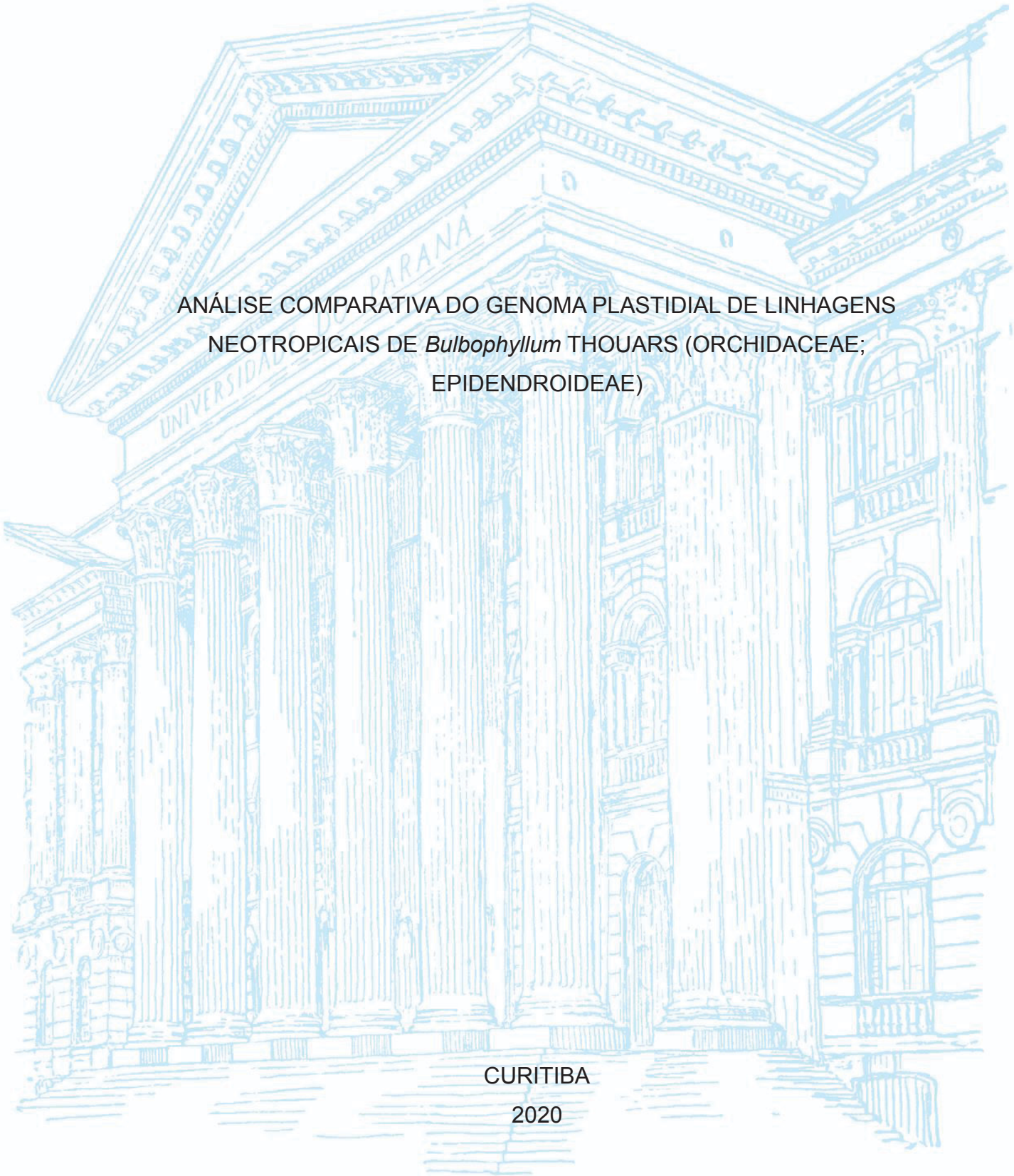
UNIVERSIDADE FEDERAL DO PARANÁ

MICHELLE ESTEFANÍA ZAVALA PÁEZ



ANÁLISE COMPARATIVA DO GENOMA PLASTIDIAL DE LINHAGENS NEOTROPICAIS DE *Bulbophyllum* THOUARS (ORCHIDACEAE; EPIDENDROIDEAE)

CURITIBA

2020

MICHELLE ESTEFANÍA ZAVALA PÁEZ

ANÁLISE COMPARATIVA DO GENOMA PLASTIDIAL DE LINHAGENS NEOTROPICAIS DE *Bulbophyllum* THOUARS (ORCHIDACEAE; EPIDENDROIDEAE)

Dissertação apresentada ao Curso de Pós-Graduação em Ecologia e Conservação, Setor de Ciências Biológicas, da Universidade Federal do Paraná, como requisito parcial à obtenção do título de Mestre em Ecologia e Conservação.

Orientador: Prof. Dr. Eric de Camargo Smidt

Coorientadora: Profa. Dra. Leila do Nascimento Vieira

CURITIBA

2020

Universidade Federal do Paraná. Sistema de Bibliotecas.
Biblioteca de Ciências Biológicas.
(Giana Mara Seniski Silva – CRB/9 1406)

Páez, Michelle Estefanía Zavala
Análise comparativa do genoma plastidial de linhagens neotropicais de *Buldophyllum* Thouars (Orchidaceae; Epidendroideae). / Michelle Estefanía Zavala Páez. – Curitiba, 2020.
58 p.: il.

Orientadora: Eric de Camargo Smidt
Coorientadora: Leila do Nascimento Vieira

Dissertação (mestrado) - Universidade Federal do Paraná, Setor de Ciências Biológicas. Programa de Pós-Graduação em Ecologia e Conservação.

1. Orquidea 2. Genomas 3. Cloroplastos 4. Evolução molecular I. Título II. Smidt, Eric de Camargo- III. Vieira, Leila do Nascimento IV. Universidade Federal do Paraná. Setor de Ciências Biológicas. Programa de Pós-Graduação em Ecologia e Conservação.

CDD (22. ed.) 584.4

MINISTÉRIO DA EDUCAÇÃO
SETOR DE CIENCIAS BIOLOGICAS
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO ECOLOGIA E CONSERVAÇÃO - 40001016048P6

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ECOLOGIA E CONSERVAÇÃO da Universidade Federal do Paraná foram convocados para realizar a arguição da dissertação de Mestrado de **MICHELLE ESTEFANÍA ZAVALA PÁEZ** intitulada: **Análise Comparativa do Genoma Plastidial de Linhagens Neotropicais de *Bulbophyllum* Thouars (Orchidaceae; Epidendroideae)**, que após terem inquirido a aluna e realizada a avaliação do trabalho, são de parecer pela sua Aprovação no rito de defesa.

A outorga do título de mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 18 de Fevereiro de 2020.

LEILA DO NASCIMENTO VIEIRA
Presidente da Banca Examinadora (UNIVERSIDADE FEDERAL DO PARANÁ)

THUANE BOCHORNY DE SOUZA BRAGA
Avaliador Externo (UNIVERSIDADE ESTADUAL DE CAMPINAS)

VIVIANE DA SILVA PEREIRA
Avaliador Externo (UNIVERSIDADE FEDERAL DO PARANÁ - BOTÂNICA)

Para Adrian e Bryan por ter estado comigo em todo este processo de aprendizagem.

AGRADECIMENTOS

Em primeiro lugar quero agradecer ao Brasil por ter me proporcionado a oportunidade de crescer do ponto de vista profissional e pessoal e à CAPES pela bolsa de mestrado fornecida durante meu processo de formação.

Além disso, quero agradecer à coordenação do programa de pós-graduação de Ecologia e Conservação que sempre facilitou cada um dos processos para todos os estrangeiros.

A meu orientador Eric Smidt por ter me dado a oportunidade de trabalhar em seu laboratório e guiado durante todo o processo. A minha coorientadora Leila Viera por todos seus ensinamentos ao longo da realização do presente trabalho, por sua paciência, carinho e especialmente por sua amizade.

A minha colega Anna Victoria Mahuad por ter me ajudado em vários processos que colaboraram para a culminação deste trabalho. A Jeison Magnabosco, por sua colaboração na edição final das imagens do trabalho. A minha amiguinha Raquel Santos por sua amizade e apoio tanto no desenvolvimento deste trabalho como na minha vida, por essas horas de desestresse na academia e as aulas gratuitas de português.

Também quero agradecer à minha família que apesar da distância sempre me brindou todo seu apoio. Obrigada mamãe e papai por ensinar-me a voar alto, porque sem vocês esta etapa não significaria nada.

Ao Adrian, o companheiro de minha vida, que sempre me encoraja a seguir em frente e enche minha vida de amor e luz. Obrigada por seu valente e incondicional amor.

RESUMO

Bulbophyllum é o maior gênero Pantropical de Orchidaceae, com *ca*. 2.200 espécies, embora sua distribuição não seja homogênea em toda a sua extensão. A região Paleotropical é a mais rica em espécies deste gênero, com centenas ocorrendo na Ásia, seguida pela África e depois pelo Neotrópico com aproximadamente 60 espécies. Na região Neotropical, o gênero possui seis linhagens, duas acima do Equador e quatro principalmente na Mata Atlântica e no Cerrado brasileiro. Embora a filogenia e taxonomia do gênero sejam bem compreendidas nas Américas, alguns complexos de espécies ainda precisam de mais estudos em relação a sua delimitação e portanto, é necessário o desenvolvimento de marcadores moleculares específicos. Pesquisas anteriores sugerem que o genoma plastidial aumenta a resolução filogenética em baixos níveis taxonômicos e é uma ferramenta eficaz para identificar sequencias com alta variação. Diante disso, neste estudo foram sequenciados o genoma plastidial de oito espécies de *Bulbophyllum*, representando cinco seções neotropicais, usando a plataforma de sequenciamento Illumina MiSeq. Todos os genomas conservaram a típica estrutura dividida em quatro partes e embora a estrutura geral dos genomas foi conservada, foram detectadas diferenças na composição dos genes da família *ndh* e no comprimento total. O comprimento total foi determinado pela contração e expansão da cópia única menor como resultado da perda independente dos genes *ndh* presentes nessa região. A análise de seleção positiva indicou que os genes codificadores de proteínas eram geralmente bem conservados. No entanto, foram identificados 95 possíveis sítios sob seleção positiva distribuídos em quatro genes. Além disso, um total de 54 microssatélites foram identificados e iniciadores específicos para o gênero *Bulbophyllum* foram desenvolvidos. As dez sequências mais variáveis (*trnR-atpA, trnM-aptE, ccsA-ndhD, clpP-psbB, trnS-trnG, psbB-psbT, atpH-atpI, psbK-psbI, rpl32-trnL* e *matK-trnK*) foram propostas como potenciais marcadores moleculares para identificar espécies deste gênero e melhorar a resolução filogenética. O presente estudo fornece ótimos recursos moleculares para *Bulbophyllum*, que inclui um total de 54 microssatélites e marcadores moleculares das dez principais regiões mais variáveis com seus primers específicos.

Palavras-chave: genoma do cloroplasto. Orquídeas neotropicais. Evolução molecular. NGS. Marcadores moleculares.

ABSTRACT

Bulbophyllum is the largest Pantropical genus of Orchidaceae, with ca. 2,200 species, although its distribution is not homogeneous over its entire range. The Paleotropical region is the richest in species of this genus, with hundreds occurring in Asia, followed by Africa and then the Neotropics that has about 60 spp. In the Neotropics, the genus has six lineages, two above the Equator and four mainly in Brazil Atlantic Rainforest and Cerrado. Although the phylogeny and taxonomy of the genus are well understood in the Americas, some complex - species need more studies in its delimitation. Therefore, the development of specific molecular markers is required. Previous research suggests that the plastid genome increases phylogenetic resolution at low taxonomic levels and considers it an effective tool for detected effective divergence sequences. In this study, we sequenced the complete plastid genome of eight *Bulbophyllum* species, representing five Neotropical sections, using an Illumina MiSeq Platform. All genomes conserve the typical quadripartite structure and although the general structure of plastid genomes is conserved, differences in *ndh* genes composition and total length were detected. The total length was determined by the contraction and expansion of the small single-copy region due to the independent loss of the seven *ndh* genes present in this region. Positive selection analyses indicated that protein-coding genes were generally well conserved, however, we identified 95 putative sites under positive selection distributed in four genes. Furthermore, a total of 54 polymorphic SSRs were identified and specific primer pairs were developed. In addition, we propose ten sequences (*trnR-atpA*, *trnM-aptE*, *ccsA-ndhD*, *clpP-psbB*, *trnS-trnG*, *psbB-psbT*, *atpH-atpI*, *psbK-psbI*, *rpl32-trnL*, and *matK-trnK*) as potential molecular markers to identify *Bulbophyllum* species and improve the phylogenetic resolution of this genus. The present study provides great molecular resources for *Bulbophyllum*, which includes a total of 54 microsatellites and molecular markers from the ten most variable regions with their specific primers.

Keywords: chloroplast genome. Neotropical orchids. Molecular evolution. NGS. Molecular markers.

SUMARIO

1 INTRODUCTION 10
2 MATERIALS AND METHODS 13
2.1 PLASTID GENOME SEQUENCING AND ASSEMBLY 13
2.2 PLASTID GENOME FEATURES 14
2.3 CODON USAGE 14
2.4 MOLECULAR EVOLUTION ANALYSIS ON PROTEIN-CODING GENES 15
2.5 CHARACTERIZATION OF REPEAT SEQUENCES AND SIMPLE SEQUENCE REPEAT (SSRS) 15
2.6 HYPERVARIABLE REGIONS 15
2.7 PHYLOGENETIC ANALYSIS 16
3 RESULTS 18
3.1 PLASTID GENOME FEATURES 18
3.2 CODON USAGE 20
3.3 MOLECULAR EVOLUTION ANALYSIS ON PROTEIN-CODING GENES 21
3.4 CHARACTERIZATION OF REPEAT SEQUENCES AND SIMPLE SEQUENCE REPEAT (SSRS) 22
3.5 HYPERVARIABLE REGIONS 24
3.6 PHYLOGENETIC ANALYSIS 26
4 DISCUSSION 27
4.1 PLASTID GENOME FEATURES 27
4.2 CODON USAGE 28
4.3 MOLECULAR EVOLUTION ANALYSIS ON PROTEIN-CODING GENES 29
4.4 CHARACTERIZATION OF REPEAT SEQUENCES AND SIMPLE SEQUENCE REPEAT (SSRS) 30
4.5 HYPERVARIABLE REGIONS 30
4.6 PHYLOGENETIC ANALYSIS 32

5 CONCLUSION.....33

REFERENCES.....34

APPENDIX.....42

Comparative plastid genome of Neotropical *Bulbophyllum* Thouars (Orchidaceae; Epidendroideae) lineages

The present work will be sent to Frontiers in Plant Science journal

1 INTRODUCTION

Bulbophyllum is the largest Pantropical genus of Orchidaceae, with ca. 2200 species (WCSP, 2018). However, its distribution is not homogeneous over its entire range. The genus reaches its highest richness in the Paleotropics, with hundreds of species occurring in Asia, followed by Africa and the Neotropics with ca. 60 species (SMIDT et al., 2011; PRIDGEON et al., 2014). In the Neotropics, the genus is represented by six lineages, two near above the Equator and four mainly in Brazil's Atlantic Rainforest and Cerrado (SMIDT; BORBA, 2007; SMIDT et al., 2011).

The *Bulbophyllum* species are not only recognized worldwide for their diversity, but also for the potential use of their compounds in the field of traditional medicine and agricultural pest management. These orchids contain aromatic compounds whose properties have been applied to human health, such as in the treatment of tumors, conception problems and as antifebriles (WU et al., 2006; CHEN et al., 2008; LALITHARANI et al., 2011). Furthermore, the volatile compounds of *Bulbophyllum* orchids are used in integrated pest management strategies. For instances, the Tephritidae flies, a pest of many economic crops, are attracted inward of the control traps using the aromas of *Bulbophyllum* orchids (TAN et al., 2006; TAN; NISHIDA, 2007; JALEEL et al., 2018).

Due to their ornamental value, species diversity and high levels of endemism, *Bulbophyllum* species have been widely studied in the Neotropics (SMIDT; BORBA 2007; SMIDT et al., 2007; MANCINELLI; SMIDT, 2012). As a result, multiple taxonomic (BORBA et al., 1998; RIBEIRO et al., 2008; NUNES et al., 2014, 2015, 2017) and phylogenic (SMIDT et al., 2011, 2013) studies have been carried out. Concerning to molecular phylogeny, several plastids (*psbA-trnH* and *trnS-trnG* intergenic spacers) and a nuclear marker (ITS spacer) have been used to investigate evolutionary relationships among the neotropical *Bulbophyllum* species (SMIDT et

al., 2011, 2013). However, despite the number of related studies and the use of molecular tools, some complex - species need deeper investigation for their delimitation due to the evidence of natural hybridization (BORBA; SEMIR, 1998; MANCINELLI; SMIDT, 2012) and introgression (AZEVEDO et al., 2006) across their species range.

Previous research suggests that the plastid genome increases phylogenetic resolution at low taxonomic levels and considers it an effective tool in population genetics (Zhang et al., 2016; Zhitao et al., 2017; Agostino et al., 2018). In addition, the study of the plastid genome to the development of specific molecular markers for phylogeographic analysis has great potential (MARIAC et al., 2014). In Orchidaceae, with the exception of several myco-heterotrophic species (FENG et al., 2016; YUAN et al., 2018), the plastid genome has the typical quadripartite structure: two inverted repeated regions (IR), which are separated by a smaller single copy region (SSC) and a larger single copy region (LSC) (CHANG et al., 2006; DELANNOY et al., 2011; PAN et al., 2012; YANG et al., 2013; LUO et al., 2014; ZHITAO et al., 2017). The size of the plastid genome within orchids is relatively small with range from 0.01 to 0.17 Mb (LIN et al., 2015; SCHELKUNOV et al., 2015). Genetic content and gene order are generally conserved, but some variations have been reported. Most of these variations are in total length, composition of the *ndh* gene complex, and the loss of many photosynthetic-related genes in myco-heterotrophic species (GRAHAM et al., 2017; ZHITAO et al., 2017).

In this study, we present the complete plastid genome of eight *Bulbophyllum* species, representing five Neotropical sections. First, we compared the gene content, gene order and codon use among them. Second, the amino acid sites under positive selection in all protein-coding genes are located. Third, we determined the distribution and location of long and short repeated sequences among these genomes, including potential microsatellite markers. Fourth, the sequence variability is calculated to determine the top ten hypervariable regions and primers are provided for these sequences. Finally, the evolutionary significance of the hypervariable regions for *Bulbophyllum* species is evaluated using phylogenetic analysis. The present results provide abundant information for species identification and to increase the phylogenetic resolution of *Bulbophyllum*. Also, the development of

specific molecular markers will help in phylogeographic analyses with implications for species management and conservation.

2 MATERIALS AND METHODS

2.1 PLASTID GENOME SEQUENCING AND ASSEMBLY

Plastid-enriched plant DNA was extracted from fresh leaves of eight *Bulbophyllum* species (Table 1) using Sakaguchi et al. (2017) methodology, aiming to reduce the amount of initial material required for plastid isolation. Afterwards, plastid-enriched DNA (ptDNA) was extracted following Doyle and Doyle (1987) protocol. Purification of ptDNA was performed with DNA Clean and Concentrator kit (Zymo Research, Orange, CA).

Table 1- *Bulbophyllum* species sequenced with their collection data.

