NUSANTARA BIOSCIENCE
Vol. 15, No. 1, pp. 12-21
May 2023

ISSN: 2087-3948
E-ISSN: 2087-3956
DOI: 10.13057/nusbiosci/n150102

In silico comparative analysis of the complete chloroplast genome sequences in different jewel orchid species

MINH PHUONG NGUYEN[1], THI HUONG TRINH[1], THI KIM ANH NGO[1], SASANTI WIDIARSIH[2], VIET THE HO[1,♥]

[1]Faculty of Biology and Environment, Ho Chi Minh City University of Food Industry. 140 Le Trong Tan Street, Ho Chi Minh City 72009, Vietnam. Tel./Fax. +84-933-025-55-23, ♥email: thehv@hufi.edu.vn.

[2]Research Center for Radiation Process Technology, Research Organization of Nuclear Energy, National Research and Innovation Agency. Jl. Lebakbulus Raya No. 49, Jakarta 12440, Indonesia

Manuscript received: 30 October 2022. Revision accepted: 27 December 2022.

Abstract. *Nguyen MP, Trinh TH, Ngo TKA, Widiarsih S, Ho VT. 2023. In silico comparative analysis of the complete chloroplast genome sequences in different jewel orchid species. Nusantara Bioscience 15: 12-21.* Jewel orchid is the common name of several orchid species which can be alike in morphological characteristics but variable in medicinal properties. As these plants are utilized to treat several diseases, their natural existence in the wild habitat is rapidly diminished. Therefore, a better understanding of the genetic information of this plant for better genetic conservation and development of these plants is necessary. In this study, a total of 18 published chloroplast genomes of 18 jewel orchid species determined by the next-generation sequencing method were retrieved from NCBI GenBank and targeted for genomic characterization and phylogenetic analyses. Different bioinformatics tools were utilized to characterize these genomes' genomic structure, repetitive sequences, interspecific variation, divergence, and phylogenetic relationships. The obtained data revealed that the chloroplast genomes of different jewel orchid species varied in length between 151,414 (*Anoectochilus formosanus* MN880624.1) and 154,375 (*Goodyera biflora* OM314910.1). Each species contains 34-87 SSR loci which could be useful as molecular markers for further genetic diversity study of this plant. Structural variations in the expansion and contraction of inverted repeat regions were also considered. Phylogenetic analysis identified a close relationship among species belonging to the *Goodyera* genus, and this genus is distinctive from other genera such as *Anoectochilus*, *Cystorchis*, *Dossinia*, *Ludisia*, and *Macodes*. The obtained results show a high potential of deeper characterizing the chloroplast genome of jewel orchids for species classification, identification, molecular breeding, and evolutionary exploration of these important herbal plants.

Keywords: Chloroplast genome, jewel orchid, next-generation sequencing, phylogenetic relationship, SSR

INTRODUCTION

Jewel orchid is a general name of several plant species in the Orchidaceae family. This name is used for plants that have smooth brocade leaves with beautiful veins. These plant species have been used for numerous purposes, such as ornamental plants for their beautiful foliage and health care purposes since they contain several valuable medicinal properties such as antioxidant, antitumor, and immunomodulatory agents (Winarto and Samijan 2018). Therefore, these herbs are used as treatments for several diseases as well as cancer prevention. However, because of their treasured medicinal values, they are exhaustedly exploited in the wild. Thus, the study for a better understanding of the genetic composition of this plant type for proper identification, conservation, and development of these plants is necessary. Since several visual characteristics of jewel orchid species are nearly similar, properly conserving species with high economic and pharmaceutical values is challenging. The current jewel orchid identification is mainly based on the morphological characteristics of leaves, flowers, and stems. Nevertheless, these methods are easily compromised by several factors such as different plant developmental stages or environmental conditions. Therefore, seeking a new marker for a more accurate classification of this plant group for better development and conservation is urgently needed.

Several phylogenetic studies have provided huge information about these herbal plants' relationships and evolutionary processes. However, these studies are mostly based on fragmentation analysis, such as RAPD and ISSR (David et al. 2020; Tran et al. 2022) or DNA barcodes (Ho et al. 2021; Raskoti and Ale 2021). Due to the methods' nature based on fragment length analysis, variations in internal DNA sequence are easily overlooked. On the other hand, DNA barcoding identification is based on the sequence of only a limited number of genome regions which does not provide sufficient discriminating power due to the similarity of sequences between species (Galimberti et al. 2014). Currently, the highest discriminating ability of DNA barcodes is only 70%, and this may be reduced in plants with complex genomes (Besse et al. 2021). Consequently, the relationship among jewel orchid species still lacks convincing evidence and needs further investigation.

The variation in chloroplast genomes in plants has been widely applied in studies on population genetics, evolutionary relationships, and genetic relationships to serve the conservation of plants under threatened extinction or developed molecular markers to accelerate the breeding process of plants with higher efficiency. Recently, next-

generation sequencing (NSG), a method simultaneously sequences several DNA or RNA molecules in a short time, with low cost and high accuracy, are widely used to replace the Sanger method for DNA sequencing in most applications that require a sequence of several target DNA or RNA molecules at the same time or identify the entire genomes. NGS enables a rapid increase in the completion of chloroplast genomes and has shifted the study of phylogenetics to phylogenomics (Behura 2015). In addition, many studies show that NGS can solve the remaining problems of DNA barcode technology, especially in determining plant origin, checking the mixing of poor-quality ingredients into products as well as traceability of plant-derived materials (Galimberti et al. 2014). At present, with the development of many new generation sequencing platforms, the sequencing of whole organism genomes, in general, and chloroplast genomes, in particular, are done easily and quickly. Consequently, several chloroplast genome sequences of different jewel orchid species have been published. A deeper understanding of the information on several chloroplast sequences simultaneously from availably published chloroplast genomes is an important basis for developing conservation and development programs for these plants.

