaceae): An overview. *Sci Hortic* 179:122-131.



- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T *et al.* (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647-1649.
- Lopes AS, Pacheco TG, Silva ON, Cruz LM, Balsanelli E, Souza EM, Pedrosa FO and Rogalski M (2019) The plastomes of *Astrocaryum aculeatum* G. Mey. and *A. murumuru* Mart. show a flip-flop recombination between two short inverted repeats. *Planta* 250:1229-1246.
- Lopes AS, Pacheco TG, Nimz T, Vieira LN, Guerra MP, Nodari RO, Souza EM, Pedrosa FO and Rogalski M (2018) The complete plastome of macaw palm [*Acrocomia aculeata* (Jacq.) Lodd. ex Mart.] and extensive molecular analyses of the evolution of plastid genes in Arecaceae. *Planta* 247:1011-1030.
- Menezes APA, Resende-Moreira LC, Buzatti RSO, Nazareno AG, Carlsen M, Lobo FP, Kalapothakis P and Lovato MB (2018) Chloroplast genomes of *Byrsonima* species (Malpighiaceae): Comparative analysis and screening of high divergence sequences. *Sci Rep* 8:1-12.
- Meerow AW, Noblick L, Salas-Leiva DE, Sanchez V, Francisco-Ortega J, Jestrow B and Nakamura K (2015) Phylogeny and historical biogeography of the cocosoid palms (Arecaceae, Arecoideae, Cocoseae) inferred from sequences of six WRKY gene family loci. *Cladistics* 31:509-534.
- Pereira PEE (2019) Filogenia de *Butia* (Arecaceae): um gênero de palmeiras sul-americano. Thesis, Universidade Federal de Pelotas, 62 p.
- Vieira LN, Faoro H, Fraga HPF, Rogalski M, Souza EM, Pedrosa FO, Nodari RO and Guerra MP (2014) An improved protocol for intact chloroplasts and cpDNA isolation in conifers. *PLoS One* 9:e84792.
- Wyman SK, Jansen RK and Boore JL (2004) Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20:3252-3255.

## Internet Resources

- Heiden G, Ellert-Pereira PE, Eslabão MP (2020) *Butia* in Flora do Brasil 2020 under construction, <http://floradobrasil.jbrj.gov.br/reflora/floradobrasil/FB26574> (accessed June 2020)
- Palmweb - Palms of the world online, <http://www.palmweb.org/node/2> (accessed December 2019).
- WFO (2019) *Butia* (Becc.) Becc., <http://www.worldfloraonline.org/taxon/wfo-4000005828> (accessed December 2019).

*Associate Editor: Marcela Uliano da Silva*

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License (type CC-BY), which permits unrestricted use, distribution and reproduction in any medium, provided the original article is properly cited.