Species	Section	Locality	Collector/ Collection number / Herbarium
Bulbophyllum weddellii (Lindl.) Rchb.f	*Didactyle*	Brazil, Minas Gerais, Serra da Piedade	C. Fiorini et. al (HBCB)
Bulbophyllum exaltatum Lindl.	*Didactyle*	Brazil, Minas Gerais, Santa Rita de Caldas	C. Fiorini, 218 (HBCB)
Bulbophyllum steyermarkii Foldats	*Furvescens*	Ecuador, Azuay, Gualaceo	M. Cerna 4403 (UPS)
Bulbophyllum epiphytum Barb.Rodr	*Micranthae*	Brazil, Paraná, Pirai do Sul	E. C. Smidt 1084 (UPCB)
Bulbophyllum mentosum Barb.Robr.	*Micranthae*	Brazil, Minas Gerais, Parque Nacional Sempre Vivas	C. Fiorini, 323 (HBCB)
Bulbophyllum regnellii Rchb.f.	*Napelli*	Brazil, Parana, Piraquara Mananciais da Serra – Reservatório Carvalho	E. C. Smidt, 1081 (UPCB)
Bulbophyllum granulosum Barb.Rodr	*Napelli*	Brazil, Santa Catarina, Pico Garuva	Mancinelli, W. S., 1059 (UPCB)
Bulbophyllum plumosum Barb.Rodr.	*Xiphizusa*	Brazil, Paraná, Tibagi – Parque Estadual do Guartelá	D. Imig et al, 606 (HAC)

Source: The author (2019)

Total ptDNA was sequenced on an Illumina MiSeq® following instruction of DNA Nextera XT Sample Prep Kit (Illumina™). The raw data were trimmed according to default parameters of CLC Genomics Workbench 8.0

(https://www.qiagenbioinformatics.com/) and a hybrid reference-guided *de novo* assembly approach was performed to assemble the genomes. First, the trimmed data was assembled according to complete plastid genome sequence of *Dendrobium officinale* Kimura & Migo (NC_024019.1) as reference to generate a consensus sequence. Then, multiples contigs were created through de *novo* assembly strategy and gap-closing process in consensus sequence was developed with high-quality contigs.

Dual Organellar GenoMe Annotator software (WYMAN et al., 2004) and Geneious R7 (KEARSE et al., 2012) were used to annotate the plastid genomes, using default values to predict genes encoding proteins, transfer RNAs (tRNAs), and ribosomal RNAs (rRNAs). All plastid genomes were checked manually, and the codon positions were investigated and determinates by BLASTX against the NCBI protein database, with *D. officinale* as reference. Gene maps and distribution were drawn by Organellar Genome DRAW V1.1 (LOHSE et al., 2007).

2.2 PLASTID GENOME FEATURES

For each plastid genome total length, numbers of genes encoding proteins (CDS), transfer RNAs (tRNAs), and ribosomal RNAs (rRNAs) were calculated. Replicated genes in both inverted repeats regions were identified and genes with two or more introns were classified. Junction positions between single copies and IR regions were compared. The GC content was calculated using Geneious R7 (KEARSE et al., 2012).

2.3 CODON USAGE

All CDS for each plastid genome were extracted using Geneious R7 (KEARSE et al., 2012), in order to determine the distribution of codon usage. Relative synonymous codon usage (RSCU) ratio was calculated using the R package SeqinR (CHARIF; LOBRY, 2007). RSCU values bigger than one represent codons used more frequently than expected. In the other hand, RSCU values smaller than one signify that a codon is utilizing less frequently than expected. The RSCU distributions were illustrated in the form of heatmaps using the Heatmapper program (BABICKI et al., 2016).

2.4 MOLECULAR EVOLUTION ANALYSIS ON PROTEIN-CODING GENES

A Bayesian inference approach was applied to identify amino acid sites under positive selection or purifying selection in all CDS. The evolutionary model M8 were run on The Selecton server web site (STERN et al., 2007). When positive selection was detected a likelihood ratio tests were run between the model M8a (null model) and M8 model (alternative model) to test whether positive selection was significant. Furthermore, gene divergence analysis was performed on Selecton using default parameters. The pairwise distances and branch lengths were computed applying the ML criterion under a codon model (NIELSEN; YANG, 1998).

2.5 CHARACTERIZATION OF REPEAT SEQUENCES AND SIMPLE SEQUENCE REPEAT (SSRS)

Direct, reverse, and palindromic repeats sequences were identified by REPuter online program (KURTZ et al., 2001) according with the following criteria: Hamming distance of 3, 90% sequence identity, and minimum repeat size of 30 bp. Simple sequence repeats (SSRs) were located using MISA-web program (BEIER et al., 2017) with the search parameters set to: ≥ ten repeats units for mononucleotide SSRs, ≥ five repeats units for dinucleotide SSRs, and ≥ three repeats units for di, tetra, pentha and hexanucleotide SSRS. Microsatellites primers were designed for polymorphic SSRs which were presented in at least four species using Websat (MARTINS et al., 2009). The following parameters were used: product length of 100 to 500 bp, primer length of 18 to 27 bp and GC content of 40 to 60 % with a 1 °C as maximum difference between fusion temperatures of the left and right primers.

2.6 HYPERVARIABLE REGIONS

In order to identify the top ten hypervariable regions all plastid genomes were alignment with the progressive Mauve algorithm (DARLING et al., 2004). After, the CDS, introns, and intergenic spacers (IGS) with minimum length of 150 bp and that were flanked by the same region were manually extracted. The number of mutations and Indel events were calculated by DnaSP v5 (ROZAS et al., 2004) and their rate was calculated as: mutations / 100 bp and Indels / 100 bp. The sequence variability was computed as: SV = (Total number of mutations + Total number of Indels events) /

(Conserved sites + Total number of mutations + Total number of Indels events) × 100%. The SV formula was based in SHAW et al. (2014) method but, due to Indels presents higher levels of homoplasy than mutations, Indels were considered as events instead of sites in the alignment (INGVARSSON et al., 2003). Also, the parsimony informative sites were counted by DnaSP v5 program.

Specific primers were designed for the top ten variable sequences using PRIMER3 (http://bioinfo.ut.ee / primer3- 0.4.0 /) with following parameters: maximum product length of 1000 bp, primer length of 18 to 27 bp and GC content of 40 to 60 % with a 2 °C as maximum difference between fusion temperatures of the left and right primers.

2.7 PHYLOGENETIC ANALYSIS

Aiming to verify the evolutionary significance of hypervariable regions found in this study, phylogenetic analyses were performed based on two data set: 1) the complete chloroplast genome sequences and 2) the top ten hypervariable regions. Individual sequence alignments were performed using MAFFT and then concatenated for each data set by SequenceMatrix (VAIDYA et al., 2011).

Maximum parsimony (MP), Bayesian inference (BI), and Maximum likelihood (ML) analyses were run for each set. MP analyses were performed by PAUP v4b10 (SWOFFORD, 2003), with *Dendrobium aphyllum* (Roxb.) C. E. C. Fisch (NC_035322.1) as an outgroup. Heuristic search was conducted with 1,000 bootstrap replicates, tree bisection-reconnection (TBR) branch swapping, and simple stepwise addition. All characters had equal weight; gaps were treated as missing.

ML trees with 1,000 bootstrap replications (BS) and 1,000 replicates were constructed using IQ-tree 1.6.11 (TRIFINOPOULOS et al., 2016). The best fit models for each data set were calculated by ModelFinder (KALYAANAMOORTHY et al., 2017). The K3Pu+F+R2 model was used for the complete chloroplast genome matrix. For the second set, the evolution model was calculated for each sequence with IQ-tree 1.6.11 (Annex 1). The resulting trees were represented and edited using FigTree software v1.4.1 (available at: http://tree.bio.ed.ac.uk).

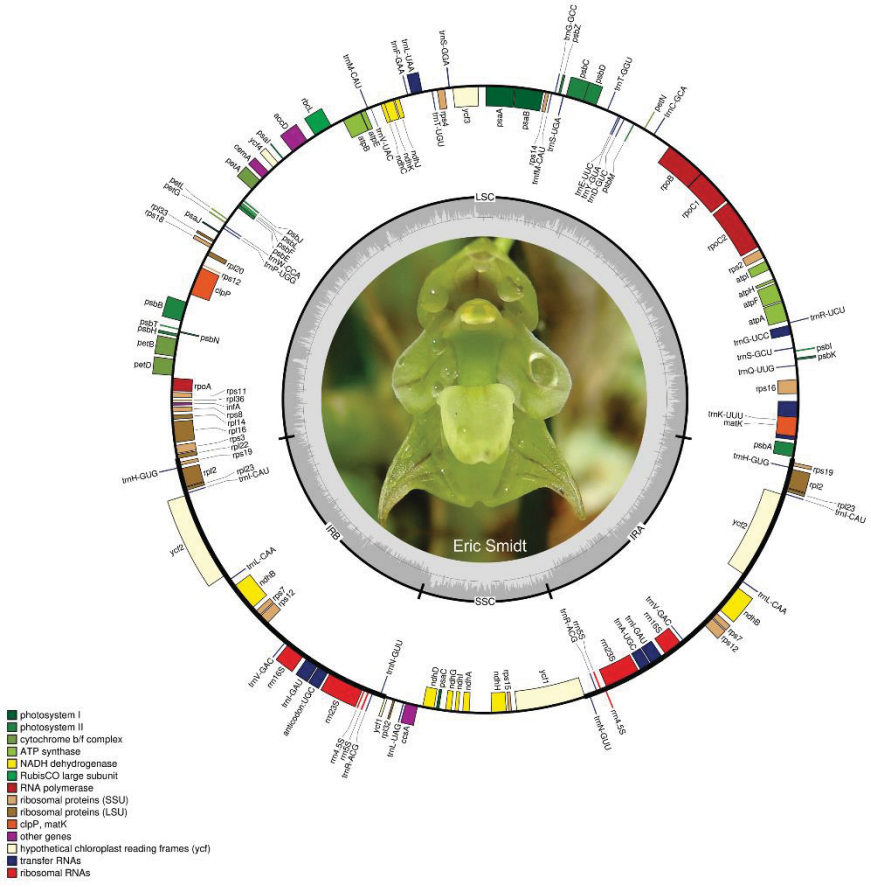
BI analysis was constructed using MrBayes 3.2.7 (RONQUIST; HUELSENBECK, 2003) and the best fit model for each set was determined by the Bayesian Information Criterion (BIC) on jModeltest version 2.1.10 (DARRIBA et al., 2012, Annex 1). Four independent Markov chains Monte Carlo (MCMC) were run twice, each with 3,000,000 million generations. The trees were sampled every 1,000 generations and the first 25% of sampling were discarded. The remaining trees were used to build a majority-rule consensus tree.

3 RESULTS

3.1 PLASTID GENOME FEATURES

The eight *Bulbophyllum* plastid genomes ranged from 146,401 bp to 151,493 bp in length, being the smallest *B. plumosum* while the largest *B. regnellii*. The genomes have a typical quadripartite structure (Figure 1; Annex 2), consisting of a pair of IRs with a length of 25,465 bp (*B. granulosum*) to 26,340 bp (*B. mentosum*), separated by the LSC with a length of 82,354 bp (*B. epiphytum*) to 84,868 bp (*B. regnellii*), and SSC with a length of 11,088 bp (*B. plumosum*) to 16,048 bp (*B. weddellii*, Annex 3). In overall, the GC content was similar between all plastid genomes (36.6 – 36.8 %; Table 2).

Figure 1 - Gene organization of the *Bulbophyllum regnellii* plastid genome. Genes shown on the outside of the circle are transcribed clockwise; genes on the inside are transcribed counterclockwise. The color of the gene boxes indicates the functional group to which the gene belongs. The thick lines indicate the length of the inverted repeats (IRA and IRB), separated by the large single-copy (LSC) and small single-copy (SSC). The dashed darker grey area in the inner circle indicates genome GC content, while the lighter grey area shows AT content.



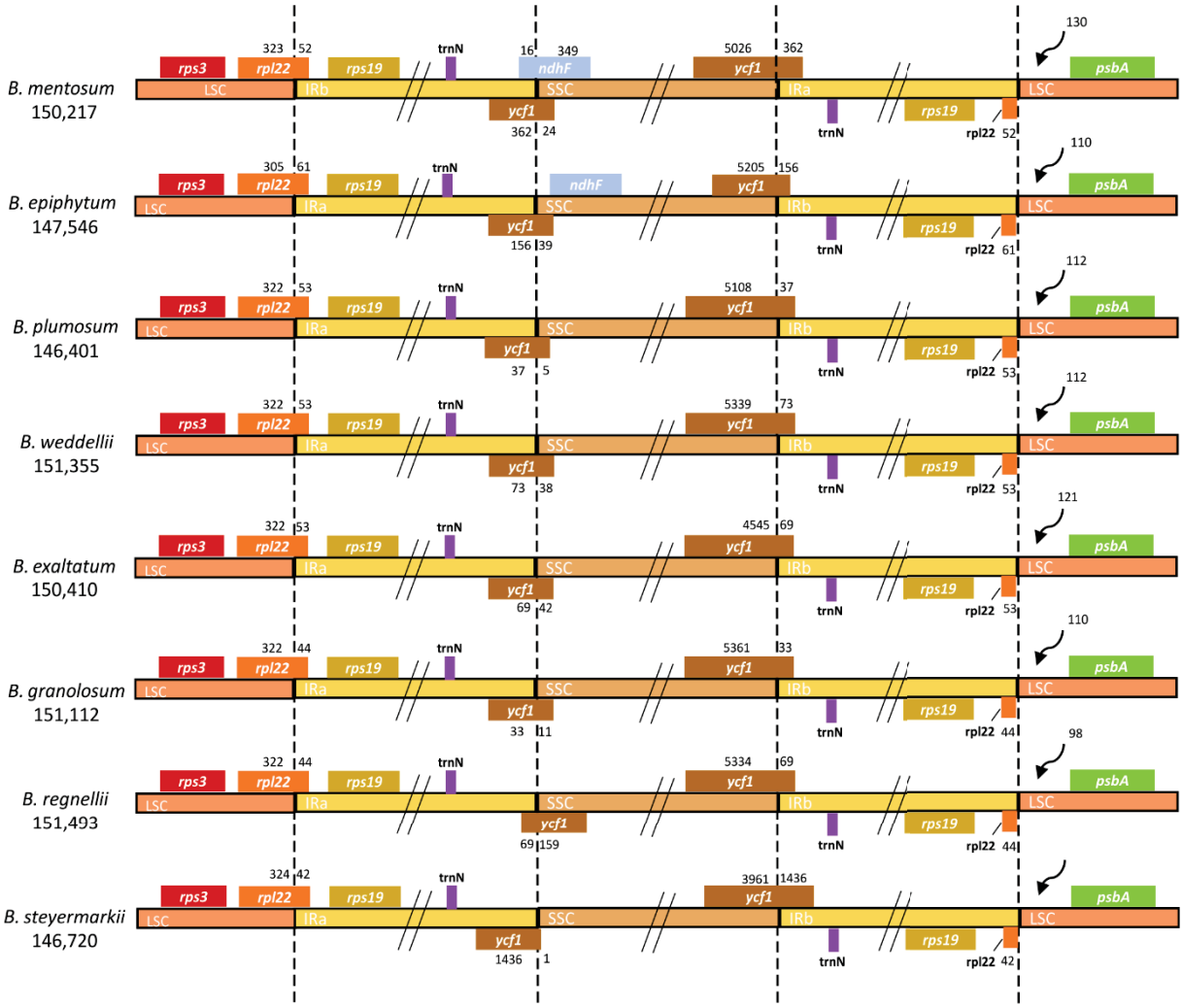
Source: The author (2019)

Table 2 - General characteristics of plastid genomes of eight *Bulbophyllum* species

Species	***B. mentosum***	***B. epiphytum***	***B. plumosum***	***B. weddellii***	***B. exaltatum***	***B. granulosum***	***B. regnellii***	***B. steyermarkii***
GenBank accession	MN604056	MN737573	MN580547	MN604059	MN604054	MN604055	MN604057	MN604058
Total length (bp)	150,217	147,546	146,401	151,355	150,410	151,112	151,493	146,720
Length of LSC region	83,642	82,354	83,260	83,450	83,335	84,492	84,868	83,488
Length of SSC region	13,895	13,497	11,089	16,049	15,380	15,690	15,541	11,321
Length of IR region	26,340	25,847	26,026	25,928	25,847	25,465	25,542	25,955
Total GC content (%)	36.7	36.7	36.6	36.6	36.8	36.7	36.7	36.7
LSC GC content (%)	34.1	34.1	33.7	34.7	34.2	34.3	34.2	34.2
SSC GC content (%)	28.3	28.5	27.6	28.9	29.1	29.1	29.1	27.9
IR GC content (%)	43.1	43.2	43	43.2	43.3	43.1	43.1	42.8
Coverage	224.17	45.26	235.29	133.38	85.94	194.7	148.57	247.55
Number of genes	102	102	102	102	103	102	102	102
Number of CDS	68	68	68	68	69	68	68	68
Number of tRNA	30	30	30	30	30	30	30	30
Number of rRNA	4	4	4	4	4	4	4	4
Number of pseudogenes	10	9	7	10	8	10	10	6

A total of 102 genes were predicted for each plastid genome, among which 68 are CDS, 30 tRNA, and four rRNA. Eight CDS (*rps12, rps7*, *ycf2*, *rpl23*, *rpl2*, *rpl22*, *rps19*, and *ycf1*), eight tRNA (*trnA-UGC*, *trnH(GUG)*, *trnI(CAU)*, *trnI(GAU), trnL(CAA)*, *trnN(GUU)*, *trnR(ACG)*, and *trnV(GAC)*), and four rRNA (*rrn16*, *rrn23*, *rrn4*.5, and *rrn5*) were duplicated in the IR regions (Annex 4). Only a small fragment of the *ycf1* was found in IRb/SSC junction (Figure 2).

Figure 2 - Comparison of LSC, IR, and SSC junction positions among eight *Bulbophyllum* plastid genomes. Genes shown below are transcribed reversely and those shown above the lines are transcribed forward.

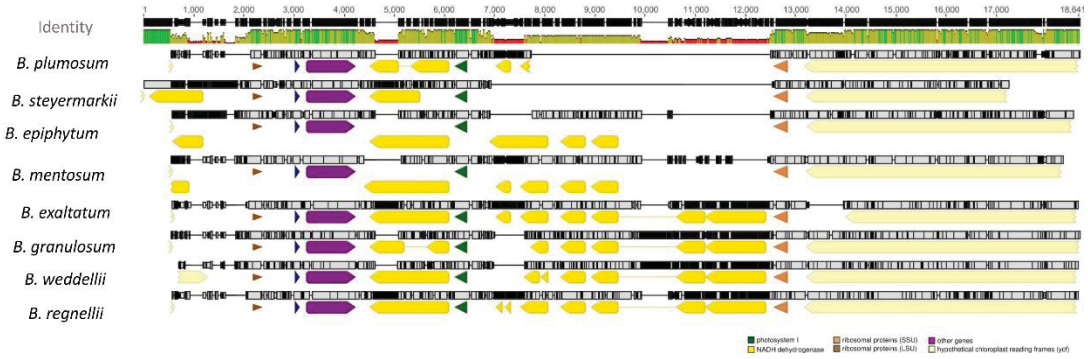


Source: The author (2019)

Additionally, the most *ndh* genes were truncated or completely lost among the eight plastid genomes. However, the *ndhG* gene showed full reading frame in *B. exaltatum*. Seven of eleven *ndh* genes were located in the SSC region and their length are correlated with the expansion and contractions of the SSC region (*r* =

0.94, p = < 0.001) and, the expansion and contractions of the SSC region determined the total length of plastid genomes (r^2 = 0.89, $F_{1, 6}$ = 57.27, p = < 0.001). The SSC length difference between the largest and smallest genomes was the 4,960 bp and this value corresponds to 97% de total length difference among the plastid genomes (Figure 3).