In this study, the complete chloroplast genome sequences from 18 jewel orchid species were obtained from public databases and used for analysis. Based on the sequence comparison results, variable DNA regions between species found in this study would be used to design specialized primer pairs to help distinguish species to serve the conservation, breeding, and development of orchid species.

MATERIALS AND METHODS

Sequence annotation and comparison of chloroplast genomes

Eighteen complete chloroplast genome sequences of different jewel orchid species were retrieved from NCBI GenBank (MW589500.1 *Anoectochilus chapaensis*; LC057212.1 *Anoectochilus emeiensis*; MN880624.1 *Anoectochilus formosanus*; MW589501.1 *Anoectochilus hainanensis*; MN880626.1 *Anoectochilus roxburghii*; MW173020.1 *Anoectochilus zhejiangensis*; MW589507.1 *Cystorchis variegata*; MW589508.1 *Dossinia marmorata*; OM314910.1 *Goodyera biflora*; OM314911.1 *Goodyera henryi*; KT886429.1 *Goodyera procera*; OM314912.1 *Goodyera pubescens*; OM314914.1 *Goodyera schlechtendaliana*; OM314915.1 *Goodyera striata*; OM314916.1 *Goodyera velutina*; MN317571.1 *Ludisia discolor*; MW589527.1 *Macodes petola*; and MW589528.1 *Macodes sanderiana*). For species with several sequences available, only one sequence was randomly selected for further analysis. The Geseq program (https://chlorobox.mpimp-golm.mpg.de/geseq.html) was used to annotate and locate genes in the chloroplast genomes (Tillich et al. 2017). Chloroplot software (https://irscope.shinyapps.io/Chloroplot/) was used to identify the number of protein-coding genes, rRNA genes, tRNA genes, and GC content in each chloroplast genome (Zheng et al. 2020).

Repeat element analysis

The whole chloroplast genome sequences of 18 jewel orchid species were aligned by the MAFFT program (https://mafft.cbrc.jp/alignment/server/) with the following parameters: BLOSUM62 for scoring matrix for amino acid sequences and 200PAM/k=2 for the scoring matrix to find sequence variation (Katoh et al. 2019). The alignment result was then used to determine the DNA polymorphism by the DnaSP software to analyze nucleotide diversity (Pi) and the total number of mutations (Eta). Evolutionary divergence for each data set and pattern of nucleotide substitution was performed by MEGA X (https://www.megasoftware.net/) with default parameters (Kumar et al. 2018). The evolutionary distance between sequences will be calculated based on the p-value (p-distance) through the Kimura two-parameter algorithm of the MEGA 7.0 software to determine the genetic differences among chloroplast genomes. Chloroplast genomes were compared using the VISTA program (https://genome.lbl.gov/vista/index.shtml) in Shuffle-LAGAN mode (Brudno et al. 2003). The comparison of the LSC/IRB/SSC/IRA junctions among these related species was visualized by IRscope (https://irscope.shinyapps.io/irapp/) based on the annotations of their available chloroplast genomes in GenBank (Amiryousefi et al. 2018). Simple Sequence Repeat (SSRs) motifs were detected by MISA (http://pgrc.ipk-gatersleben.de/misa/misa.html) using parameters of the minimum repeats of ten for mononucleotides, six for dinucleotides, five for trinucleotides, four for tetra-nucleotides, and three each for penta- and hexa-nucleotides (Beier et al. 2017). Long repeat regions were defined using REPuter software (https://bibiserv.cebitec.uni-bielefeld.de/reputer) using default parameters such as repeat size of ≥ 30 bp and 90% minimum identity to find four types of repeats, namely forward (F), reverse (R), complementary (C), and palindromic (P) (Kurtz et al. 2001).

Phylogenetic analysis

The MAFFT alignment results were then used to determine the phylogenetic relationship among genomes. Phylogenetic trees of 18 chloroplast genomes were constructed based on Neighbor Joining (NJ), which represents distance methods (Kang et al. 2017) using 1000 bootstrap replicates with chloroplast genome of *Oryza sativa* (MK348618.1) and *Zea mays* (KP966116.1) belonging to Poaceae family as outgroups. Kimura 2-parameter nucleotide substitution model was applied for phylogenetic trees as this is one of the most widely used models for estimating genetic differences due to nucleotide substitution (Nishimaki and Sato 2019). To evaluate the classification resolution of given chloroplast genomes, a genus was considered as clear resolution if all its species are grouped into one monophyletic branch of dendrogram with strong bootstrap support and if species in a specific genus are separated in different branches that genus was considered as unresolved (Sikdar et al. 2018).

RESULTS AND DISCUSSION

Sequence annotation and comparison of chloroplast genomes

Orchidaceae, one of the largest and species-richest families in flowering plants, comprises approximately 880 genera, with 26,000 species distributed worldwide (Fay and Chase 2009). In this study, 21 chloroplast genomes of jewel orchids were obtained from GenBank from 18 species. There are 2 chloroplast sequences in 3 species: *Anoectochilus roxburghii*, *Goodyera schlechtendaliana*, and *Ludisia discolor*. For the remaining 15 species, only one chloroplast sequence is available for each species. Therefore, for equality, each chloroplast sequence from each species was kept for further analysis (Table 1). By using the Geseq program, the structural characteristics and gene contents of 18 chloroplast genomes were obtained (Figure 1). Similar to other chloroplast genomes, all chloroplast genomes in this study have a four-part structure consisting of Large Single Copy (LSC) region, Small Single Copy (SSC) region, and two Inverted Repeat (IRs) regions.

The genome size of 18 jewel orchids ranged from 151,414 bp in *A. formosanus* to 154,375 bp in *G. biflora*, which is slightly smaller than the chloroplast genome of other species in Orchidaceae in *Paphiopedilum* genus such as *P. barbigerum* (155,965 bp), *P. bellatulum* (156,567 bp), *P. henryanum* (155,886 bp), *P. hirsutissimum* (156,571 bp), and the hybrid cultivar *P. 'GZSLKY' Youyou* (160,503 bp) (Liu et al., 2022) or *Cypripedium* genus such as *C. palangshanense* (207,142 bp), *C. debile* (162,773 bp), *C. subtropicum* (212,668 bp), *C. tibeticum* (197,815 bp), *C. japonicum* (174,417 bp), *C. formosanum* (178,131 bp) and *C. calceolus* (175,122 bp) (Zhang et al., 2022). The protein coding gene numbers vary from 81 (*G. procera)* to 93 (*A. emeiensis*). Eight rRNA genes were detected in all chloroplast genomes, similar to rRNA gene numbers in different genera in the Orchidaceae family, such as *Paphiopedilum* (Liu et al. 2022); *Cypripedium* (Zhang et al. 2022). Similarly, the number of genes encoding for tRNA was negligibly different among chloroplast genomes ranging from 37 to 39 genes, except *A. emeiensis*, with up to 46 genes. The average GC content of the chloroplast genomes in the 18 species was comparable and ranged from 37% to 38%. GC content is an important parameter in the DNA sequence that directly alters protein amino acid composition in plants to cope with specific environments. The genome with high GC content will be more conserved, making it more stable and harder to transcribe. Thus, the differences in GC content could be due to the different pressure of natural selection among species. The LSC lengths ranged from 81,879 bp (*A. formosanus*) to 83,596 bp (*G. henryi*)), the SSC lengths ranged from 17,026 bp (*C. variegata*) to 18,406 bp (*G. procera*) and the length of the IR region was enlarged to 26,069 bp (*A. hainanensis)* to 26,572 bp (*G. striata)*. However, the tRNA number is conserved in all chloroplast genomes of 18 species with 35 tRNA for each chloroplast genome.

Using the Dnasp program, 10,788 polymorphic sites were detected and the nucleotide diversity value is at 0,01712 lower than similar values from four other orchid species, namely *D. densiflorum*, *G. densiflorum*, *C. aloifolium* and *R. retusa* (Roy et al. 2016). However, the nucleotide diversity value obtained from our study is almost two times higher than reported in the *Paphiopedilum* orchid with an average of 0.00962 (Liu et al. 2022) and even higher than other plants in distant taxonomy such as *Pennisetum* (0.00638) (Xu et al. 2021). Generally, plants' low level of nucleotide diversity is due to the selection pressure of humans with economical plants. On the other hand, the low nucleotide diversity of wild plants could result from collecting samples in a narrow area. The divergence of 18 chloroplast genomes ranged from 0.001 to 0.028 (Table 2).

Table 1. Size comparison of plastome features of 18 jewel orchid species

Accession code	Scientific name	Genome size (bp)	LSC size (bp)	SSC size (bp)	IR size (bp)	Coding genes	rRNA	tRNA	GC content (%)
MW589500	*A. chapaensis*	152,395	82,630	17,125	26,320	90	8	38	37
LC057212.1	*A. emeiensis*	152,650	82,670	17,342	26,319	93	8	46	37
MN880624	*A. formosanus*	151,414	81,879	17,342	26,313	90	8	37	37
MW589501	*A. hainanensis*	152,645	82,881	17,626	26,069	90	8	38	37
MN880626.1	*A. roxburghii*	152,821	82,683	17,478	26,324	91	8	37	37
MW173020.1	*A. zhejiangensis*	152,509	82,247	17,026	26,498	90	8	38	37
MW589507	*C. variegata*	152,269	82,336	17,443	26,551	90	8	38	37
MW589508	*D. marmorata*	152,881	83,466	17,893	26,508	90	8	38	37
OM314910	*G. biflora*	154,375	83,596	17,720	26,488	89	8	38	37
OM314911	*G. henryi*	154,292	82,496	18,406	26,169	89	8	38	37
KT886429.1	*G. procera*	153,240	82,101	17,876	26,220	81	8	39	38
OM314912.1	*G. pubescens*	152,417	82,674	17,999	26,535	89	8	38	37
OM314914	*G.schlechtendaliana*	153,743	82,081	17,871	26,395	89	8	38	37
OM314915.	*G. striata*	152,742	82,922	17,258	26,572	89	8	38	37
OM314916	*G. velutina*	153,997	82,659	17,513	26,438	89	8	38	37
MN317571.1	*L. discolor*	153,324	82,777	17,413	26,463	87	8	38	37
MW589527	*M. petola*	153,048	82,670	17,342	26,319	90	8	38	37
MW589528	*M. sanderiana*	153,116	81,879	17,342	26,313	90	8	38	37

Table 2. Estimates of evolutionary divergence among chloroplast genome sequences of 18 jewel orchid species