Figure 3 – MAFFT alignment of the SSC region of all plastid genomes. The yellow blocks are representing the NDH genes. The green lines above represent the high similarity between the sequences while the red indicates the contrary. The genomes are organized in descending order according to their total length



Source: The author (2019)

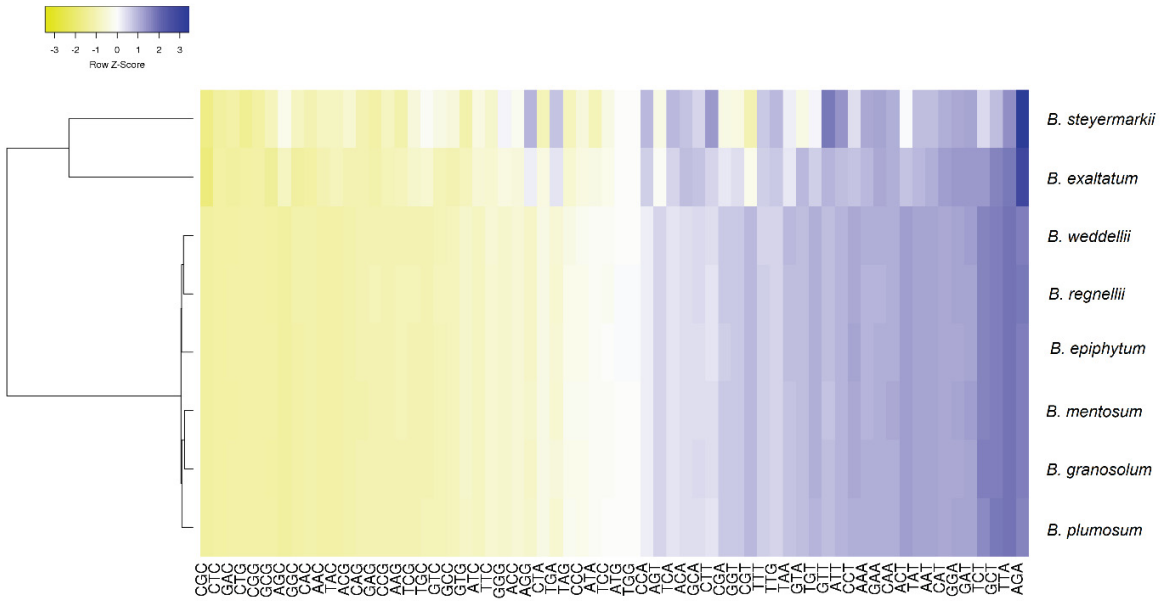
The IRs junctions of the eight plastid genomes were compared. In all plastid genomes, the LSC/IRb and IRb/SSC junctions were *rpl22* gene and *ycf1* pseudogene respectively, while in the IRa/SSC and IRa/LSC junctions were *ycf1* and *rpl22* gene. At the LSC/IRb border, the IRb expanded with up to 61 bp towards the *rpl22* gene and with up to 1,436 bp towards the *ycf1* pseudogene in the IRb/SSC border. Furthermore, only in the IRb/SSC border of *B. mentosum* was located the *ndhF* pseudogene with 16 bp located in the IRb region. The *ndhF* pseudogene was closed but not overlapping the IRb/SSC junction in *B. epiphytum* (Figure 2).

3.2 CODON USAGE

The codon usage frequency and relative synonymous codon usage (RSCU) were performed with 68 unique CDS shared among the eight plastid genomes. A total range of 18,831 to 19,306 codons were counted, among these the most abundant codon was *AAA* (4.27%), while *CGC* (0.36%) was the least abundant. The plastid genomes encode *ATG* as the initiation codon for the most protein-coding genes. However, in the *matK* (*ATT*), *rps19* (*GTG*), and *rpl2* genes (*ACG*) alternative initiation codons were found. Leucine was the most abundant amino acid, ranged between

1,258 and 2,687 codons in *B. mentosum* and *B. plumosum*, respectively. While, cysteine was the least abundant ranged between 209 codons in *B. exaltatum* and 776 codons in *B. mentosum* (Annex 5). A relative synonymous codon usage (RSCU) indicates that 29 of 32 codons that ended with G/C were not frequently used in the plastid genomes, whereas 28 of 32 codons that ended with A/T had high RSCU values and were frequently used among the eight plastid genomes (Figure 4, Annex 6).

Figure 4 - Relative synonymous codon usage ratio (RSCU) of protein-coding genes of the eight *Bulbophyllum* plastid genomes. Yellow indicates a high RSCU value and blue indicates a low RSCU value.

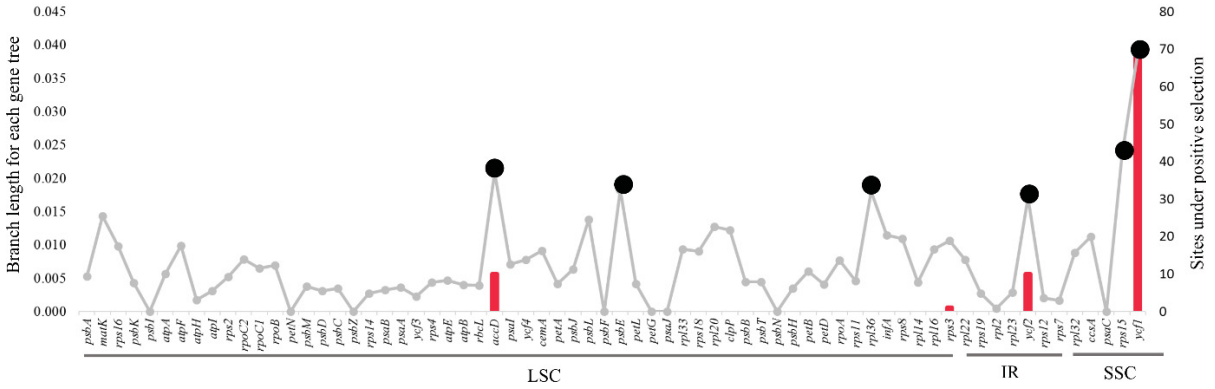


Source: The author (2019)

3.3 MOLECULAR EVOLUTION ANALYSIS ON PROTEIN-CODING GENES

The analysis conducted in the Selecton Server to investigate sites under positive selection indicated that 89 putative sites were under positive selection. These sites were distributed in four out of 68 shared proteins coding genes within the plastid genomes. The most putative sites under positive selection (adaptative selection) were found in *ycf1* gene (68 sites) followed by *ycf2, accD,* and *rps3* genes with 10, 10, and 1 site respectively. Moreover, according to the average branch length of each protein-coding gene tree, the most divergent genes were *ycf1*, *rps15*, *accD*, *psbE*, *rpl36*, and *ycf2* (Figure 5).

Figure 5 - Molecular evolution of 68 protein-coding genes shared among the eight plastid genomes. The grey lines represent de branch length average for each gene tree and the pink bars the number of putative sites under positive selection. The black circles indicated the six most divergent genes.

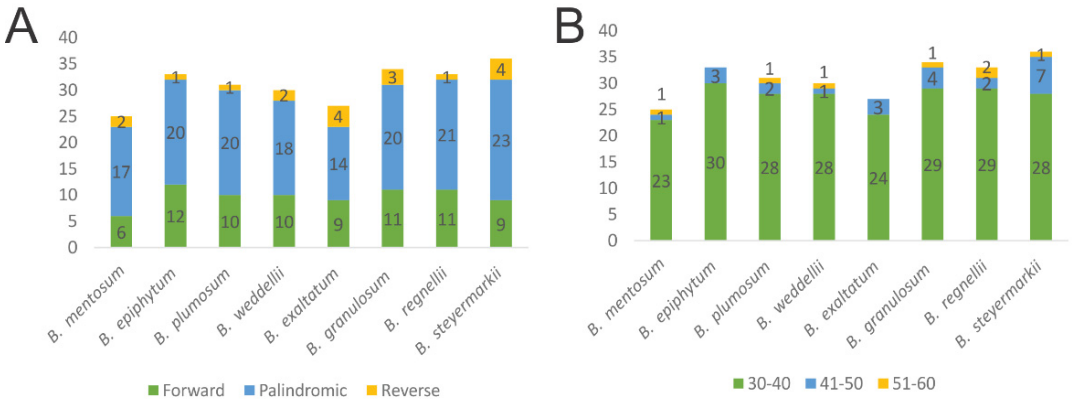


Source: The author (2019)

3.4 CHARACTERIZATION OF REPEAT SEQUENCES AND SIMPLE SEQUENCE REPEAT (SSRS)

The forward, reverse and palindromic repeats (> 30 bp) were counted and compared among the eight plastid genomes. In total, 249 repeat sequences were identified among the all plastid genomes, including 78 forward repeats, 153 palindromic repeats, and 18 reverse repeats (Figure 6A). *B. steyermarkii* presented the highest number of repeats (36), while *B. mentosum* had the fewest (25; Annex 7). The majority of repeat sequences were distributed in the intergenic spacers (54.62 %) and the most frequent length was 30 to 40 bp (Figure 6B, Annex 8) Four palindromic repeats were shared among the eight plastid genomes (Table 3).

Figure 6 - A) Frequency of forward, reverse, and palindromic repeats. B) Length frequency of tandem repeats.



Source: The author (2019)

Table 3 - Palindromic repeat sequences shared between the all *Bulbophyllum* plastid genomes.

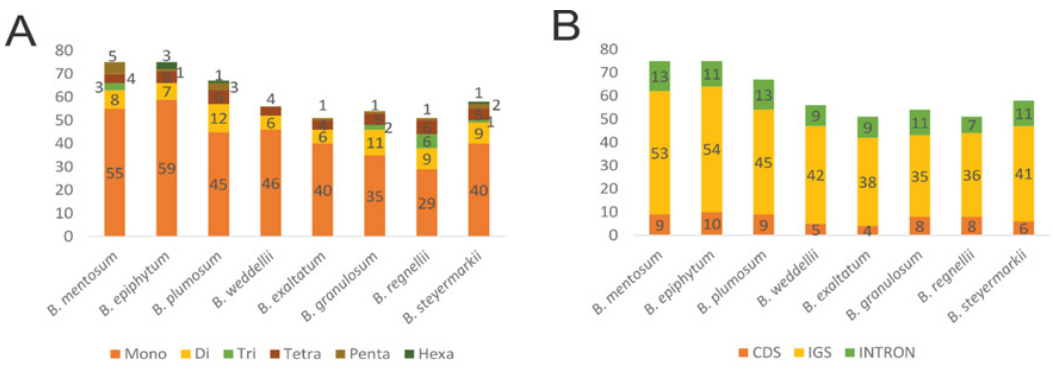
Size (bp)	Repeat Sequence	Location Region
30	ATTTCAAAAAAAAATCGATTTTTTTTTGAA	Gene (*ccsA, ccsA*)
37	TTATGTTTTGTATATTTGGATCAAATATACAAAACAT	IGS (*rps15-ycf1, rps15-ycf1*)
30	GTCGGAGAGAGAGGGATTCGAACCCTCGGT	IGS [*trnS (GCU), trnS-(GGA)*]
32	CAAAAGGAGAGAGGGGGATTCGAACCCTCGAT	IGS [*trnS (UGA), trnS (GGA*)]

*IGS = intergenic spacer

Source: The author (2019)

MISA analysis was performed in order to identify the Simple Single Sequence Repeats (SSRs) among the plastid genomes. As a result, we detected a total of 527 SSRs in the eight plastid genomes (Annex 9). The number of SSRs per species ranging from 51 to 75. Among these, there were 31-61 mononucleotide repeats, 8-14 dinucleotide repeats, 1-6 trinucleotide repeats, 5-7 tetranucleotide repeats, 1–5 pentanucleotide repeats, and 1-3 hexanucleotide repeats. The trinucleotide, pentanucleotide, and hexanucleotide repeats were absent in several plastid genomes (Figure 7A). The most SSRs were in the LSC region (78 %), following by SSC (14 %) and IR (8 %) region. These 527 SSRs were mainly located in intergenic spacers (347 SSRs), following by CDS (95 SSRs) and introns (84 SSRs, Figure 7B). We totally deigned 46 primers to 54 microsatellites, which are present in at least four species (Annex 10). Seven of 54 microsatellites were in the *clpP-psbB, matK-trnK, psbB-psbT, psbK-psbI,* and *trnR-atpA* hypervariable regions.

Figure 7 - A) Distribution and frequency of SSR types. B) Distributions of SSR for each region.



Source: The author (2019)

3.5 HYPERVARIABLE REGIONS

Using *B. steyermarkii* as reference, the number of mutations and Indels events were compared among *Bulbophyllum* plastid genomes. A total of 148 sequences (CDS, IGS and introns) flanked by the same exon and longer than 150 bp were extracted. As a result, 4,064 mutations and 624 Indels were identified. The IGS had the highest mutations rate with 5.21 mutations for each 100 bp, following by introns and CDS with 3.56 and 2.58, respectively (Table 4).

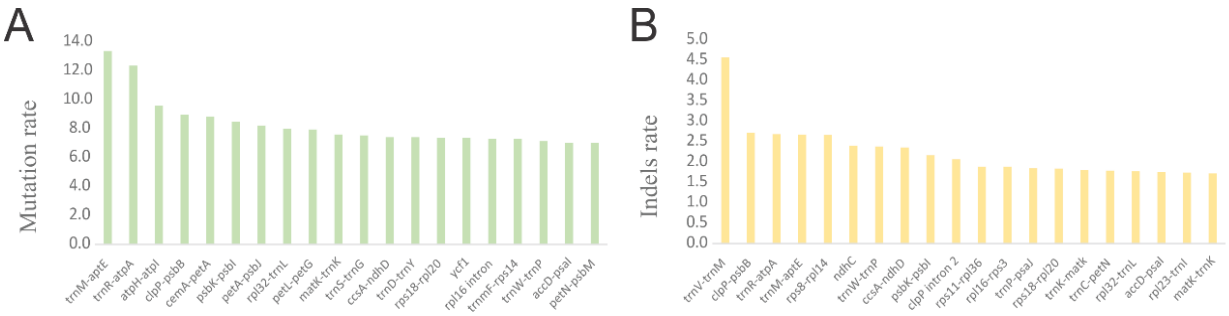
Table 4 - Mutations and Indels events for each region: protein coding genes (CDS), intergenic spacers (IGS), and introns. Rate mutation each 100 pairs bases among all *Bulbophyllum* plastid genomes.

Region	Mutations	Indels	Total length	Mutations / 100 bp	Indels / 100 bp
CDs	1,554	83	60,206	2.58	0.14
IGS	1,927	410	36,968	5.21	1.11
Intron	490	104	13,781	3.56	0.75

Source: The author (2019)

The sequences with the highest mutations rate were *trnM-aptE, trnR-atpA, atpH-atpI, clpP-psbB, cemA-petA, psbK-psbI, petA-psbJ, rpl32-trnL, petL-petG,* and *matK-trnK* (Figure 8A). Also, when the Indels were considered, the IGS presented the highest rate with 1.11 Indels for each 100 pb. The *trnV-trnM, clpP-psbB, trnR-atpA, trnM-aptE, rps8-rpl14, ndhC, trnW-trnP, ccsA-ndhD, psbK-psbI,* and *clpP intron2* sequences had the highest Indels rates (Figure 8B).

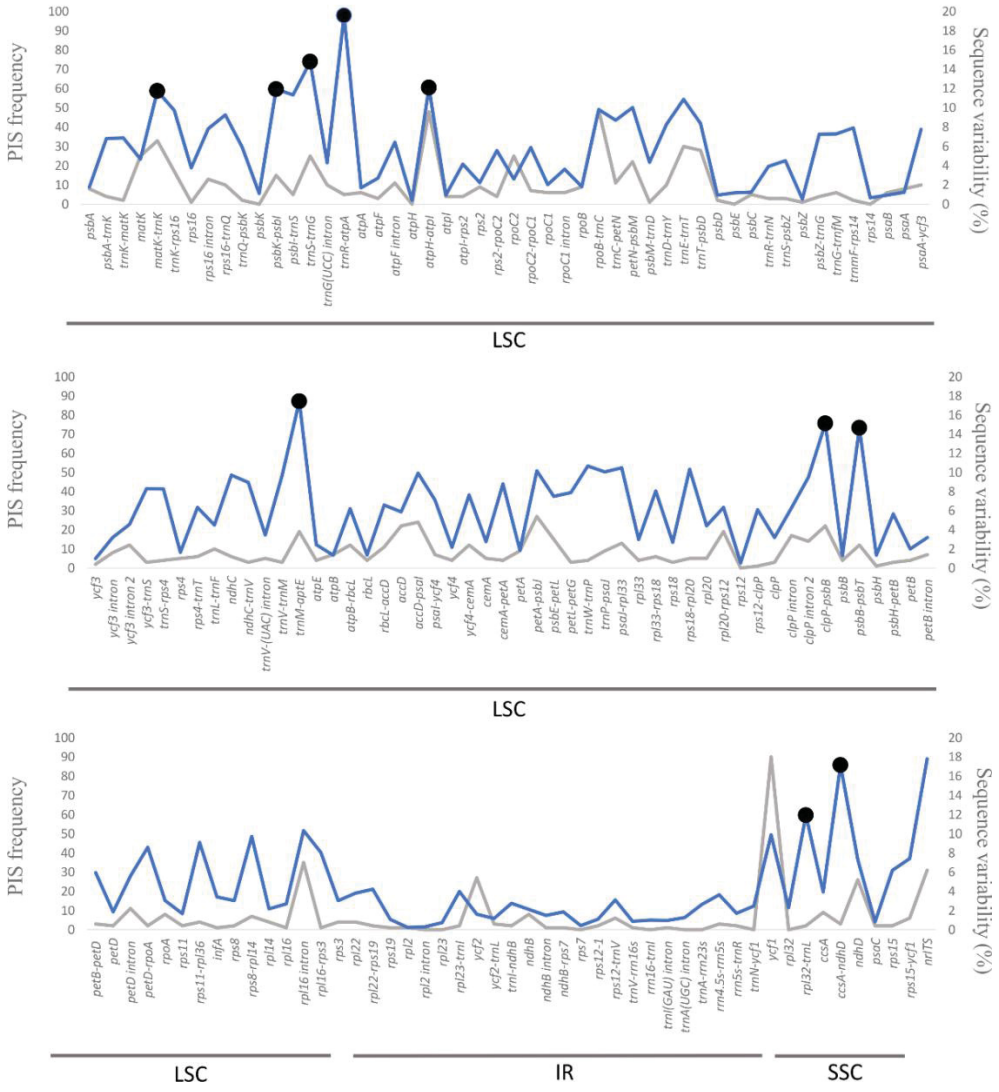
Figure 8 – A) The 20 sequences with the highest mutation and B) Indels rate



Source: The author (2019)

Furthermore, the most PIS were found in the *ycf1, rpoB-trnC, atpH-atpI, rpl16-I, matK-trnK, trnE-trnT, trnT-psbD, petA-psbJ, ycf2,* and *matK* sequences with 90, 49, 48, 35, 31, 30, 28, 27, 27, and 25 sites, respectively (Figure 9). Finally, the sequence variability was calculated for each sequence. As a result, the ten most variable sequences were *trnR-atpA*, *trnM-aptE*, *ccsA-ndhD*, *clpP-psbB*, *trnS-trnG*, *psbB-psbT*, *atpH-atpI*, *psbK-psbI*, *rpl32-trnL*, and *matK-trnK* (Figure 9, Annex 11) and had 15% (184 sites) of total PIS. The correlation analysis showed that the sequence variability was positively correlated with the AT content ($r = 0.81$, $p < 0.001$). Primer pairs were developed for the ten hypervariable regions (Annex 12).

Figure 9 - Sequence variability and parsimony-informative sites (PIS) of 148 plastid sequences and ITS. The blue line represents the SV and the top ten of the hypervariable regions are marked with the black circles. The grey lines indicated the frequency of PIS of each sequence.



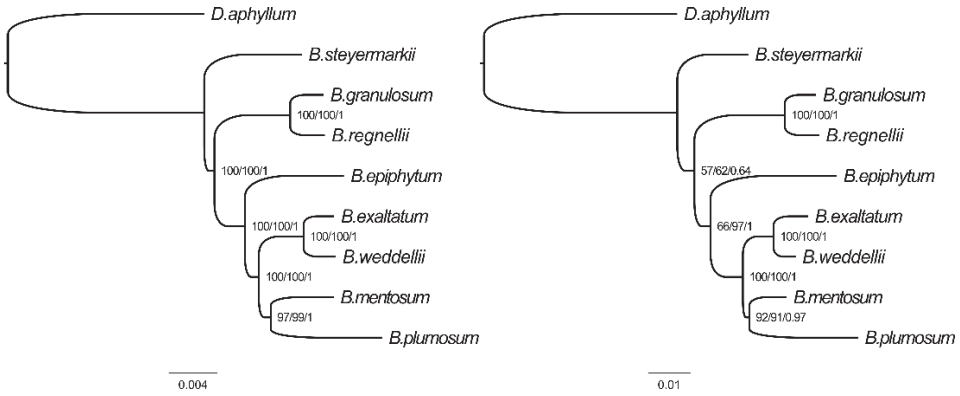
Source: The author (2019)

3.6 PHYLOGENETIC ANALYSIS

Maximum parsimony (MP), Bayesian inference (BI), and Maximum likelihood (ML) analyses were performed using the complete plastid genomes of the eight *Bulbophyllum* species and the ten most variable region (*trnR-atpA*, *trnM-aptE*, *ccsA-ndhD*, *clpP-psbB*, *trnS-trnG*, *psbB-psbT*, *atpH-atpI*, *psbK-psbI*, *rpl32-trnL*, and *matK-trnK)*, with *D. aphyllum* selected as an outgroup. The eight complete chloroplast genomes and the outgroup were aligned in a single data matrix with a total of 163,916 nucleotide sites, among them 2,429 (1.48 %) were identified as parsimony informative sites. The top ten variable regions matrix had 7,087 nucleotide sites with 285 (4%) parsimony-informative characters.

The tree topology generated in the MP, BI, and ML analyses were identical for the two data sets. In *Bulbophyllum*, *B. steyermarkii* was the first species to diverge, followed by *B. granulosum* and *B. regnellii* clade. *B. exaltatum* and *B. weddellii* were clustered into a clade that is close related to the monophyletic group formed by *B. mentosum* and *B. plumosum*. *B. epiphytum* appeared as sister species of the two last clades mentioned. The majority of nodes received high support. Only two nodes in the phylogenetic tree of the top 10 variable regions presented moderate support, which were the relationship between *B. granulosum* + *B. regnellii* clade with the other species (57 – 62), and among *B. epiphytum* with the *B. exaltatum* + *B. weddellii* and *B. mentosum* + *B. plumosum* clades (66; Figure 10).

Figure 10 - Phylogenetic relationships of the eight *Bulbophyllum* species constructed from whole genome (left) and the top 10 variable regions (right) with Maximum parsimony, Maximum likelihood, and Bayesian inference methods. The ML topology and length branches were represented. Numbers associated with branches are MP/ ML/ BI bootstrap support values respectively



Source: The author (2019)

4 DISCUSSION

4.1 PLASTID GENOME FEATURES

In this study, we sequenced the first eight complete plastid genomes of the neotropical *Bulbophyllum* species. All genomes conserved the typical quadripartite structure previously reported in Orchidaceae (ZHITAO et al., 2017), with exception of *Aphyllorchis montana* and *Gastrodia elata*, heterotrophic orchids, who lost one IR copy (FENG et al., 2016; YUAN et al., 2018). Further, the gene arrangement and content were similar within the *Bulbophyllum* genomes and with its sister genus *Dendrobium*, containing 68 CDS, 30 tRNA, and four ribosomal rRNA (ZHITAO et al., 2017).

Even though the general structure of *Bulbophyllum* plastid genomes are conserved, differences in *ndh* genes composition and total length were detected. The *ndh* family is composed of eleven genes which are involved in the respiratory electron transport and chlororespiration (OHYAMA et al., 1986; MARTÍN; SABATER, 2010; UEDA et al., 2012). Ten of the eleven *ndh* genes in *Bulbophyllum* were deleted or truncated by some punctual mutations or deletion that generated premature stop codons. The only gene with full reading frame, i.e. *ndhG,* was observed in *B. exaltatum*. The loss of *ndh* genes has been previously reported in orchids (CHANG et al., 2006; YANG et al., 2013; LUO et al., 2014; KIM et al., 2015; ZHITAO et al., 2017; MAUAD et al., 2019) and is useful in comparative analyses but not in phylogenetics due to high levels of homoplasy depicted by the independent loss of some *ndh* genes in a number of species of Orchidaceae (KIM et al., 2015).

Previous studies conclude that the total length of plastid genome is influenced by contractions and expansions of the IRs in Angiosperms (CHUMLEY et al., 2006; GREEN, 2011; WENG et al., 2017). In contrast, here we found that the length of *Bulbophyllum* plastid genomes is determined by the contraction and expansion of the SSC due to the independent loss of the seven *ndh* genes present in this region. Previous orchids studies found similar results (CHANG et al., 2006; JHENG et al., 2012; KIM et al., 2015; ZHU et al., 2018) suggesting that the gain/loss of *ndh* genes have a more pronounced effect in the SSC length, but only Zhitao et al. (2017) associated the total length of plastid genome with the SSC length and conclude that

total length is medially correlated with the SSC region and contrary with our findings they suggest that the total length of plastid genome depends on the length of LSC.

In general, the gene organization of IRs borders among *Bulbophyllum* orchids were similar. However, some differences in the IRs/SSC junctions were detected. In all *Bulbophyllum* plastid genomes we identified the presence of the *trnH-rps19* cluster within the IRs and the IR/LSC junctions were located upstream of the *rpl22* gene. In agreement with our findings Luo et al. (2014) and Kim et al. (2015) detected the same IRs/LSC pattern among some species of orchids. This type of gene organization was named as type III by Wang et al. (2008) based in the *trnH-rps19* cluster position among 123 Angiosperms. These results suggest that IRs/LSC junction is conserved among orchids.

On the other hand, different IR/SSC patterns have been reported in Orchidacae (YANG et al., 2013; LUO et al., 2014; NIU et al., 2017a; DONG et al., 2018; ZHU et al., 2018). According to Kim et al. (2015) the stability of the IR/SSC junctions are highly affected by the loss and gain of the *ndh* genes, especially of *ndhF* gene, the absence of this gene could create complicated modifications like shifts in the IR border gene composition, reduction of *ycf1* (less than 1kb) and changes in the *ycf1* position within the IRs. When we considered the presence/absent of the *ndhF* pseudogene, three types of IR/SSC junctions were identified. The first was found in *B. epiphytum*, where the *ndhF* pseudogene is present and near the IRb but it is not overlapping the IRb/SSC junction. *B. mentosum* showed the second type with the IRb expanded with 16 and 362 bp towards the *ndhF* and *ycf1*, respectively. The other six genomes present the third type in which the *ndhF* gene was loss, and the IRs expands towards *ycf1* with 33 to 1,436 bp (Figure 2). Due to the *Bulbophyllum* species were not clustered in their sections when were classified according to the borders of IRs, we suggest that the IR borders have not phylogenetic signal.

4.2 CODON USAGE

The codon usage analysis revealed the presence of several alternative start codons and preference for codons ending in A/T among the *Bulbophyllum* plastid genomes. In the same way to other orchids, *Bulbophyllum* plastid genomes showed ATG as a start codon, however three alternative start codon ATT, GTG, and ACG for *matK,*

rps19 and *rpl2* genes were detected. These and other RNA editing sites in *rps12* and *ycf15* genes has been previously reported in Orchidaceae (CHANG et al., 2006; YANG et al., 2013; LUO et al., 2014; BARTHET et al., 2015; SCHELKUNOV et al., 2015) and they are probably involved in the adaption of some functional process (HE et al., 2016). Additionally, a RSCU analysis indicated that the codons ending in A/T are more frequently used than of those ending in G/C. Numerous Angiosperms show this high preference of A/T at third position of codon (MENG et al., 2018; TIAN et al., 2018; RAMAN et al., 2019) and some author suggests that it is a result of compositional bias due to the high AT content in the plastid genomes (MORTON, 1993; ZHOU et al., 2008; NIU et al., 2017b).

4.3 MOLECULAR EVOLUTION ANALYSIS ON PROTEIN-CODING GENES

The protein-coding genes with putative sites under positive selection, in general, coincided with the most divergent genes. Eighty-nine putative sites under positive selection were located and distributed in the *ycf1*, *ycf2, accD,* and *rps3* genes. The plastid *ycf1* gene is the second largest gene in the plastid genome and is essential for photosynthetic protein import (KIKUCHI et al., 2013). Also, it is one of the most variable genes among the land plants therefore it has been used in phylogenetic inference analyses (DONG et al., 2015). Similar with other plastid genomes of Orchidaceae (NIU et al., 2017b; DONG et al., 2018; LI et al., 2019) here we detected 68 sites under positive selection in the *ycf1*.

Additionally, the *ycf2* and *accD* gene showed 10 putative sites under adaptative selection. The function of *ycf2* is still unknown, and the *accD* gene regulates fatty acid synthesis enzyme and is essential to maintain the plastid compartment (KODE et al., 2005). Furthermore, the *rps3* gene showed only one site under positive selection and is related to translation process. These findings could be a result of the adaption process due to some natural selective pressures like temperature, substrate, water, light and could help *Bulbophyllum* orchids in their diversification. However, more studies are needed to understand why these genes are probably under adaptative selection.

4.4 CHARACTERIZATION OF REPEAT SEQUENCES AND SIMPLE SEQUENCE REPEAT (SSRS)

Repeat sequences may play an important role in the genome recombination and rearrangements, size and structure, through illegitimate recombination and slipped-strand mispairing (GEMAYEL et al., 2010). In this study, 249 repeats sequences were located and the majority of those were in the non-coding regions. This finding was similar with Yang et al. (2013) work in which the comparative analyses of eight *Cymbidium* species showed 232 repeats sequences and equally with our results the most repeat sequences were ubicated in the non-coding regions. However, in *Cymbidium* orchids, the reverse repeat sequences were the most abundant and in *Bulbophyllum* plastid genomes were palindromic repeats.

The simple repeat sequences have been extensively used in populations genetics and phylogenetics studies (KARTZINEL et al., 2013; MINASIEWICZ et al., 2018) due to their high mutation rates (GEMAYEL et al., 2010). Here, a total of 54 polymorphic SSRs were identified in the *Bulbophyllum* plastid genomes. The most abundant type of SSR were mononucleotide repeats (A/T). These results were similar with the 47 polymorphic SSRs located among 25 *Dendrobium* species with the presence of 44 mononucleotide SSRs (A/T; ZHITAO et al., 2017). Also, as equal with other plastid genomes studies (ZHANG et al., 2016; NIU et al., 2017a), the majority of SSRs located in *Bulbophyllum* are composed of poly-thymine (polyT) or poly-adenine (polyA) repeats, while the G/C SSRs were rare (Annex 9). Thus, we suggest that these SSRs contribute to the AT richness composition of *Bulbophyllum* plastid genomes. Due to, the SSR have a great potential to be used in genetic diversity studies, here we provided a complete set of microsatellites primers pairs to be used in future *Bulbophyllum* populations genetics studies.

4.5 HYPERVARIABLE REGIONS

A total of 148 sequences longer than 150 bp were extracted and for each sequences the mutation and Indels rates were calculated. As expected, the highest mutation and Indels rates were located in the non-coding regions. These findings are probably the result of high AT content in the non-coding regions (Annex 13) and high AT values may contribute in deleterious replication errors as mutations or deletions (NIU et al.,

2017a). Another possible reason is because non-coding regions present fainter selective pressure than the coding regions.

Based on the number of mutations, Indels events, and conserved sites, we propose ten sequences as potential molecular markers to identify *Bulbophyllum* species and improve the phylogenetic resolution for this genus (Figure 9). In the last years, numerous orchids plastid genomes have been sequenced. As a result, various plastid molecular markers have been proposed for Orchidaceae (YANG et al., 2013; NIU et al., 2017a, 2017c; ZHITAO et al., 2017; DONG et al., 2018; ZHU et al., 2018; LI et al., 2019). For instances, Niu et al. (2017c) suggest that *trnK- rps16*, *trnS-trnG*, and *rps16-trnQ* intergenic spacer could be used for genera within Epidendroideae, and *clpP-psbB* and *rps16- trnQ* for Cypripedioideae. Besides that, Niu et al. (2017a) recommend the following sequences: *ndhA* intron, *matK-trnK, clpP-psbB, rps8-rpl14, trnT-trnL, trnK-matK, clpP* intron*, psbK-trnK, trnS-psbC,* and *ndhF-rpl32* to be used in Apostasioideae. The identification of different molecular markers among the orchid subfamilies suggests that the variable sequences differ among *taxa* and our results corroborated it. Only four of ten variable sequences identified for *Bulbophyllum* are shared with its sister genus, *Dendrobium* (*trnR-atpA, psbB-psbT, rpl32-trnL,* and *clpP-psbB*). Among the other six regions, two sequences (*matK-trnK* and *trnS-trnG*) have been previously reported in Niu et al. (2017c, 2017a) and Zhu et al. (2018) studies and the last four molecular markers (*atpH-atpI, ccsA-ndhD, psbK-psbI,* and *trnM-aptE*) are first reported for Orchidaceae.

The previous molecular markers used in the last *Bulbophyllum* researches have no high sequence variability. For example, in the molecular phylogeny of the Neotropical sections of *Bulbophyllum* (SMIDT et al., 2011) used two plastid markers *psbA-trnH* and *trnS-trnG* and a nrITS marker. Although, the *trnS-trnG* and nrITS had high values in this study, the SV of *psbA-trnH* cluster was only 6.5% variable. Other example is Fischer et al. (2007) study in which they reconstructed the phylogeny of Madagascan *Bulbophyllum* using four plastids intergenic spacer and the nrITS. Among the four markers, one (*trnF–ndhJ*) was not considered in this study because the sequence is not shared among the neotropical *Bulbophyllum* species, due to the independent loss of *ndh* genes. The *trnL–trnF* and *psbA–trnH* intergenic spacer had 4.4 and 6.5 % of SV, respectively, and the last one, *trnE–trnD* is a cluster forming by *trnD-trnY*, *trnY* and *trnY-trnD* sequences and the SV value was 5.9 %. Probably, the use of

molecular markers with low SV is one of the reasons why in both studies, some terminal nodes were not resolved.

In general, the nrITS spacer has been widely used in phylogenetics (FISCHER et al., 2007; SMIDT et al., 2011, 2018), whereas for its high variability and PIS. Here we found one intergenic spacer, namely *trnR-atpA*, with greater variability than nrITS and two sequences (*atpH-atpI, matK-trnK*) showed more PIS than the nrITS, but with fewer variability (Annex 14). These findings show the existence of plastid markers as variable as the nuclear markers that have not yet been used in *Bulbophyllum* studies.

4.6 PHYLOGENETIC ANALYSIS

Aiming to test the evolutionary significance of the ten hypervariable regions, we reconstructed the relationship among the eight *Bulbophyllum* species with the complete plastid genomes and the top ten variable sequences. Both data set showed the same topology and the most branches had stronger bootstrap supports in the three approaches used (MP, BI, and ML). Also, the topology is congruent with the molecular phylogeny of the neotropical *Bulbophyllum* (SMIDT et al., 2011). Hence, we reaffirmed that mentioned above and strongly suggest the use of these sequences in future molecular phylogenetic analyses for *Bulbophyllum* neotropical sections.

Furthermore, in the molecular phylogeny of neotropical sections of *Bulbophyllum,* the *B. mentosum* position was unclear. According to Smidt et al. (2011), with the nrITS analysis, *B. mentosum* belongs to *Bulbophyllum* section *Micranthae*. Nonetheless, with the plastid sequences, this species appears in the *B.* section *Xiphizusa*. Here, using the maternal inherited (plastid genome) *B. mentosum* was located in the *B.* section *Xiphizusa* and the two plastid markers used in Smidt et al. (2011) work (*psbA-trnH* and *trnS-trnG*) agree with the whole plastid genome phylogenetic signal. However, we suggest that the incongruences between nuclear and plastid data find in the previous phylogeny may be a result of ancient hybridizations process (BARBER et al., 2007).

5 CONCLUSION

In this project we sequenced the first eight complete plastomes of *Bulbophyllum* orchids, representing five of the six Neotropical sections. In general, the plastomes were similar in gene content and structure, excepting for *ndh* genes composition. However, despite this general high similarity, we detected several sequences with higher variability than the previous nuclear and plastid molecular markers used in this genus.

We provided great molecular resources for *Bulbophyllum*, which includes 54 microsatellites and molecular markers of the ten top most variable regions with their specific primers. These molecular resources were useful to improve our understanding about the phylogenic relationship among the species and will help species identification in problematic *taxons* like *B. epiphytum x B. rupicolum*, *B. micranthum x B. macroceras*, and the study of population genetics and phylogeography of *Bulbophyllum*.

REFERENCES

AZEVEDO, C. O.; BORBA, E. L.; VAN DEN BERG, C. Evidence of natural hybridization and introgression in *Bulbophyllum involutum* Borba, Semir & F. Barros and *B. weddellii* (Lindl.) Rchb. f. (Orchidaceae) in the Chapada Diamantina, Brazil, by using allozyme markers. **Brazilian Journal of Botany**, 2006. scielo.

BABICKI, S.; ARNDT, D.; MARCU, A.; et al. Heatmapper: web-enabled heat mapping for all. **Nucleic Acids Research**, v. 44, n. 1, p. 147–153, 2016. Disponível em: <https://doi.org/10.1093/nar/gkw419>.

BARBER, J. C.; FINCH, C. C.; FRANCISCO-ORTEGA, J.; SANTOS-GUERRA, A.; JANSEN, R. K. Hybridization in Macaronesian *Sideritis* (Lamiaceae): evidence from incongruence of multiple independent nuclear and chloroplast sequence datasets. **TAXON**, v. 56, n. 1, p. 74–88, 2007. John Wiley & Sons, Ltd. Disponível em: <https://doi.org/10.2307/25065737>.