No.	Accession code	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	*A. chapaensis*																	
2	*A. emeiensis*	0.004																
3	*A. formosanus*	0.005	0.004															
4	*A. hainanensis*	0.005	0.004	0.001														
5	*A. roxburghii*	0.005	0.005	0.001	0.001													
6	*A. zhejiangensis*	0.005	0.005	0.002	0.002	0.002												
7	*C. variegata*	0.017	0.017	0.017	0.017	0.017	0.017											
8	*D. marmorata*	0.016	0.015	0.016	0.015	0.016	0.016	0.012										
9	*G. biflora*	0.016	0.016	0.016	0.016	0.016	0.016	0.012	0.006									
10	*G. henryi*	0.017	0.017	0.017	0.017	0.017	0.017	0.016	0.015	0.015								
11	*G. procera*	0.025	0.025	0.025	0.025	0.025	0.025	0.026	0.025	0.025	0.026							
12	*G. pubescens*	0.022	0.022	0.022	0.022	0.022	0.022	0.023	0.022	0.022	0.023	0.011						
13	*G. schlechtendaliana*	0.023	0.023	0.023	0.023	0.023	0.023	0.023	0.022	0.022	0.024	0.012	0.006					
14	*G. striata*	0.024	0.024	0.025	0.025	0.025	0.025	0.025	0.024	0.024	0.025	0.017	0.013	0.013				
15	*G. velutina*	0.026	0.026	0.026	0.026	0.027	0.027	0.027	0.026	0.026	0.027	0.019	0.015	0.016	0.015			
16	*L. discolor*	0.027	0.027	0.028	0.028	0.028	0.028	0.028	0.027	0.027	0.028	0.020	0.017	0.017	0.018	0.020		
17	*M. petola*	0.021	0.021	0.021	0.021	0.022	0.021	0.022	0.021	0.020	0.022	0.026	0.022	0.023	0.025	0.027	0.028	
18	*M. sanderiana*	0.026	0.026	0.026	0.026	0.026	0.026	0.027	0.026	0.026	0.027	0.020	0.017	0.018	0.020	0.022	0.022	0.027

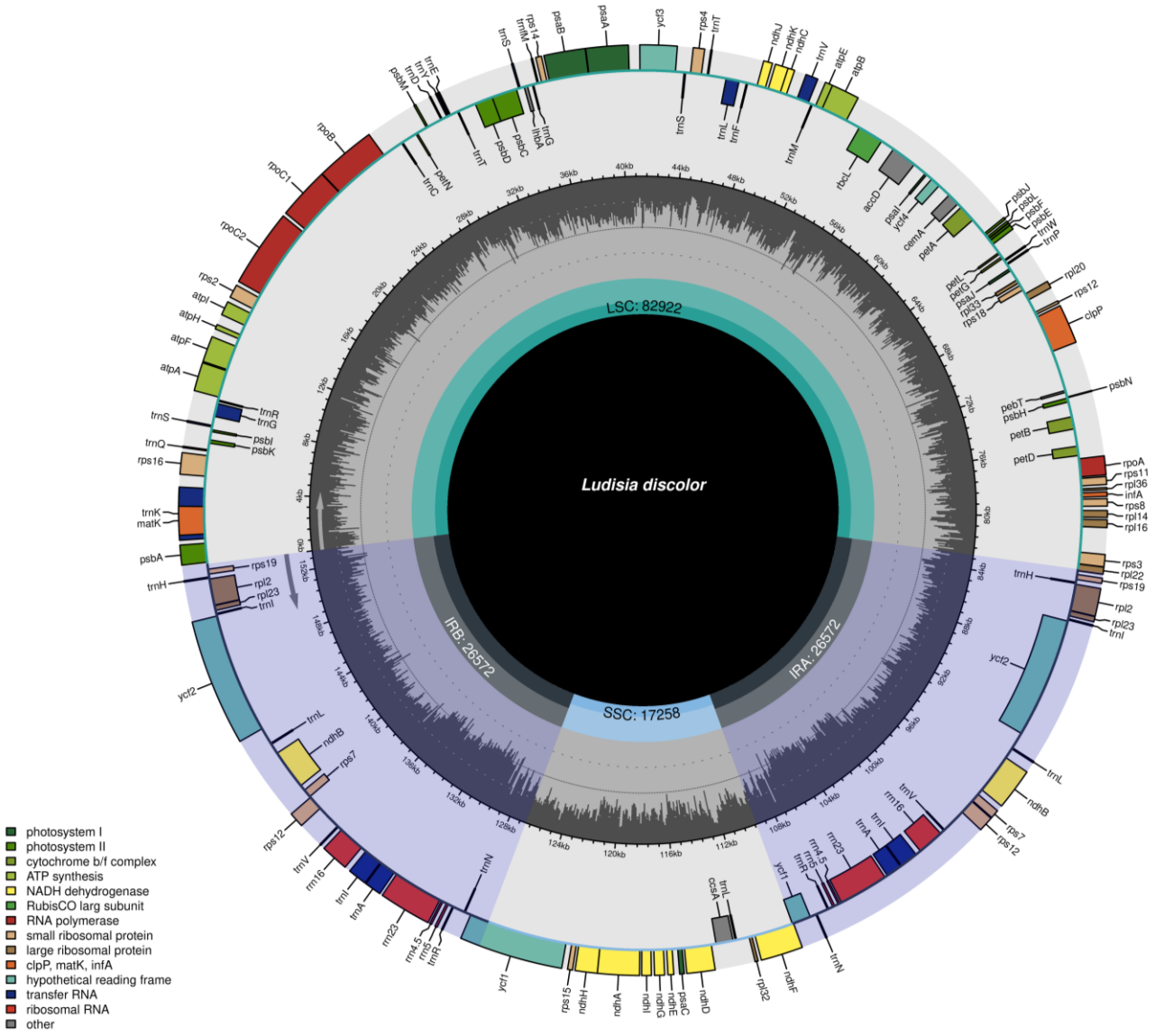

Figure 1. Typical map of jewel orchid chloroplast genome using MN317571.1 of *Ludisia discolor* as example (the genes drawn outside and inside of the circle are transcribed in clockwise and counterclockwise directions, respectively. The main parts of chloroplast genome are written as LSC, SSR, IRA, and IRB. The dark gray color and the light gray color of the inner circle shows the GC content and AT content, respectively)