BARTHET, M. M.; MOUKARZEL, K.; SMITH, K. N.; PATEL, J.; HILU, K. W. Alternative translation initiation codons for the plastid maturase *matK*: Unraveling the pseudogene misconception in the Orchidaceae. **BMC Evolutionary Biology**, v. 15, n. 1, p. 1–15, 2015. BMC Evolutionary Biology. Disponível em: <http://dx.doi.org/10.1186/s12862-015-0491-1>.

BEIER, S.; THIEL, T.; MÜNCH, T.; SCHOLZ, U.; MASCHER, M. MISA-web: a web server for microsatellite prediction. **Bioinformatics**, v. 33, n. 16, p. 2583–2585, 2017. Disponível em: <http://dx.doi.org/10.1093/bioinformatics/btx198>.

BORBA, E. L.; SEMIR, J.; BARROS, F. *Bulbophyllum involutum* Borba, Semir & F. Barros (Orchidaceae), a new species from the Brazilian “campos rupestres”. **Novon**, v. 8, n. 3, p. 225–229, 1998. Disponível em: <http://www.jstor.org/stable/3392005>.

BORBA, E.; SEMIR, J. *Bulbophyllum x cipoense* (Orchidaceae), a new natural hybrid from the Brazilian" campos rupestres": description and biology. **Lindleyana**, v. 13, n. 2, p. 113–120, 1998.

CHANG, C; LIN, H. C.; LIN, I. P.; et al. The chloroplast genome of *Phalaenopsis aphrodite* (Orchidaceae): Comparative analysis of evolutionary rate with that of grasses and its phylogenetic implications. **Molecular Biology and Evolution**, v. 23, n. 2, p. 279–291, 2006.

CHARIF, D.; LOBRY, J. R. SeqinR 1 . 0-2 : A Contributed Package to the R Project for Statistical Computing Devoted to Biological Sequences Retrieval and Analysis. In: V. M. Bastolla U, Porto M, Roman HE (Org.); **Structural approaches to sequence evolution**. Springer B ed., p.1–26, 2007. New York.

CHEN, Y.; XU, J.; YU, H.; et al. Cytotoxic phenolics from *Bulbophyllum odoratissimum*. **Food Chemistry**, v. 107, n. 1, p. 169–173, 2008.

CHUMLEY, T. W.; PALMER, J. D.; MOWER, J. P.; et al. The complete chloroplast genome sequence of *Pelargonium* × *hortorum*: Organization and evolution of the

largest and most highly rearranged chloroplast genome of land plants. **Molecular Biology and Evolution**, v. 23, n. 11, p. 2175–2190, 2006.

DARLING, A.; MAU, B.; BLATTNER, F.; PERNA, N. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. **Genome Res**, v. 14, p. 1394–1403, 2004.

DARRIBA, D.; TABOADA, G.; DOALLO, R.; POSADA, D. jModelTest 2: more models, new heuristics and parallel computing. **Nature Methods**, v. 9, n. 8, p. 772, 2012.

DELANNOY, E.; FUJII, S.; COLAS DES FRANCS-SMALL, C.; BRUNDRETT, M.; SMALL, I. Rampant Gene loss in the underground orchid *Rhizanthella gardneri* highlights evolutionary constraints on plastid genomes. **Molecular Biology and Evolution**, v. 28, n. 7, p. 2077–2086, 2011.

DONG, W. L.; WANG, R. N.; ZHANG, N. Y.; et al. Molecular evolution of chloroplast genomes of orchid species: Insights into phylogenetic relationship and adaptive evolution. **International Journal of Molecular Sciences**, v. 19, n. 3, 2018.

DONG, W.; XU, C.; LI, C.; et al. ycf1, the most promising plastid DNA barcode of land plants. **Scientific Reports**, v. 5, p. 8348, 2015.

DOYLE, J.; DOYLE, J. A rapid DNA isolation procedure for small amounts of leaf tissue. **Phytochem Bull**, v. 19, p. 810–815, 1987.

FENG, Y.-L.; WICKE, S.; LI, J.-W.; et al. Lineage-Specific Reductions of Plastid Genomes in an Orchid Tribe with Partially and Fully Mycoheterotrophic Species. **Genome Biology and Evolution**, v. 8, n. 7, p. 2164–2175, 2016. Disponível em: <https://doi.org/10.1093/gbe/evw144>.

FISCHER, G. A.; GRAVENDEEL, B.; SIEDER, A.; et al. Evolution of resupination in Malagasy species of *Bulbophyllum* (Orchidaceae). **Molecular Phylogenetics and Evolution**, v. 45, n. 1, p. 358–376, 2007.

GEMAYEL, R.; VINCES, M. D.; LEGENDRE, M.; VERSTREPEN, K. J. Variable Tandem Repeats Accelerate Evolution of Coding and Regulatory Sequences. **Annual Review of Genetics**, v. 44, n. 1, p. 445–477, 2010. Annual Reviews. Disponível em: <https://doi.org/10.1146/annurev-genet-072610-155046>.

GRAHAM, S. W.; LAM, V. K. Y.; MERCKX, V. S. F. T. Plastomes on the edge: the evolutionary breakdown of mycoheterotroph plastid genomes. **New Phytologist**, v. 214, n. 1, p. 48–55, 2017. John Wiley & Sons, Ltd (10.1111). Disponível em: <https://doi.org/10.1111/nph.14398>.

GREEN, B. R. Chloroplast genomes of photosynthetic eukaryotes. **Plant Journal**, v. 66, n. 1, p. 34–44, 2011.

HE, P.; HUANG, S.; XIAO, G.; ZHANG, Y.; YU, J. Abundant RNA editing sites of chloroplast protein-coding genes in *Ginkgo biloba* and an evolutionary pattern analysis. **BMC Plant Biology**, v. 16, n. 1, p. 1–12, 2016. BMC Plant Biology. Disponível em: <http://dx.doi.org/10.1186/s12870-016-0944-8>.

INGVARSSON, P. K.; RIBSTEIN, S.; TAYLOR, D. R. Molecular Evolution of Insertions and Deletion in the Chloroplast Genome of Silene. **Molecular Biology and Evolution**, v. 20, n. 11, p. 1737–1740, 2003.

JALEEL, W.; LU, L.; HE, Y. Biology, taxonomy, and IPM strategies of *Bactrocera tau* Walker and complex species (Diptera; Tephritidae) in Asia: a comprehensive review. **Environmental Science and Pollution Research**, v. 25, p. 19346–19361, 2018. Environmental Science and Pollution Research. Disponível em: <http://link.springer.com/10.1007/s11356-018-2306-6>.

JHENG, C. F.; CHEN, TIEN CHIH; LIN, J. Y.; et al. The comparative chloroplast genomic analysis of photosynthetic orchids and developing DNA markers to distinguish *Phalaenopsis* orchids. **Plant Science**, v. 190, p. 62–73, 2012. Elsevier Ireland Ltd. Disponível em: <http://dx.doi.org/10.1016/j.plantsci.2012.04.001>.

KALYAANAMOORTHY, S.; MINH, B. Q.; WONG, T. K. F.; VON HAESELER, A.; JERMIIN, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. **Nature Methods**, v. 14, p. 587, 2017. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved. Disponível em: <https://doi.org/10.1038/nmeth.4285>.

KARTZINEL, T. R.; SHEFFERSON, R. P.; TRAPNELL, D. W. Relative importance of pollen and seed dispersal across a Neotropical mountain landscape for an epiphytic orchid. **Molecular Ecology**, v. 22, n. 24, p. 6048–6059, 2013.

KEARSE, M.; MOIR, R.; WILSON, A.; et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. **Bioinformatics**, v. 28, p. 1647–1649, 2012.

KIKUCHI, S.; BÉDARD, J.; HIRANO, M.; et al. Uncovering the protein translocon at the chloroplast inner envelope membrane. **Science**, v. 339, n. 6119, p. 571–574, 2013.

KIM, H. T.; KIM, J. S.; MOORE, M. J.; et al. Seven new complete plastome sequences reveal rampant independent loss of the ndh gene family across orchids and associated instability of the inverted repeat/small single-copy region boundaries. **PLoS ONE**, v. 10, n. 11, p. 1–18, 2015.

KODE, V.; MUDD, E. A.; IAMTHAM, S.; DAY, A. The tobacco plastid accD gene is essential and is required for leaf development. **Plant Journal**, v. 44, n. 2, p. 237–244, 2005.

KURTZ, S.; CHOUDHURI, J. V; OHLEBUSCH, E.; et al. REPuter: The Manifold Applications of Repeat Analysis on a Genomic Scale. **Nucleic Acids Res.**, 2001. Disponível em: <http://nar.oxfordjournals.org/content/29/22/4633.full.pdf+html>.

LALITHARANI, S.; MOHAN, V. R.; MARUTHUPANDIAN, A. Pharmacognostic investigations on B*ulbophyllum albidum* (Wight) Hook. F. **International Journal of PharmTech Research**, v. 3, n. 1, p. 556–562, 2011.

LI, Z.-H.; MA, X.; WANG, D.-Y.; et al. Evolution of plastid genomes of *Holcoglossum* (Orchidaceae) with recent radiation. **BMC Evolutionary Biology**, v. 19, n. 1, p. 1–10, 2019. BMC Evolutionary Biology.

LIN, C. S.; CHEN, J. J. W.; HUANG, Y. T.; et al. The location and translocation of ndh genes of chloroplast origin in the Orchidaceae family. **Scientific Reports**, v. 5, p. 1–10, 2015.

LOHSE, M.; DRECHSEL, O.; BOCK, R. OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. **Current Genetics**, v. 52, n. 5, p. 267–274, 2007. Disponível em: <https://doi.org/10.1007/s00294-007-0161-y>.

LUO, J.; HOU, B.-W.; NIU, Z.-T.; et al. Comparative Chloroplast Genomes of Photosynthetic Orchids: Insights into Evolution of the Orchidaceae and Development of Molecular Markers for Phylogenetic Applications. **PLOS ONE**, v. 9, n. 6, p. e99016, 2014. Public Library of Science. Disponível em: <https://doi.org/10.1371/journal.pone.0099016>.

MANCINELLI, W.; SMIDT, E. O gênero *Bulbophyllum* (Orchidaceae) na Região Sul do Brasil. **Rodriguésia**, v. 63, n. 4, p. 803–815, 2012.

MARIAC, C.; SCARCELLI, N.; POUZADOU, J.; et al. Cost-effective enrichment hybridization capture of chloroplast genomes at deep multiplexing levels for population genetics and phylogeography studies. **Molecular Ecology Resources**, v. 14, n. 6, p. 1103–1113, 2014.

MARTÍN, M.; SABATER, B. Plastid ndh genes in plant evolution. **Plant Physiology and Biochemistry**, v. 48, n. 8, p. 636–645, 2010. Elsevier Masson SAS. Disponível em: <http://dx.doi.org/10.1016/j.plaphy.2010.04.009>.

MARTINS, W. S.; LUCAS, D. C. S.; NEVES, K. F. DE S.; BERTIOLI, D. J. WebSat--a web software for microsatellite marker development. **Bioinformation**, v. 3, n. 6, p. 282–283, 2009. Biomedical Informatics Publishing Group. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/19255650>.

MENG, J.; LI, X.; LI, H.; et al. Comparative analysis of the complete chloroplast genomes of four *Aconitum* medicinal species. **Molecules**, v. 23, n. 5, p. 2–15, 2018.

MINASIEWICZ, J.; ZNANIECKA, J. M.; GÓRNIAK, M.; KAWIŃSKI, A. Spatial genetic structure of an endangered orchid *Cypripedium calceolus* (Orchidaceae) at a regional scale: limited gene flow in a fragmented landscape. **Conservation Genetics**, v. 19, n. 6, p. 1449–1460, 2018. Springer Netherlands. Disponível em: <http://dx.doi.org/10.1007/s10592-018-1113-4>.

MORTON, B. R. Chloroplast DNA codon use: Evidence for selection at the *psb A* locus based on tRNA availability. **Journal of Molecular Evolution**, v. 37, n. 3, p. 273–280, 1993. Disponível em: <https://doi.org/10.1007/BF00175504>.

NIELSEN, R.; YANG, Z. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. **Genetics**, v. 148, p. 929–936, 1998.

NIU, Z.; PAN, J.; ZHU, S.; et al. Comparative analysis of the complete plastomes of *Apostasia wallichii* and *Neuwiedia singapureana* (Apostasioideae) reveals different evolutionary dynamics of IR/SSC boundary among photosynthetic orchi. **Frontiers in Plant Science**, v. 8, n. October, p. 1–11, 2017.

NIU, Z.; XUE, Q.; WANG, H.; et al. Mutational biases and GC-biased gene conversion affect GC content in the plastomes of dendrobium genus. **International Journal of Molecular Sciences**, v. 18, n. 11, 2017.

NIU, Z.; XUE, Q.; ZHU, S.; et al. The complete plastome sequences of four orchid species: Insights into the evolution of the orchidaceae and the utility of plastomic mutational hotspots. **Frontiers in Plant Science**, v. 8, n. May, p. 1–11, 2017.

NUNES, E. L. P.; MALDONADO, P. E.; SMIDT, E. C.; STÜTZEL, T.; COAN, A. I. Floral micromorphology and anatomy and its systematic application to Neotropical *Bulbophyllum* section Micranthae (Orchidaceae). **Botanical Journal of the Linnean Society**, v. 183, n. 2, p. 294–315, 2017.

NUNES, E. L. P.; SMIDT, E. C.; STÜTZEL, T.; COAN, A. I. What do floral anatomy and micromorphology tell us about Neotropical *Bulbophyllum* section Didactyle (Orchidaceae: Bulbophyllinae)? **Botanical Journal of the Linnean Society**, v. 175, n. 3, p. 438–452, 2014.

NUNES, E. L. P.; SMIDT, E. C.; STÜTZEL, T.; COAN, A. I. K. E. Comparative floral micromorphology and anatomy of species of *Bulbophyllum* section Napelli (Orchidaceae), a Neotropical section widely distributed in forest habitats. **Botanical Journal of the Linnean Society**, v. 177, n. 2, p. 378–394, 2015.

OHYAMA, K.; FUKUZAWA, H.; KOHCHI, T.; et al. Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. **Nature**, v. 322, n. 6079, p. 572–574, 1986. Disponível em: <https://doi.org/10.1038/322572a0>.

PAN, I.-C.; LIAO, D.-C.; WU, F.-H.; et al. Complete Chloroplast Genome Sequence of an Orchid Model Plant Candidate: *Erycina pusilla* Apply in Tropical *Oncidium* Breeding. **PLOS ONE**, v. 7, n. 4, p. e34738, 2012. Public Library of Science. Disponível em: <https://doi.org/10.1371/journal.pone.0034738>.

PRIDGEON AM, CRIBB PJ, CHASE MW, R. F. **Genera Orchidacearum, Epidendroideae (Part 3)**. 6o ed. Oxford: Oxford University Press, 2014.

RAMAN, G.; PARK, S.; LEE, E. M.; PARK, S. J. Evidence of mitochondrial DNA in the chloroplast genome of *Convallaria keiskei* and its subsequent evolution in the Asparagales. **Scientific Reports**, v. 9, n. 1, p. 1–11, 2019.

RIBEIRO, P. L.; BORBA, E. L.; DE CAMARGO SMIDT, E.; et al. Genetic and morphological variation in the *Bulbophyllum exaltatum* (Orchidaceae) complex occurring in the Brazilian “campos rupestres”: Implications for taxonomy and biogeography. **Plant Systematics and Evolution**, v. 270, n. 1–2, p. 109–137, 2008.

RONQUIST, F.; HUELSENBECK, J. MRBAYES 3: Bayesian phylogenetic inference under mixed models., 2003. Bioinformatics.

ROZAS, J.; C SÁNCHEZ-DELBARRIO, J.; MESSEGUER, X.; ROZAS, R. **DnaSP, DNA Polymorphism Analyses by the Coalescent and Other Methods**. 2004.

SAKAGUCHI, S.; UENO, S.; TSUMURA, Y.; et al. Application of a Simplified Method of Chloroplast Enrichment to Small Amounts of Tissue for Chloroplast Genome Sequencing. **Applications in Plant Sciences**, v. 5, n. 5, p. 1700002, 2017. Disponível em: <http://www.bioone.org/doi/10.3732/apps.1700002>.

SCHELKUNOV, M. I.; SHTRATNIKOVA, V. Y.; NURALIEV, M. S.; et al. Exploring the limits for reduction of plastid genomes: A case study of the mycoheterotrophic orchids *Epipogium aphyllum* and *Epipogium roseum*. **Genome Biology and Evolution**, v. 7, n. 4, p. 1179–1191, 2015.

SHAW, J.; SHAFER, H. L.; LEONARD, O. R.; et al. Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: The tortoise and the hare IV. **American Journal of Botany**, v. 101, n. 11, p. 1987–2004, 2014. John Wiley & Sons, Ltd. Disponível em: <https://doi.org/10.3732/ajb.1400398>.

SILVÉRIO RIGHETTO MAUAD, A. V.; DO NASCIMENTO VIEIRA, L.; BOLSON, M.; et al. Complete chloroplast genome of *Anathallis obovata* (Orchidaceae: Pleurothallidinae). **Revista Brasileira de Botanica**, v. 42, n. 2, p. 345–352, 2019. Springer International Publishing. Disponível em: <https://doi.org/10.1007/s40415-019-00524-3>.

SMIDT, E. C.; BORBA, E. L.; GRAVENDEEL, B.; FISCHER, G. A.; BERG, C. VAN DEN. Molecular phylogeny of the Neotropical sections of *Bulbophyllum* (Orchidaceae) using nuclear and plastid spacers. **TAXON**, v. 60, n. 4, p. 1050–1064, 2011.

SMIDT, E. C.; DE BRITO, A. L. V. T.; MARTINS, A. C.; et al. Phylogenetics, biogeography and character evolution in the *Ornithocephalus* clade (Orchidaceae, Oncidiinae). **Botanical Journal of the Linnean Society**, v. 188, n. 4, p. 339–354, 2018.

SMIDT, E. C.; SILVA-PEREIRA, V.; BORBA, E. L.; VAN DEN BERG, C. Richness, distribution and important areas to preserve *Bulbophyllum* in the neotropics. **Lankesteriana**, v. 7, n. 1–2, p. 107–113, 2007.

SMIDT, E. C.; BORBA, E. *Bulbophyllums* in Brazil: collection history and distribution. **Orchids**, , n. October 2015, p. 130–133, 2007.

SMIDT, E. C.; GALLO, L. W.; SCATENA, V. L. Leaf anatomical and molecular studies in *Bulbophyllum* section Micranthae (Orchidaceae) and their implications for systematics. **Revista Brasileira de Botanica**, v. 36, n. 1, p. 75–82, 2013.

STERN, A.; DORON-FAIGENBOIM, A.; EREZ, E.; et al. Selecton 2007: advanced models for detecting positive and purifying selection using a Bayesian inference approach. **Nucleic Acids Research**, v. 35, n. suppl_2, p. W506–W511, 2007. Disponível em: <https://doi.org/10.1093/nar/gkm382>.

SWOFFORD, D. L. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). , 2003. Sunderland, Massachusetts.: Sinauer Associates.

TAN, K. H.; NISHIDA, R. Zingerone in the floral synomone of *Bulbophyllum baileyi* (Orchidaceae) attracts Bactrocera fruit flies during pollination. **Biochemical Systematics and Ecology**, v. 35, n. 6, p. 334–341, 2007.