Table 3. The pattern of nucleotide substitution among chloroplast genome sequences of 18 jewel orchid species (in percentage)

Nitrogenous bases	A	T	C	G
A	-	6.69	4.02	***10.81***
T	6.52	-	***10.81***	3.87
C	6.52	***18.0***	-	3.87
G	***18.19***	6.69	4.02	-

(Patterns and rates of substitutions were estimated under the Tamura-Nei model. Rates of different transitional substitutions and those of transversionsal substitutions are shown in bold and italics, respectively)

The chloroplast genome of *L. discolor* shows the largest difference with the remaining chloroplast genomes. The substitution of different nucleotides in whole genomes was evaluated on the entire codon position and shown in Table 3. Theoretically, there are 4 types of transitions; the substitution of a purine for a purine (A or G) nucleotide or a pyrimidine for a pyrimidine (C and T) nucleotide), also 8 types of transversions (the substitution of a purine (A or G) nucleotide for a pyrimidine nucleotide or vice versa). The expected ratio between transition and transversion is 0.5. In this study, the transitional substitution (57,8%) was significantly higher than the transversional substitution (43,2%). The high frequency of transitional substitution was also found among different species in the *Dracunculus* clade (Abdullah et al. 2021) or *Catalpa* genus (Li et al. 2022). Substitution is the most common mutation that causes variation and diversity among individuals and functions as a force for species evolution. It is also vital for phylogenetic construction since the transition bias in nucleotide transitions provides important information for clustering analysis.

Repeat element analysis

With the default parameters of MISA program of tandem repeat sequences consisting of 1-6 nucleotide repeat units, the relative abundance of SSR is detected. Microsatellites or SSRs are commonly used to identify the variable in the genomes of species. In total, 1,078 SSRs were detected among 18 jewel orchid species from 34 SSRs (*G. schlechtendaliana*) to 87 SSR (*D. marmorata*) with an average of approximately 60 SSRs per chloroplast genomes (Table 4). Seven SSR motifs were detected, namely A, T, C, G, AT, TA, and TTC. Two mononucleotide types consisting of T and A are the most dominant, with a frequency of 676 (62.7%) and 339 (31.4%), respectively. In contrast, C and G mononucleotide types are rarely detected, with only 3 and 1. Another dinucleotide (AT and TA) and trinucleotide (TTC) motifs were also identified with a low percentage. The A and T motifs seem common among plant chloroplast genomes. Similarly, Liu and colleagues reported the appearance of A and T motifs up to 66.39% in six oak species (*Quercus* L.) (Liu et al. 2021). The repeat motif type of chloroplast genomes in this study is less than those from other plants, such as two species in the *Morus* genus, which possess up to 18 motif types (Li et al. 2016) or up to 27 motif types in oak (Liu et al. 2021).

In addition, the 18 chloroplast sequences were analyzed with the REPuter program to determine the abundance of four oligonucleotide repeat types, namely forward (F), palindromic (P), reverse (R), and complementary (C). The number and type of repeat elements are largely variable among 18 jewel orchid species (Figure 2), ranging from 37 (*G. schlechtendaliana*) to 50 units (*A. chapaensis, A. emeiensis, A. formosanus, A. hainanensis*, *C. variegate*, *G. biflora*, *G. procera*, and *G. striata*). A total of 852 repeat elements were identified, palindromic repeats are the most commonly found, accounted up to 355 (41.67%) of the number of repeat elements. The second position was 274 (32.16%) forward, followed by 176 (20.66%) reverse and 47 (5.52%) complement repeat elements. These SSRs have a high potential to be used as candidate genetic markers. They are distributed widely in chloroplast genomes and serve as molecular markers for phylogenetic relationship inference. Moreover, SSRs are also associated with different types of genome rearrangement, recombination, and large inversions, which are useful for further phylogenetic studies.

Although the structure of chloroplast genomes is highly conserved among terrestrial plants, significant variation in the expansion and contraction of IR regions affects different genome sizes among plants. The LSC/IRb/SSC/IRa/LSC borders and adjacent genes were characterized to find similarities and differences among 18 jewel orchid species (Figure 3). Although the genomic structure and size were highly conserved in the 18 chloroplast genomes, the IR/SC boundary regions still showed considerable differences. The four regions are varied in length, of which *rps*3, *rpl*22, *rpl*19, *ycf*1, and *ndh*F genes were present at the junctions of the LSC/IR and SSC/IR borders. Notable variations were observed in the expansion and contraction of the IR regions. For the LSC/IR borders, *rpl*22 genes of 17/18 species are extended 12-95 bp into the IRb regions, whereas only this gene of *G. procera* was localized completely in LSC region. It indicates that this border has moved toward the LSC region compared to *G. procera* (Huang et al. 2020). On the contrary, only *ycf*1 gene in *L. discolor* stays extended from IRb to SSC regions. Interestingly, they are missing in *ndh*F (chloroplast NADH dehydrogenase F) genes in IRb/SSC regions of *D. marmorata*, *A. hainanensis* and *M. petola*, suggesting that the loss of this gene should have occurred independently among jewel orchids species. A previous study also reported that this gene is present in *Viburnum dilatatum* but not in at least six other species in the *Viburnum* genus (Park et al. 2020). This gene is often commonly pseudogenized or lost in different species in the *Paphiopedilum* genus (Liu et al. 2022).

Table 4. The different repeat types in the chloroplast genomes of 18 jewel orchid species

Scientific name	Repeat motifs							Total SSRs
	A	T	C	G	AT	TA	TTC	
A. chapaensis	21	43	0	0	2	3	0	69
A. emeiensis	21	49	1	0	2	3	1	77
A. formosanus	19	45	0	0	3	2	1	70
A. hainanensis	19	48	0	0	2	2	0	71
A. roxburghii	17	48	1	0	2	3	1	72
A. zhejiangensis	21	49	0	0	2	3	1	76
C. variegata	13	24	0	0	2	4	0	43
D. marmorata	32	54	0	0	0	1	0	87
G. biflora	18	24	0	0	0	1	0	43
G. henryi	11	25	0	0	1	1	0	38
G. procera	13	27	0	0	2	0	0	42
G. pubescens	15	34	0	0	0	1	0	50
G.schlechtendaliana	10	20	0	1	0	3	0	34
G. striata	17	31	0	0	0	1	0	49
G. velutina	14	32	1	0	0	4	0	51
L. discolor	22	38	0	0	1	1	0	62
M. petola	29	49	0	0	0	1	0	79
M. sanderiana	27	36	0	0	1	1	0	65
Total	676	339	3	1	20	35	5	1,078

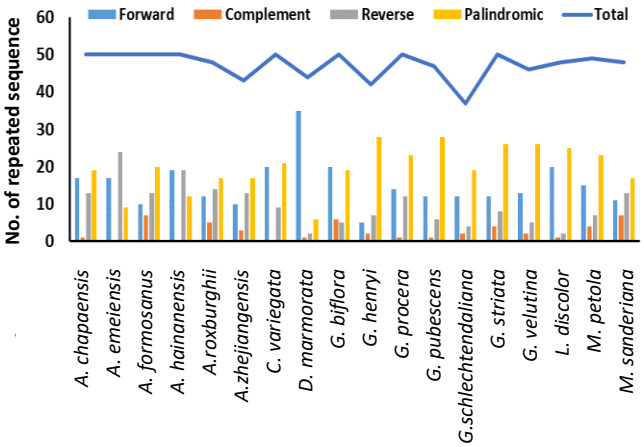

Figure 2. Number of repeated sequences in 18 jewel orchid chloroplast genomes

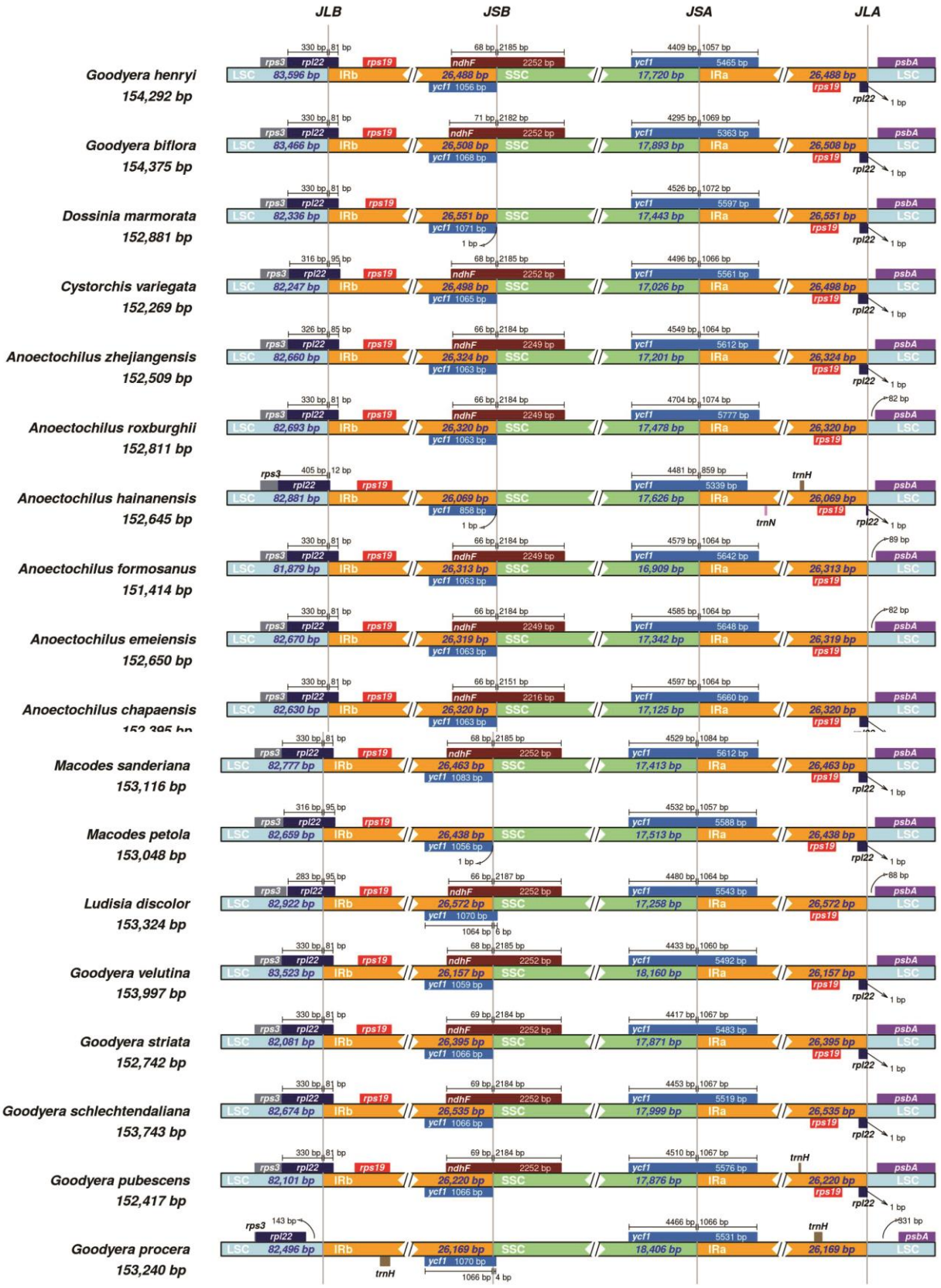

Figure 3. The comparison of the Large Single Copy (LSC), inverted repeat (IR) and Small Single Copy (SSC) border regions among 18 jewel orchid chloroplast genomes. Boxes above or below the main lines represent the genes at the IR/SC borders whereas the numbers above the gene indicate the distance from the gene terminal to the boundary region

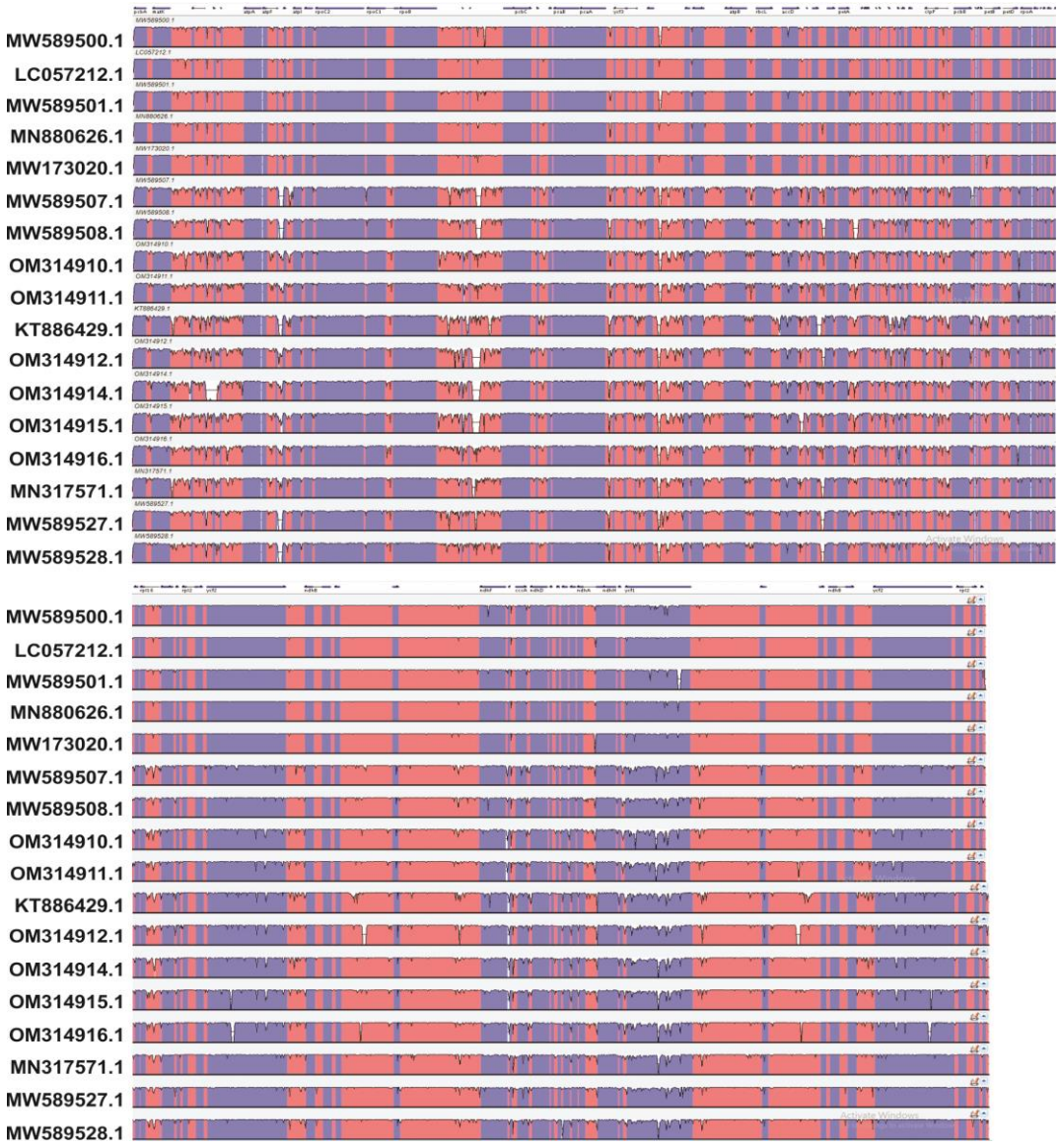

Figure 4. Comparison of chloroplast genomes of 18 jewel orchid species with *Anoectochilus formosanus* (MN880624.1) as the reference using mVISTA program. Coding regions are in blue and non-coding regions are in orange.