TAN, K. H.; TAN, L. T.; NISHIDA, R. Floral phenylpropanoid cocktail and architecture of *Bulbophyllum vinaceum* orchid in attracting fruit flies for pollination. **Journal of Chemical Ecology**, v. 32, n. 11, p. 2429–2441, 2006.

TIAN, N.; HAN, L.; CHEN, C.; WANG, Z. The complete chloroplast genome sequence of *Epipremnum aureum* and its comparative analysis among eight Araceae species. **PLoS ONE**, v. 13, n. 3, p. 1–21, 2018.

TRIFINOPOULOS, J.; NGUYEN, L.-T.; VON HAESELER, A.; MINH, B. Q. IQTREE: a fast online phylogenetic tool for maximum likelihood analysis. **Nucleic Acids Research**, v. 44, n. 1, p. 232–235, 2016. Disponível em: <https://doi.org/10.1093/nar/gkw256>.

UEDA, M.; KUNIYOSHI, T.; YAMAMOTO, H.; et al. Composition and physiological function of the chloroplast NADH dehydrogenase-like complex in Marchantia polymorpha. **Plant Journal**, v. 72, n. 4, p. 683–693, 2012.

VAIDYA, G.; LOHMAN, D. J.; MEIER, R. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. **Cladistics**, v. 27, n. 2, p. 171–180, 2011. John Wiley & Sons, Ltd (10.1111). Disponível em: <https://doi.org/10.1111/j.1096-0031.2010.00329.x>.

VERMEULEN, J. J.; PHELPS, J.; THAVIPOKE, P. Notes on *Bulbophyllum* (Dendrobiinae; Epidendroideae; Orchidaceae): two new species and the dilemmas of species discovery via illegal trade. **Phytotaxa**, v. 184, n. 1, p. 12–22, 2014.

WANG, R.-J.; CHENG, C.-L.; CHANG, C.-C.; et al. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. **BMC Evolutionary Biology**, v. 8, n. 1, p. 36, 2008. Disponível em: <https://doi.org/10.1186/1471-2148-8-36>.

WCSP. World Checklist of Selected Plant Families. Facilitated by the Royal Botanic Gardens, Kew. Disponível em: <http://wcsp.science.kew.org/>. Acesso em: 20/7/2018.

WENG, M. L.; RUHLMAN, T. A.; JANSEN, R. K. Expansion of inverted repeat does not decrease substitution rates in *Pelargonium* plastid genomes. **New Phytologist**, v. 214, n. 2, p. 842–851, 2017.

WU, B.; HE, S.; PAN, Y. J. New dihydrodibenzoxepins from *Bulbophyllum kwangtungense*. **Planta Medica**, v. 72, n. 13, p. 1244–1247, 2006.

WYMAN, S.; JANSEN, R.; BOORE, J. Automatic annotation of organellar genomes with DOGMA. **Bioinformatics**, 2004.

YANG, J.-B.; TANG, M.; LI, H.-T.; ZHANG, Z.-R.; LI, D.-Z. Complete chloroplast genome of the genus *Cymbidium*: lights into the species identification, phylogenetic implications and population genetic analyses. **BMC Evolutionary Biology**, v. 13, n. 1, p. 84, 2013. Disponível em: <https://doi.org/10.1186/1471-2148-13-84>.

YUAN, Y.; JIN, X.; LIU, J.; et al. The *Gastrodia elata* genome provides insights into plant adaptation to heterotrophy. **Nature Communications**, v. 9, n. 1, p. 1615, 2018. Springer US. Disponível em: <http://dx.doi.org/10.1038/s41467-018-03423-5>.

ZHANG, Y.; DU, L.; LIU, A.; et al. The Complete Chloroplast Genome Sequences of Five *Epimedium* Species: Lights into Phylogenetic and Taxonomic Analyses. **Frontiers in Plant Science**, v. 7, n. March, p. 1–12, 2016. Disponível em: <http://journal.frontiersin.org/Article/10.3389/fpls.2016.00306/abstract>.

ZHITAO, N.; SHUYING, Z.; JIAJIA, P.; et al. Comparative analysis of *Dendrobium* plastomes and utility of plastomic mutational hotspots. **Scientific Reports**, v. 7, n. 1, p. 1–11, 2017. Springer US. Disponível em: <http://dx.doi.org/10.1038/s41598-017-02252-8>.

ZHOU, M.; LONG, W.; LI, X. Patterns of synonymous codon usage bias in chloroplast genomes of seed plants. **Forestry Studies in China**, v. 10, n. 4, p. 235, 2008. Disponível em: <https://doi.org/10.1007/s11632-008-0047-1>.

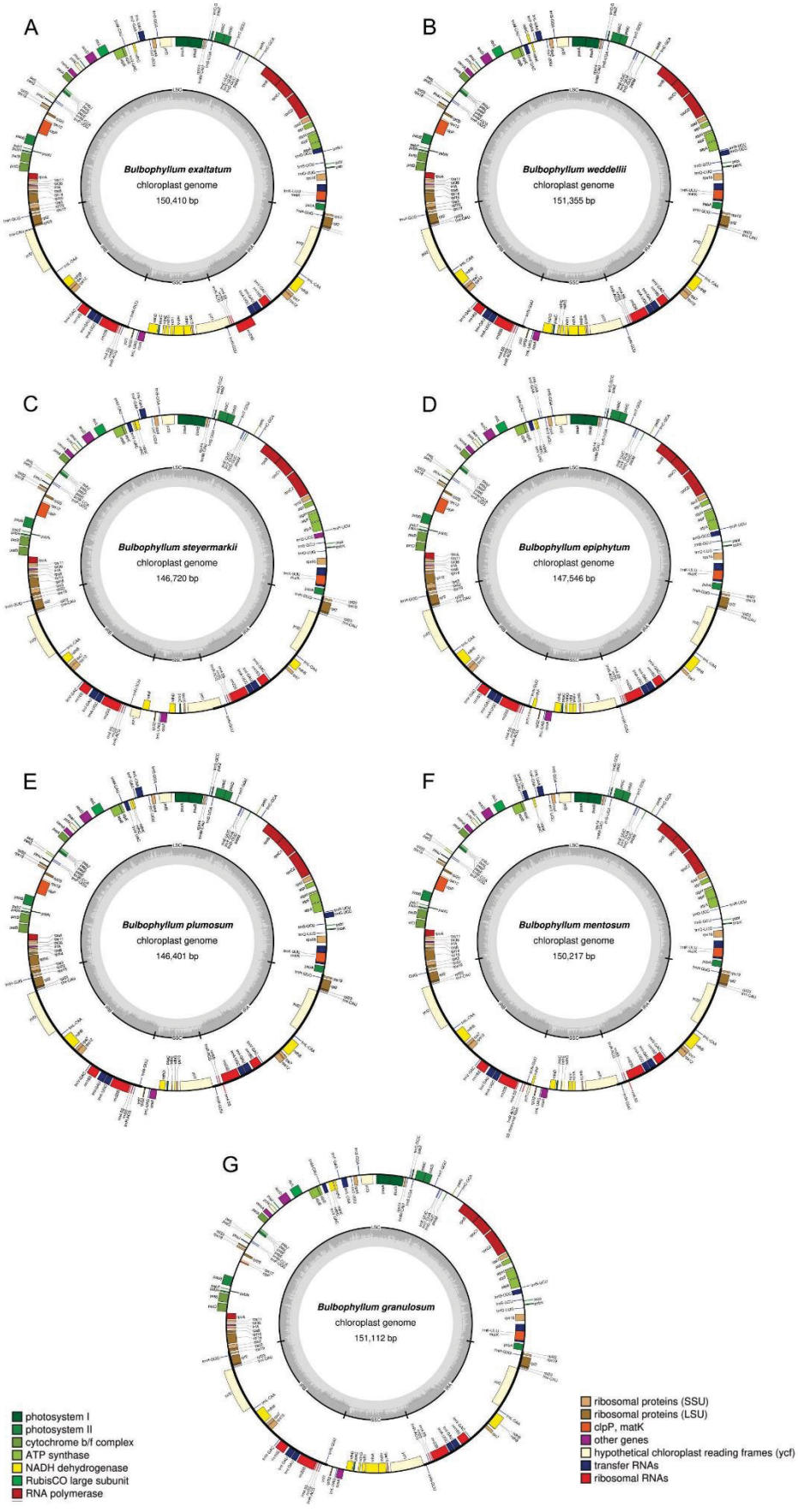
ZHU, S.; NIU, Z.; XUE, Q.; et al. Accurate authentication of *Dendrobium officinale* and its closely related species by comparative analysis of complete plastomes. **Acta Pharmaceutica Sinica B**, v. 8, n. 6, p. 969–980, 2018. Elsevier B.V. Disponível em: <http://dx.doi.org/10.1016/j.apsb.2018.05.009>.

APPENDIX

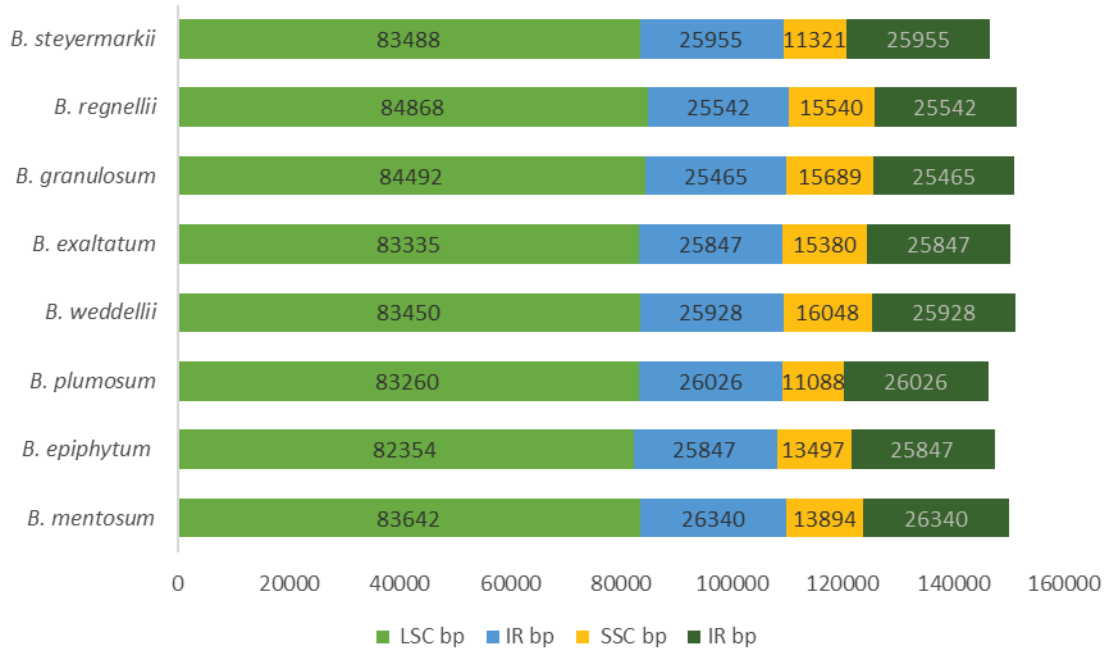
ANNEX 1 – The best fit model chosen for phylogenetic analysis for each set and sequence

Set	Sequences	Bayesian inference (BIC)	Maximum likelihood (AIC)
Whole genome		GTR+I+G	K3Pu+F+R2
Top 10 variable sequences	*matK-trnK*	GTR+G	K3Pu+F+G4
	atpH-atpI	GTR	K3Pu+F
	ccsA-ndhD	F81+G	K3Pu+F+I
	clpP-psbB	F81+G	F81+F+G4
	psbB-psbT	GTR+G	K3Pu+F+I
	psbK-psbI	GTR	K3Pu+F
	rpl32-trnL	HKY	K3Pu+F
	trnM-atpE	F81	F81+F
	trnR-atpA	F81	K3Pu+F
	trnS-trnG	GTR+G	K3Pu+F+G4

ANNEX 2 – Gene organization of *Bulbophyllum* plastid genomes. A) *B. exaltatum*, B) *B. weddellii*, C) *B. steyermarkii*, D) *B. epiphytum*, E) *B. plumosum*, F) *B. mentosum* and G) *B. granulosum*.



ANNEX 3 – Comparison among the eight cp genomes. The figure represents the total length of plastomes and the length of LSC, SSC and IR region.



Annex 4 - A list of genes found in the plastid genomes of eight *Bulbophyllum* species

Function	Gene group	Gene name
Photosynthesis pathways	Photosystem I	*psa* A, B, C, I, J
	Photosystem I assembly	*ycf* 3**,4
	Photosystem II	*psb A, B, C, D, E, F, G, H, I, J, K, L, M, N, T, Z*
	F-type ATP synthase	*atp* A, B, E, F*, H, I
	NAD(P)H-dehydrogenase Complex	*ndh*G (only in *B. exaltatum*)
	Component of cytochrome b6/f Complex	*pet* A, B*, D*, G, L, N
	Cytochrome c biogenesis protein	*ccs*A
	Large subunit of Rubisco	*rbcL*
Structural RNAs	Transfer RNAs	*trn* A-UGC*▪, C-GCA, D-GUC, E-UUC, F-GAA, fM-CAU, G-GCC, G-UCC▪, H-GUG, I-CAU▪, I-GAU*▪, K-UUU*, L-CAA▪, L-UAA*, L-UAG, M-CAU, N-GUU▪, P-UGG, Q-UUG, R-ACG▪, R-UCU, S-GCU, S-GGA, S-UGA, T-GGU, T-UGU, V-GAC▪, V-UAC*, W-CCA, Y-GUA
	Ribosomal RNAs	*rrn 4.5▪, 5▪, 16▪, 23▪*
Transcription and translation related genes	Transcription	*rpo* A, B, C1*, C2
	Ribosomal proteins	*rps* 11, 12**▪+, 14, 15, 16*, 18, 19▪, 2, 3, 4, 7▪, 8; rpl 2*▪. 14, 16*, 20, 22▪, 23▪, 32, 33, 36
	Translation initiation factor	*inf*A
Other genes	RNA processing	*mat*K
	Proteolysis	*clp*P **
	Fatty acid synthesis	*acc*D
	Carbon metabolism	*cem*A
Unknown Proteins	Conserved reading frames	*ycf* 1▪*, 2▪
Pseudogenes***		*ndh*A, B**▪, C, D, E, F, H, I, J, K; one copy of *ycf*1

*, 1 intron in gene; **, 2 introns in gene; ▪, gene repeated in IR region; +, gene has 2 separate transcription units. *** The pseudogenes are not present in all species for more details see Annex 2

Annex 5 - The codon usage frequency of 68 CDS shared among the eight *Bulbophyllum* plastid genomes.

Amino acid	Codon	% of AA								Count							
		B.men	*B. epi*	*B.plu*	*B.wed*	*B.exa*	*B.gra*	*B.reg*	*B.est*	*B.men*	*B. epi*	*B.plu*	*B.wed*	*B.exa*	*B.gra*	*B.reg*	*B.est*
Ala	GCA	0.71%	1.61%	0.66%	0.64%	1.69%	1.66%	1.67%	0.76%	136	309	126	124	321	321	323	144
Ala	GCC	0.45%	0.68%	0.35%	0.35%	0.74%	0.74%	0.74%	0.45%	87	131	67	67	141	143	143	85
Ala	GCG	0.39%	0.55%	0.40%	0.42%	0.51%	0.51%	0.48%	0.40%	74	105	77	81	97	99	93	75
Ala	GCT	0.51%	2.48%	0.38%	0.37%	2.57%	2.53%	2.53%	0.61%	97	475	73	72	489	489	488	114
Arg	AGA	3.32%	1.98%	1.44%	1.48%	1.91%	1.94%	1.96%	3.16%	636	379	276	286	364	374	379	595
Arg	AGG	2.08%	0.65%	1.08%	1.09%	0.66%	0.65%	0.64%	2.06%	398	124	207	211	125	126	123	388
Arg	CGA	1.14%	1.33%	0.63%	0.62%	1.42%	1.39%	1.39%	1.19%	219	254	120	120	270	269	269	225
Arg	CGC	0.48%	0.36%	0.32%	0.31%	0.33%	0.33%	0.34%	0.47%	92	68	62	60	63	64	65	89
Arg	CGG	0.71%	0.43%	0.84%	0.84%	0.39%	0.40%	0.40%	0.73%	137	82	162	162	74	77	77	138
Arg	CGT	0.68%	1.49%	0.36%	0.35%	1.51%	1.50%	1.51%	0.75%	130	285	69	68	288	290	291	141
Asn	AAC	1.73%	1.01%	1.71%	1.72%	0.96%	0.95%	0.95%	1.62%	331	193	329	331	183	184	183	306
Asn	AAT	3.74%	3.92%	2.67%	2.67%	3.79%	3.83%	3.84%	3.77%	718	750	513	514	721	739	740	709
Asp	GAC	0.70%	0.75%	0.89%	0.89%	0.77%	0.77%	0.76%	0.70%	135	144	171	172	146	148	146	132
Asp	GAT	2.29%	3.30%	1.77%	1.75%	3.28%	3.29%	3.30%	2.41%	440	632	339	338	624	635	636	454
Cys	TGC	1.95%	0.34%	0.81%	0.81%	0.26%	0.26%	0.26%	1.92%	374	66	155	156	50	51	50	361
Cys	TGT	2.10%	0.98%	0.83%	0.83%	0.84%	0.84%	0.85%	2.06%	402	187	159	160	159	163	164	388
Gln	CAA	1.69%	2.82%	2.09%	2.09%	2.84%	2.85%	2.85%	1.86%	324	539	402	402	540	551	549	350
Gln	CAG	0.78%	0.83%	1.66%	1.65%	0.84%	0.85%	0.86%	0.77%	149	159	319	319	160	164	165	145
Glu	GAA	2.24%	4.23%	2.36%	2.42%	4.26%	4.25%	4.27%	2.34%	430	809	453	466	810	820	823	440
Glu	GAG	0.91%	1.31%	1.81%	1.79%	1.33%	1.35%	1.36%	0.97%	174	250	347	345	253	261	263	183
Gly	GGA	1.82%	2.67%	1.26%	1.27%	2.70%	2.69%	2.70%	1.83%	349	511	242	244	513	519	521	345
Gly	GGC	0.97%	0.71%	0.49%	0.49%	0.72%	0.70%	0.69%	0.97%	186	136	95	94	137	135	134	183
Gly	GGG	1.53%	1.09%	1.33%	1.32%	1.12%	1.12%	1.11%	1.52%	293	209	255	254	214	216	214	287
Gly	GGT	0.92%	2.29%	0.80%	0.80%	2.28%	2.27%	2.27%	0.97%	176	438	153	155	434	439	438	183

Annex 5. Cont.

Amino acid	Codon	% of AA								Count							
		B.men	*B. epi*	*B.plu*	*B.wed*	*B.exa*	*B.gra*	*B.reg*	*B.est*	*B.men*	*B. epi*	*B.plu*	*B.wed*	*B.exa*	*B.gra*	*B.reg*	*B.est*
His	CAC	0.65%	0.54%	0.82%	0.84%	0.55%	0.53%	0.52%	0.64%	124	104	157	161	104	103	101	121
His	CAT	1.58%	2.11%	1.31%	1.32%	2.04%	2.02%	2.01%	1.68%	303	403	252	255	388	390	388	317
Ile	ATA	1.37%	2.59%	3.41%	3.36%	2.57%	2.60%	2.58%	1.46%	262	495	654	648	490	502	497	274
Ile	ATC	1.73%	1.56%	2.36%	2.36%	1.62%	1.61%	1.61%	1.70%	331	298	454	454	308	311	311	321
Ile	ATT	3.34%	4.17%	2.88%	2.84%	4.13%	4.15%	4.16%	3.44%	640	798	552	548	785	802	803	647
Leu	CTA	0.65%	1.35%	2.21%	2.25%	1.38%	1.38%	1.40%	0.69%	124	259	425	434	263	267	271	130
Leu	CTC	0.49%	0.60%	1.66%	1.65%	0.59%	0.60%	0.60%	0.50%	93	114	318	319	112	115	115	95
Leu	CTG	0.46%	0.64%	1.98%	1.98%	0.68%	0.69%	0.68%	0.45%	88	122	380	381	130	133	131	85
Leu	CTT	1.71%	2.03%	2.43%	2.46%	1.99%	1.99%	1.99%	1.74%	328	388	466	474	379	384	383	327
Leu	TTA	1.71%	3.11%	2.36%	2.25%	3.23%	3.25%	3.25%	1.85%	327	596	454	434	615	627	626	348
Leu	TTG	1.55%	2.16%	3.35%	3.34%	2.14%	2.15%	2.15%	1.68%	298	413	644	643	407	415	415	316
Lys	AAA	4.05%	4.37%	4.30%	4.28%	4.36%	4.43%	4.40%	4.03%	777	837	826	826	829	856	848	759
Lys	AAG	1.69%	1.48%	3.17%	3.25%	1.39%	1.40%	1.40%	1.61%	324	284	609	627	264	270	270	303
Met	ATG	0.98%	2.20%	2.86%	2.86%	2.25%	2.28%	2.29%	1.05%	188	420	549	551	428	441	442	197
Phe	TTC	2.52%	1.98%	2.89%	2.89%	1.98%	1.95%	1.94%	2.39%	483	379	554	558	376	376	375	451
Phe	TTT	4.69%	3.66%	3.34%	3.33%	3.60%	3.59%	3.62%	4.48%	900	701	642	642	685	694	698	844
Pro	CCA	0.98%	1.08%	1.30%	1.29%	1.09%	1.08%	1.07%	0.92%	187	207	249	248	208	208	207	174
Pro	CCC	0.58%	0.85%	0.66%	0.68%	0.91%	0.88%	0.89%	0.64%	112	162	126	132	173	169	171	121
Pro	CCG	0.49%	0.44%	0.92%	0.93%	0.48%	0.46%	0.47%	0.47%	93	84	176	179	91	89	90	88
Pro	CCT	0.74%	1.57%	0.85%	0.83%	1.60%	1.57%	1.58%	0.79%	141	300	164	160	304	304	305	148
Ser	AGC	1.97%	0.42%	0.37%	0.38%	0.39%	0.39%	0.39%	1.91%	378	81	71	74	74	76	75	359
Ser	AGT	1.81%	1.65%	0.60%	0.59%	1.60%	1.58%	1.59%	1.78%	348	315	116	114	304	305	307	335
Ser	TCA	2.63%	1.38%	1.46%	1.49%	1.45%	1.46%	1.46%	2.51%	505	265	281	287	275	281	282	472
Ser	TCC	1.79%	1.23%	1.12%	1.12%	1.18%	1.17%	1.14%	1.75%	344	236	215	216	225	225	219	329
Ser	TCG	1.66%	0.60%	0.95%	0.95%	0.60%	0.61%	0.61%	1.59%	319	115	183	183	114	117	118	299
Ser	TCT	2.25%	2.25%	1.49%	1.51%	2.31%	2.32%	2.30%	2.30%	432	431	287	291	440	447	443	433

Annex 5. Cont.

Amino acid	Codon	% of AA								Count							
		B.men	*B. epi*	*B.plu*	*B.wed*	*B.exa*	*B.gra*	*B.reg*	*B.est*	*B.men*	*B. epi*	*B.plu*	*B.wed*	*B.exa*	*B.gra*	*B.reg*	*B.est*
Stp	TAA	2.45%	0.19%	2.23%	2.22%	0.17%	0.16%	0.17%	2.22%	469	37	428	428	32	31	32	418
Stp	TAG	1.31%	0.13%	2.61%	2.58%	0.11%	0.10%	0.10%	1.22%	252	25	501	498	20	20	20	230
Stp	TGA	2.64%	0.10%	1.89%	1.95%	0.08%	0.09%	0.08%	2.42%	507	20	363	376	16	17	16	455
Thr	ACA	1.77%	1.51%	1.12%	1.13%	1.52%	1.51%	1.50%	1.77%	339	288	215	217	289	291	290	334
Thr	ACC	1.16%	0.90%	0.63%	0.62%	0.92%	0.91%	0.90%	1.16%	223	172	121	119	176	176	174	219
Thr	ACG	1.06%	0.53%	0.64%	0.66%	0.53%	0.54%	0.54%	1.00%	203	102	123	127	100	104	104	188
Thr	ACT	1.24%	2.04%	0.68%	0.68%	2.08%	2.06%	2.07%	1.32%	237	391	131	131	396	398	399	249
Trp	TGG	2.49%	1.78%	1.99%	2.02%	1.77%	1.77%	1.77%	2.38%	477	340	383	390	336	341	341	449
Tyr	TAC	1.77%	0.75%	1.67%	1.66%	0.70%	0.68%	0.72%	1.66%	340	144	320	320	133	132	139	313
Tyr	TAT	3.64%	2.87%	2.30%	2.28%	2.75%	2.79%	2.79%	3.58%	697	550	441	439	523	539	538	675
Val	GTA	0.50%	1.90%	1.81%	1.79%	2.02%	1.98%	1.97%	0.62%	95	364	347	345	385	383	380	116
Val	GTC	0.59%	0.67%	1.28%	1.26%	0.69%	0.71%	0.72%	0.62%	114	128	245	242	132	138	138	116
Val	GTG	0.36%	0.86%	1.39%	1.38%	0.89%	0.89%	0.90%	0.42%	69	164	266	267	169	172	174	79
Val	GTT	1.13%	1.91%	1.78%	1.76%	1.98%	1.97%	1.95%	1.26%	216	366	341	339	376	380	377	237

* B. men: *B. mentosum*, B. epi: B. *epiphytum*, B: plu: B. *plumosum,* B. wed: *B. weddellii*, B. exa: *B. exaltatum*, B. gra: *B. granulosum,* B. reg: *B. regnellii*, B. ste: *B. steyermarkii*

Annex 6 - The Relative Synonymous Codon Usage (RSCU) of 68 CDS shared among the eight *Bulbophyllum* plastid genomes.

Amino acid	Codon	RSCU							
		B.men	***B. epi***	***B.plu***	***B.wed***	***B.exa***	***B.gra***	***B.reg***	***B.est***
Ala	GCA	1.38	1.21	1.47	1.44	1.23	1.22	1.23	1.38
Ala	GCC	0.88	0.51	0.78	0.78	0.54	0.54	0.55	0.81
Ala	GCG	0.75	0.41	0.90	0.94	0.37	0.38	0.36	0.72
Ala	GCT	0.98	1.86	0.85	0.84	1.87	1.86	1.86	1.09
Arg	AGA	2.37	1.91	1.85	1.89	1.84	1.87	1.89	2.27
Arg	AGG	1.48	0.62	1.39	1.40	0.63	0.63	0.61	1.48
Arg	CGA	0.82	1.28	0.80	0.79	1.37	1.35	1.34	0.86
Arg	CGC	0.34	0.34	0.42	0.40	0.32	0.32	0.32	0.34
Arg	CGG	0.51	0.41	1.08	1.07	0.38	0.39	0.38	0.53
Arg	CGT	0.48	1.43	0.46	0.45	1.46	1.45	1.45	0.54
Asn	AAC	0.63	0.41	0.78	0.78	0.40	0.40	0.40	0.60
Asn	AAT	1.37	1.59	1.22	1.22	1.60	1.60	1.60	1.40
Asp	GAC	0.47	0.37	0.67	0.67	0.38	0.38	0.37	0.45
Asp	GAT	1.53	1.63	1.33	1.33	1.62	1.62	1.63	1.55
Cys	TGC	0.96	0.52	0.99	0.99	0.48	0.48	0.47	0.96
Cys	TGT	1.04	1.48	1.01	1.01	1.52	1.52	1.53	1.04
Gln	CAA	1.37	1.54	1.12	1.12	1.54	1.54	1.54	1.41
Gln	CAG	0.63	0.46	0.88	0.88	0.46	0.46	0.46	0.59
Glu	GAA	1.42	1.53	1.13	1.15	1.52	1.52	1.52	1.41
Glu	GAG	0.58	0.47	0.87	0.85	0.48	0.48	0.48	0.59
Gly	GGA	1.39	1.58	1.30	1.31	1.58	1.59	1.59	1.38
Gly	GGC	0.74	0.42	0.51	0.50	0.42	0.41	0.41	0.73
Gly	GGG	1.17	0.65	1.37	1.36	0.66	0.66	0.65	1.15
Gly	GGT	0.70	1.35	0.82	0.83	1.34	1.34	1.34	0.73
His	CAC	0.58	0.41	0.77	0.77	0.42	0.42	0.41	0.55
His	CAT	1.42	1.59	1.23	1.23	1.58	1.58	1.59	1.45
Ile	ATA	0.64	0.93	1.18	1.18	0.93	0.93	0.93	0.66
Ile	ATC	0.81	0.56	0.82	0.83	0.58	0.58	0.58	0.78
Ile	ATT	1.56	1.50	1.00	1.00	1.49	1.49	1.50	1.56
Leu	CTA	0.59	0.82	0.95	0.97	0.83	0.83	0.84	0.60
Leu	CTC	0.44	0.36	0.71	0.71	0.35	0.36	0.36	0.44
Leu	CTG	0.42	0.39	0.85	0.85	0.41	0.41	0.40	0.39
Leu	CTT	1.56	1.23	1.04	1.06	1.19	1.19	1.18	1.51
Leu	TTA	1.56	1.89	1.01	0.97	1.94	1.94	1.94	1.60
Leu	TTG	1.42	1.31	1.44	1.44	1.28	1.28	1.28	1.46

Annex 6 - Cont.

Amino acid	Codon	RSCU							
		B.men	***B. epi***	***B.plu***	***B.wed***	***B.exa***	***B.gra***	***B.reg***	***B.est***
Lys	AAA	1.41	1.49	1.15	1.14	1.52	1.52	1.52	1.43
Lys	AAG	0.59	0.51	0.85	0.86	0.48	0.48	0.48	0.57
Met	ATG	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Phe	TTC	0.70	0.70	0.93	0.93	0.71	0.70	0.70	0.70
Phe	TTT	1.30	1.30	1.07	1.07	1.29	1.30	1.30	1.30
Pro	CCA	1.40	1.10	1.39	1.38	1.07	1.08	1.07	1.31
Pro	CCC	0.84	0.86	0.70	0.73	0.89	0.88	0.88	0.91
Pro	CCG	0.70	0.45	0.98	1.00	0.47	0.46	0.47	0.66
Pro	CCT	1.06	1.59	0.92	0.89	1.57	1.58	1.58	1.11
Ser	AGC	0.98	0.34	0.37	0.38	0.31	0.31	0.31	0.97
Ser	AGT	0.90	1.31	0.60	0.59	1.27	1.26	1.28	0.90
Ser	TCA	1.30	1.10	1.46	1.48	1.15	1.16	1.17	1.27
Ser	TCC	0.89	0.98	1.12	1.11	0.94	0.93	0.91	0.89
Ser	TCG	0.82	0.48	0.95	0.94	0.48	0.48	0.49	0.81
Ser	TCT	1.11	1.79	1.49	1.50	1.84	1.85	1.84	1.17
Stp	TAA	1.15	1.35	0.99	0.99	1.41	1.37	1.41	1.14
Stp	TAG	0.62	0.91	1.16	1.15	0.88	0.88	0.88	0.63
Stp	TGA	1.24	0.73	0.84	0.87	0.71	0.75	0.71	1.24
Thr	ACA	1.35	1.21	1.46	1.46	1.20	1.20	1.20	1.35
Thr	ACC	0.89	0.72	0.82	0.80	0.73	0.73	0.72	0.88
Thr	ACG	0.81	0.43	0.83	0.86	0.42	0.43	0.43	0.76
Thr	ACT	0.95	1.64	0.89	0.88	1.65	1.64	1.65	1.01
Trp	TGG	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Tyr	TAC	0.66	0.41	0.84	0.84	0.41	0.39	0.41	0.63
Tyr	TAT	1.34	1.59	1.16	1.16	1.59	1.61	1.59	1.37
Val	GTA	0.77	1.42	1.16	1.16	1.45	1.43	1.42	0.85
Val	GTC	0.92	0.50	0.82	0.81	0.50	0.51	0.52	0.85
Val	GTG	0.56	0.64	0.89	0.90	0.64	0.64	0.65	0.58
Val	GTT	1.75	1.43	1.14	1.14	1.42	1.42	1.41	1.73

* B. men: *B. mentosum*, B. epi: B. *epiphytum*, B plu: B. *plumosum,* B. wed: *B. weddellii*, B. exa: *B. exaltatum*, B. gra: *B. granulosum,* B. reg: *B. regnellii*, B. ste: *B. steyermarkii*

Annex 7 - Distribution and frequency of the repeat sequences find in the *Bulbophyllum* plastid genomes.

Species	Direction			Total general
	Forward	Palindromic	Reverse	
B. mentosum	6	17	2	25
B. epiphytum	12	20	1	33
B. plumosum	10	20	1	31
B. weddellii	10	18	2	30
B. exaltatum	9	14	4	27
B. granulosum	11	20	3	34
B. regnellii	11	21	1	33
B. steyermarkii	9	23	4	36
Total	78	153	18	249

Annex 8 -. Frequency of repeat sequence length for each *Bulbophyllum* specie.

Length	*B. men*	*B. epi*	*B. plu*	*B. wed*	*B. exa*	*B. gra*	*B. reg*	*B. ste*	Total
30	12	12	14	10	9	10	13	12	92
31	1	1	1	2	2	2	1	3	13
32	2	10	3	4	3	8	3	2	35
33	1	0	0	2	1	0	4	2	10
34	1	1	3	2	2	1	2	0	12
35	0	0	0	0	0	0	1	1	2
36	1	2	2	2	2	1	1	2	13
37	1	1	2	1	1	2	1	3	12
38	0	1	0	0	0	1	0	0	2
39	3	2	3	5	4	3	3	3	26
40	1	0	0	0	0	1	0	0	2
41	0	0	0	0	0	1	0	0	1
43	0	0	0	0	0	1	0	0	1
44	0	0	0	0	0	0	0	4	4
46	1	1	1	1	1	1	1	1	8
47	0	1	1	0	2	1	1	1	7
48	0	0	0	0	0	0	0	1	1
50	0	1	0	0	0	0	0	0	1
52	0	0	0	0	0	0	1	1	2
53	1	0	0	0	0	0	0	0	1
54	0	0	0	0	0	1	0	0	1
59	0	0	1	1	0	0	1	0	3

* B. men: *B. mentosum*, B. epi: B. *epiphytum*, B: plu: B. *plumosum,* B. wed: *B. weddellii*, B. exa: *B. exaltatum*, B. gra: *B. granulosum,* B. reg: *B. regnellii*, B. ste: *B. steyermarkii.*

Annex 9 - Distribution and frequency of the simple Single Sequence Repeats (SSRs) among the *Bulbophyllum* genomes.

Specie	Total	Unit size						Total per region			Total per region			A/T SSRs	% A/T SSRs
		Mono	Di	Tri	Tetra	Penta	Hexa	CDS	IGS	INTRON	IR	LSC	SSC		
B. mentosum	80	57	10	3	5	5	0	14	53	13	6	63	11	74	92.50
B. epiphytum	80	61	9	0	6	1	3	15	54	11	8	60	12	73	91.25
B. plumosum	72	47	14	0	7	3	1	14	45	13	6	58	8	67	93.06
B. weddellii	61	48	8	0	5	0	0	10	42	9	4	49	8	57	93.44
B. exaltatum	56	42	8	0	5	1	0	9	38	9	4	46	6	52	92.86
B. granulosum	59	37	13	2	6	1	0	13	35	11	4	44	11	51	86.44
B. regnellii	56	31	11	6	7	1	0	13	36	7	4	40	12	46	82.14
B. steyermarkii	63	42	11	1	6	2	1	11	41	11	7	51	5	54	85.71

*CDS: gene encoding proteins, IGS: intergenic spacer, LSC: larger single copy region, SSC: small single copy region, IR: invert region.

Annex 10 - Distribution of the polymorphic SSR which are presented in at least four species with their specific primer pairs.

Primer name	Position	Region	Location	SSR Type	Primer sequence (5′-3′)	Tm	Length
Bulbo ssr1	*psbA-matK*	LSC	IGS	(A) 11, (T)17	CCATAGGAATRACCAAAACT	54.1	240
					TATACATAGGGAAAGTCGTGTG	54.5	
Bulbo ssr2	*trnK-matK*	LSC	IGS	(A) 13	GCTTGCARTTTTCATTGCAC	59.5	267
					CGGATTACATAGAGAACGTATTTGG	60.2	
Bulbo ssr3	*matK-trnK*	LSC	IGS	(AAAG) 4	GAGACAGAGAGCCCAATCTA	54.4	232
					KAGACRGATGTGTAGAAGAAAT	5.1	
Bulbo ssr4	*matK-trnK*	LSC	IGS	(A) 12	TCTTCTACACATCYGTCTMCAA	59.7	181
					GCAACGAGCTTCCTTTCTTAAT	59.1	
Bulbo ssr5	*rps16*	LSC	INTRON	(AT) 5, (TA) 6	TAGATGGCTCATTGGGATA	54	226
					ACTTGAGTTAGGAGTACGAATG	53	

Annex 10. Cont.

Primer name	Position	Region	Location	SSR Type	Primer sequence (5′-3′)	Tm	Length
Bulbo ssr6	*rps16*	LSC	INTRON	(T) 10	AGGCAGCAACATACCTTT GTKACAATCAATACGTTAGACC	53.2 53.9	218
Bulbo ssr7	*rps16-trnQ*	LSC	IGS	(A) 14	GCTTTCTACCACATCGTTT GKATCTYTTATGATGGACAA	52.8 53.4	299
Bulbo ssr8	*rps16-trnQ*	LSC	IGS	(A) 11, (A) 12	GGAATTGTCCATCATAARAG GTCCCAGATCGTTTGYATM	52 52.7	180
Bulbo ssr9	*psbK-psbI*	LSC	IGS	(A) 12	GTTATGCCGRTTATACCTTTAC ATCGAGGGTTYCTAATGTCYA	54.8 55.4	194
Bulbo ssr10	*psbK-psbI*	LSC	IGS	(A) 12	CWTTTGAATAARGGAAGGG CACCTATTTTACGACACACACT	54.7 54.5	295
Bulbo ssr11	*psbK-psbI*	LSC	IGS	(A) 10	GTGTGTGTCGTAAAATAGGTG CGAAGATGAAGAGAGAAACA	53.2 53.1	129
Bulbo ssr12	*trnG-trnR*	LSC	IGS	(T) 12	CGACTATAACCCCTAGCCTTCC GACGCCCTTCATTCCTATTTTC	60.3 61.1	183
Bulbo ssr13	*trnR-atpA*	LSC	IGS	(A)15	TAATGGATAGGACAKAGGTCTT AAAYCCTTTTGAGAGAAGC	54.6 53.7	299
Bulbo ssr14	*atpF*	LSC	INTRON	(A) 15	GACTCTTCAGACCAGACAAAA CATTAACTGGAAAGATGGGT	54.5 54.2	299
Bulbo ssr15	*atpI-rps2*	LSC	IGS	(A) 11	CGGATACCGAKAAATCACA GGCCAACGATGATGCTAT	57 57	275
Bulbo ssr16	*rps2-rpoC2*	LSC	IGS	(T) 12	ATGACCAAAATGAACTCCTGCT TCCATGAGACATCAGACCAATC	60 59.9	317
Bulbo ssr17	*rpoC2*	LSC	CDS	(T) 11	AGATYCTGAGAGAAAAGTGTTG AAGAAGTCGAGTAGGTGGATTA	52.3 54.9	300
Bulbo ssr18	*rpoB-trnC*	LSC	IGS	(AAGT)3, (AAAT) 4	CGGGCTCRATATCTTATCTACGTAT ATCTCAAAACACGCRTCGTT	58 56.4	380

Annex 10 - Cont.

Primer name	Position	Region	Location	SSR Type	Primer sequence (5′-3′)	Tm	Length
Bulbo ssr19	*rpoB-trnC*	LSC	IGS	(ATTT) 3	AACGAYGCGTGTTTTGAGAT GGATTCGTSAAGTTCGATCA	58.2 58.6	400
Bulbo ssr20	*petN-psbM*	LSC	IGS	(TATAT) 4	CTCTTTCACTTGTAGTATGGG CCAACTGCTTKAATTCCTGA	53.4 53.5	376
Bulbo ssr21	*trnS-psbZ*	LSC	IGS	(A) 10	ACTCAGCCATCTCTCCCAAATA CCAACCATCAGAAGAAGCAAAT	60.1 60.5	314
Bulbo ssr22	*psaA-ycf3*	LSC	IGS	(T) 14	CRAAGTGCCCATAGACTTTAAT AGGCGTTTTGAATAAGACCA	57.5 57.3	261
Bulbo ssr23	*ycf3-2*	LSC	INTRON	(T) 13	TTCACCAAGCGTAGGTTTCTTT AATTCCGACAACTTCAGGAGAA	60.2 60.1	170
Bulbo ssr24	*trnS-rps4*	LSC	IGS	(A) 11	CTCTCTCCGTTTMTTTTMAT ATGAGTTGTTGGTYGTAGAA	53.7 53.1	278
Bulbo ssr25	*atpB-rbcL*	LSC	IGS	(T) 12	CCACYCAATCGAATCCAA AGAAAGAATAGAGGACCACCCC	60.3 59.8	232
Bulbo ssr26	*accD-psaI*	LSC	IGS	(A) 12	TTMARCTACATGGTTTCCTTCC TCRATAKAGGGATSCTTATTCC	60.5 59.6	297
Bulbo ssr27	*petA-psbJ*	LSC	IGS	(A) 11	TTTGTGGGATGTCTGAAACT AGACGATTAPGAAGGCTAAKAA	55.5 56.5	133
Bulbo ssr28	*petA-psbJ*	LSC	IGS	(A) 12	ATTGTGTCGTCTTGTATCGAAAG TKAAACATCCTGCAAATMCT	59 58.2	300
Bulbo ssr29	*trnP-psaJ*	LSC	IGS	(T) 13	ATGAATCTCTCGTAACTKBCY TCTGMTGGGTCATTATACAT	53.4 53.1	269
Bulbo ssr30	*psaJ-rpl33*	LSC	IGS	(A) 13	GTAACACGGCTCCAAWTTTGC RATTTMCCCTAKAATGGATCCACA	58.7 58.4	143
Bulbo ssr31	*psaJ-rpl33*	LSC	IGS	(A) 14, (TAT) 4	AAATGTGGATCCATTMTAGGGK GCCATGAACCTCCTTTTCTTTT	57.5 57.4	396

Annex 10 - Cont.

Primer name	Position	Region	Location	SSR Type	Primer sequence (5′-3′)	Tm	Length
Bulbo ssr32	*clpP1*	LSC	INTRON	(A) 12, (T) 10	GTGACGCTKAAATTGACTCTTG TACCAAACGTCTAGCATTCCCT	59.2 60	262
Bulbo ssr33	*clpP2*	LSC	INTRON	(T) 13	TCTCGATGAAGTCGGTTGATTA GCCCATTMAGGAACAAGAAA	59.7 60.1	286
Bulbo ssr34	*clpP2*	LSC	INTRON	(T) 12	TTAGGTTCTTGTTCTACTCCG CATCGAATATTTTGGGGAAAGG	54.7 55.5	127
Bulbo ssr35	*clpP-psbB*	LSC	IGS	(TA) 8, (T) 10, (TAT) 4	CAAGTCGCACTATACGTCAACC TCCTCTATGAACCTTCCAGTCC	59.7 59.6	379
Bulbo ssr36	*psbB-psbT*	LSC	IGS	(T) 13	TTTGYCATCTTTTGCCTCTC CACACCTATTYGTTTTGGATCA	60.3 59.9	496
Bulbo ssr37	*petB*	LSC	INTRON	(A) 13	CATAGGGTCTCAACAAGAGAATCC TCAAGCCGAAAMACACAAATAC	60.4 60	259
Bulbo ssr38	*petD*	LSC	INTRON	(T) 14	GGCTCCGTAAGATCCCTAGAAT CCCKGTTCTTCCTTAGATCCCT	60 60	236
Bulbo ssr39	*rpoA*	LSC	CDS	(T) 10	GCAAAATGTTTCTGTARAGTGYC ATGCAGAGGAAGAGGACATGAA	60.2 61.2	295
Bulbo ssr40	*rps8-rpl14*	LSC	IGS	(T) 12	ATGTCCYTACCCATGACGAAC AGTTTCATTAGCACCCGAAGTA	58.8 58	265
Bulbo ssr41	*rpl16*	LSC	INTRON	(T) 11	CAATGGAGCTCTCAACAAAATMTG AGGCAGTGTTATAAAGCATCAACA	58.6 59	374
Bulbo ssr42	*rpl16*	LSC	INTRON	(T) 11	TCTCTMTTTCTATCATCYTTCC TCAGTGTGTGACTCGTTAGTTT	54.4 55.1	286
Bulbo ssr43	*rps12-trnV*	IR	IGS	(T) 14	TCTGGGCTCTTCTATCTTCTAC CTACAGGATCAACAAACCTATG	54.7 54.1	161
Bulbo ssr44	*rps12-trnV*	IR	IGS	(TATTA) 3	GAGATCCTTTCGATGACCTAT AGAGACAAAATGTAGGACTGGT	54.9 54.6	335

Annex 10 - Cont.

Primer name	Position	Region	Location	SSR Type	Primer sequence (5′-3′)	Tm	Length
Bulbo ssr45	*ndhD*	SSC	CDS	(T) 11	CAAGATCAAGGYACAAATTC	55.8	380
					TTATCTHYCTGGTTGGTTATAG	55.1	
Bulbo ssr46	*ycf1*	SSC	CDS	(ATTTTC) 3	MTWCMATTYCTKAAAATCCY	55.3	282
					ATGRGCAGAGAAMAAGAAT	54.7	

*CDS: gene encoding proteins, IGS: intergenic spacer, LSC: larger single copy region, SSC: small single copy region, IR: invert region, TM: Primer Melting Temperature.

Annex 11 - The sequences variability (SV), AT and GC content, and length of 148 sequences (CDS, IGS and introns) that flanked by the same exon and longer than 150 pairs. The number of mutations and Indels events were counted using a *B. steyermarkii* cp genome as reference.

Sequence	SV	AT %	GC %	Length	*Sequence*	SV	AT %	GC %	Length
trnR-atpA	19.58	81.3	18.7	186	*rpl22-rps19*	4.23	76	24	263
trnM-aptE	17.44	73.8	26.2	187	*atpI-rps2*	4.17	78.4	21.6	304
ccsA-ndhD	17.16	77.2	22.8	297	*rpl23-trnI*	3.97	66.5	33.5	172
clpP-psbB	15.14	80.5	19.5	847	*trnR-trnN*	3.93	58.7	41.3	563
trnS-trnG	14.79	81.4	18.6	1075	*ccsA*	3.93	62.5	37.5	978
psbB-psbT	14.67	86.5	13.5	692	*rpl22*	3.81	68.4	31.6	375
atpH-atpI	12.12	72.4	27.6	782	*rps16*	3.78	65.8	34.2	288
psbK-psbI	11.95	76.1	23.9	507	*rrn4.5s-rrn5s*	3.65	55.8	44.2	221
rpl32-trnL	11.93	77.8	22.2	676	*rpoC1 intron*	3.64	63.9	36.1	771
matK-trnK	11.76	77.9	22.1	1277	*trnV (UAC) intron*	3.46	62.3	37.7	583
psbI-trnS	11.35	80.4	19.6	203	*infA*	3.42	63.5	36.5	234
trnE-trnT	10.89	80.5	19.5	1140	*clpP*	3.19	59.1	40.9	615
trnW-trnP	10.67	69.3	30.7	168	*ycf3 intron*	3.19	67.5	32.5	756
psaJ-rpl33	10.48	79.5	20.5	704	*petB intron*	3.18	66.5	33.5	728
rps18-rpl20	10.33	72	28	272	*rps12-trnV*	3.09	63.7	36.3	2058
rpl16 intron	10.32	75.5	24.5	1380	*rpoA*	3.05	65.2	34.8	1029
petA-psbJ	10.16	72	28	1011	*rps3*	3.03	66.4	33.6	657
trnP-psaJ	10.06	76.7	23.3	378	*rps8*	3.02	67.9	32.1	396
petN-psbM	10.04	79.2	20.8	883	*rpl33*	2.99	65	35	201
accD-psaI	9.93	75	25	911	*cemA*	2.75	66.5	33.5	690
ycf1	9.89	72.4	27.6	5484	*trnI-ndhB*	2.75	62.7	37.3	550
rpoB-trnC	9.84	73.6	26.4	1553	*atpF*	2.70	63.9	36.1	555
trnK-rps16	9.73	76.6	23.4	603	*rpl16*	2.70	56.7	43.3	408
ndhC	9.73	63.6	36.4	375	*rps18*	2.68	64	36	297
trnV-trnM	9.71	68.7	31.3	175	*trnA-rrn23s*	2.63	58.7	41.3	157
rps8-rpl14	9.71	81.1	18.9	225	*rpoC2*	2.62	63.7	36.3	4170
clpP intron 2	9.50	75.5	24.5	1111	*trnN-ycf1*	2.47	62.5	37.5	329
rps16-trnQ	9.27	86.3	13.7	1079	*atpE*	2.44	59.8	40.2	411
rps11-rpl36	9.09	76.7	23.3	159	*rpl32*	2.30	69.3	30.7	174
ndhC-trnV	8.96	74.1	25.9	323	*rps2*	2.25	62.5	37.5	711
cemA-petA	8.81	71.8	28.2	226	*rpl14*	2.17	60.9	39.1	369
trnC-petN	8.72	69.6	30.4	838	*ycf4*	2.16	60.3	39.7	555
petD-rpoA	8.59	70.4	29.6	175	*ndhB*	2.11	63.5	36.5	1542
trnT-psbD	8.38	69.4	30.6	947	*rpoC1*	2.04	61.4	38.6	2067
trnD-trnY	8.30	70.8	29.2	270	*petB*	2.00	60.6	39.4	648
ycf3-trnS	8.29	68.7	31.3	216	*petA*	1.87	61.7	38.3	963
trnS-rps4	8.26	73.3	26.7	323	*petD*	1.86	62.2	37.8	492

Annex 11. Cont.

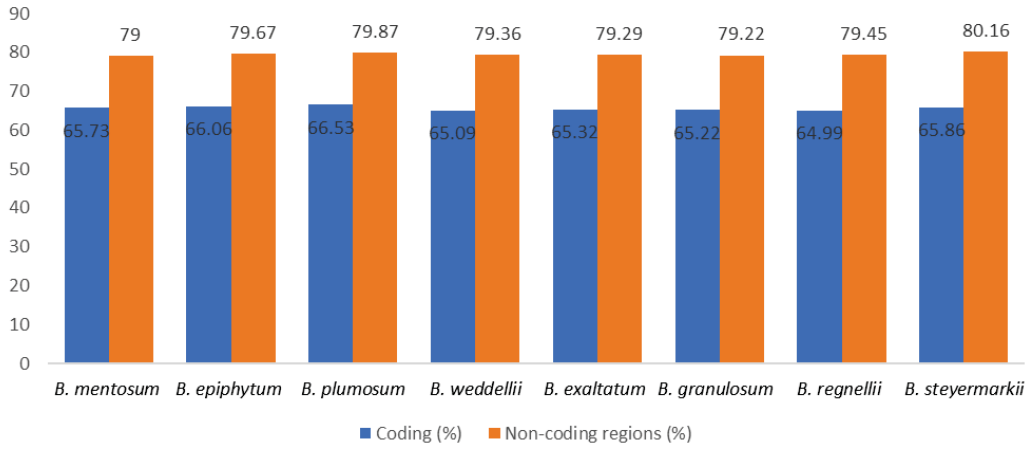
Sequence	SV	AT %	GC %	Length	*Sequence*	SV	AT %	GC %	Length
rpl33-rps18	8.06	73	27	189	*ndhB-rps7*	1.86	64.9	35.1	323
rpl16-rps3	8.05	83.3	16.7	212	*rpoB*	1.84	62.1	37.9	3213
trnmF-rps14	7.93	64.3	35.7	164	*psbA*	1.79	58.3	41.7	1062
petL-petG	7.89	70.4	29.6	189	*rrn5s-trnR*	1.71	58.6	41.4	234
rps16 intron	7.83	70.5	29.5	912	*atpA*	1.71	60.6	39.4	1524
psaA-ycf3	7.78	73.3	26.7	591	*rps11*	1.68	57.1	42.9	417
ycf4-cemA	7.64	74.8	25.2	750	*rps4*	1.65	61.3	38.7	606
psbE-petL	7.50	74.6	25.4	1222	*ycf2*	1.62	62.2	37.8	6909
rps15-ycf1	7.41	78.1	21.9	467	*ndhB intron*	1.48	61.4	38.6	699
trnG-trnfM	7.30	68.1	31.9	177	*rbcL*	1.37	56.4	43.6	1470
ndhD	7.29	73	27	1577	*psbH*	1.35	62.7	37.3	222
psbZ-trnG	7.28	70.4	29.6	226	*atpB*	1.34	57.2	42.8	1497
psaI-ycf4	7.11	70.3	29.7	434	*psbB*	1.31	56.2	43.8	1527
trnK-matK	6.88	74.1	25.9	278	*psbC*	1.27	55.9	44.1	1422
psbA-trnK	6.81	76.7	23.3	277	*trnA (UGC) intron*	1.25	50.1	49.9	815
rbcL-accD	6.58	70.5	29.5	730	*psaA*	1.24	57.1	42.9	2253
atpF intron	6.45	70.7	29.3	937	*psbE*	1.19	61.2	38.8	252
rps4-trnT	6.34	72.8	27.2	342	*ycf2-trnL*	1.16	66.8	33.2	977
rpl20-rps12	6.34	67.2	32.8	766	*psbK*	1.11	65.7	34.3	180
clpP intron	6.28	70.1	29.9	697	*rps12 intron*	1.10	60.6	39.4	547
atpB-rbcL	6.20	80.1	19.9	1133	*rps19*	1.08	60.8	39.2	279
rps15	6.18	69.2	30.8	273	*rrn16-trnI*	1.00	52.1	47.9	309
rps12-clpP	6.11	70.6	29.4	154	*ycf3*	0.99	62.7	37.3	507
petB-petD	5.95	71.8	28.2	185	*trnI (GAU) intron*	0.96	51	49	952
trnQ-psbK	5.92	75.2	24.8	379	*psaB*	0.95	58.9	41.1	2205
rpoC2-rpoC1	5.88	63.5	36.5	187	*psbD*	0.94	58.2	41.8	1062
accD	5.87	68	32	1482	*atpI*	0.94	63.4	36.6	744
psbH-petB	5.65	73.6	26.4	184	*trnV-rrn16s*	0.87	53.1	46.9	229
petD intron	5.58	67.2	32.8	812	*psaC*	0.81	59.1	40.9	246
rps2-rpoC2	5.58	74.3	25.7	268	*rpl23*	0.74	61.9	38.1	270
matK	4.68	68.6	31.4	1536	*rps14*	0.66	59.5	40.5	303
ycf3 intron 2	4.58	65.4	34.6	734	*rps12*	0.54	57.9	42.1	372
trnS-psbZ	4.52	69.5	30.5	215	*psbZ*	0.53	66.3	33.7	189
trnL-trnF	4.49	72.9	27.1	254	*rps7*	0.43	58.8	41.2	468
rpl20	4.42	64.7	35.3	387	*atpH*	0.41	56.6	43.4	246
psbM-trnD	4.35	76.2	23.8	1113	*rpl2 intron*	0.30	59.3	40.7	664
trnG (UCC) intron	4.30	68.1	31.9	683	*rpl2*	0.24	55.6	44.4	822

Annex 12 - Specific primer pairs for the top ten hypervariable regions find in neotropical *Bulbophyllum*.

Nro.	Hypervariable sequence	Region	Location	Primer sequence (5′-3′)	Length	Tm	Product
1	*trnR-atpA*	LSC	IGS	CCTTTGGTATAGGTTCAAATCC	22	56.7	261
				CCAAGATATTTACCGAAGAAGC	22	57.1	
2	*trnM-atpE*	LSC	IGS	AGAGTATTGCTTTCATACGG	20	52.2	226
				GCTGTCAAAGTKATTTCTTC	20	52.5	
3	*ccsA-ndhD*	SSC	IGS	GGGATCAATCTATTAGGAATCG	22	56.7	984
				TTATCTHYCTGGTTGGTTATAG	22	55.1	
4	*clpP-psbB*	LSC	IGS	ACAAGTCGCACTATACGTCAA	21	56	870
				CAACAGTATGAACACGATACCA	22	56.1	
5	*trnS-trnG*	LSC	IGS	CTTTAATCCACTCAGCCATC	20	54.9	757
				TGTCCAAYCAAAGTAATCAK	20	53.6	
6	*psbB-psbT*	LSC	IGS	TCCAACTACAAGGAGACAA	19	51.3	631
				GTCGAAACTAAGAGGAATGT	20	51	
7	*atpH-atpI*	LSC	IGS	GGATTCATGGTAAGTTCCTY	20	53.6	738
				GGATTCACCTATATAAGCCG	20	54	
8	*rpl23-trnL*	SSC	IGS	TCTTTAGAGGCAGTAAAAGC	20	51.9	764
				GTGTCTACCAATTTCACCAT	20	52.3	
9	*psbk-psbI*	LSC	IGS	TCTCTTAGCCTTTGTTTGG	19	53.2	571
				CGAAGATGAAGAGAGAAACA	20	53	
10	*matK-trnK*	LSC	IGS	GATATARGAAGTTTTGTTSCCG	22	55.5	1,050
				GGGTCTAGTKAATAAATGGATAGAG	25	56.3	

*LSC: larger single copy region, SSC: small single copy region, IGS: intergenic spacer, TM: Primer Melting Temperature.

Annex 13 - Adenine-Thymine content expressed as a percentage of the coding and non-coding regions.



Annex 14 - The sequence variability (SV) and Parsimony informative sites (PIS) of nrITS and the top ten variable regions.