In contrast, the structure of SSC/IRa boundary regions is relatively stable. The gene *ycf*1 in the SSC region exhibited an interesting astride at the border of SSC/IRa with the extension from 859 bp (*A. roxburghii*) to 1,084 bp (*M. sanderiana*) into the IRa regions. The related expansions and contractions at SSR and LSC junctions with IRs suggest that the relationships among jewel orchid species may play evolutionary signals. Furthermore, the contractions and expansions at these positions may contribute to the variations in the chloroplast genomes and the IR expansions or contractions are likely to result from the gene conversion during plant speciation (Huang et al. 2020).

The annotated MN880624.1 chloroplast genome was used as a reference in mMISTA for alignment of the chloroplast genome among 18 jewel orchid species (Figure 4). Generally, the size and gene order of 18 analyzed chloroplast genomes are conserved. Nevertheless, some identified divergent regions are *acc*D, *ccs*A, *ycf*1, and *ycf*2.

Phylogenetic analyses

The results of phylogenetic analysis among 18 chloroplast genomes show a significant relationship among jewel orchid species with high bootstrap values (Figure 5). The *Goodyera* and *Anoectochillus* genera were found to be the best conserved clustered in two monophyletic groups. However, this result is opposite to Zhou and colleague's report, which is based on the phylogenetic dendrogram of the chloroplast genome, *G. velutina*, in the same cluster with *A. emeiensis* and *L. discolor* and separated from other species in the *Goodyera* genus such as *G. schlechtendaliana*, *G. goliosa*, *G. fumata* and *G. procera* (Zhou et al. 2019).

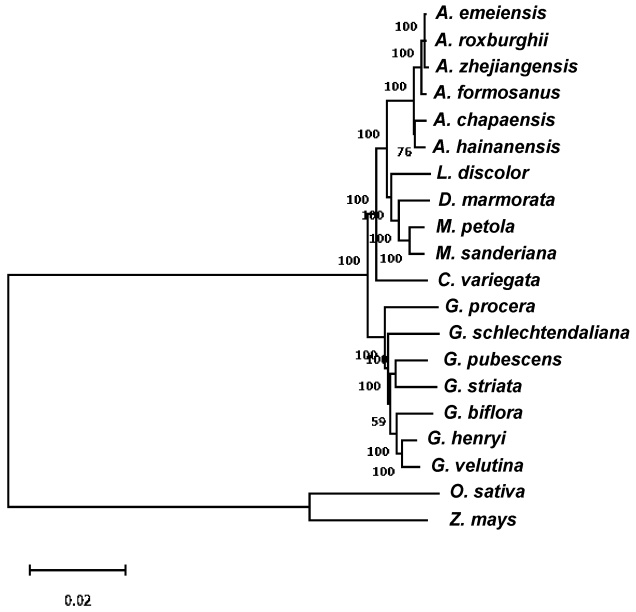

Figure 5. Phylogenetic tree of 18 chloroplast genomes of jewel orchid (The chloroplast sequences of *O. sativa* (rice) and *Z. mays* (maize) are used as outgroups. Numbers near branches are bootstrap values)

In conclusion, in silico analysis of many chloroplast genomes may play crucial roles in studying phylogeny, gene flow, and population genetics among different jewel orchid species. This study reveals the typical structure and content of the chloroplast genomes among the 18 jewel orchid species, an economically important herbal plant. This information regarding similarities and divergence among chloroplast genomes would enrich our understanding of jewel orchid genetic structure. Moreover, information about highly polymorphic regions from *acc*D, *ccs*A, *ycf*1, and *ycf*2 genes would also contribute to molecular markers and highly divergent regions, which might be useful for further studies of the taxonomy and phylogeographic of jewel orchid species.

ACKNOWLEDGEMENTS

This work was supported by the Ho Chi Minh City University of Food Industry- Vietnam through the HUFI fund for Science and Technology under the Contract No. 157/HD-DCT.

REFERENCES

Abdullah, Henriquez CL, Mehmood F, Hayat A, Sammad A, Weseem S, Waheed MT, Matthews PJ, Croat TB, Poczai P, Ahmed I. 2021. Chloroplast genome evolution in the *Dracunculus* clade (Aroideae, Araceae). Genomics 113: 183-192. DOI: 10.1016/j.ygeno.2020.12.016.

Amiryousefi A, Hyvönen J, Poczai P. 2018. IRscope: an online program to visualize the junction sites of chloroplast genomes. Bioinformatics 34 (17): 3030-3031. DOI: 10.1093/bioinformatics/bty220.

Behura S. 2015. Insect phylogenomics. Insect Mol Biol 24 (4): 403-411. DOI: 10.1111/imb.12174.

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

Besse P, Da Silva D, Grisoni M. 2021. Plant DNA barcoding principles and limits: A case study in the genus *Vanilla*. Methods Mol Biol 2222: 131-148. DOI: 10.1007/978-1-0716-0997-2_8.

Brudno M, Malde S, Poliakov A, Do CB, Couronne O, Dubchak I, Batzoglou S. 2003. Glocal alignment: Finding rearrangements during alignment. Bioinformatics 19S1: i54-i62. DOI: 10.1093/bioinformatics/btg1005.

David D, Rusdi NA, Mokhtar RAM, Faik AAM, Gansau JA. 2020. Establishment of in vitro regeneration protocol for Sabah's jewel orchid, *Macodes limii* J.J. Wood & A.L. Lamb. Horticulturae 8: 155. DOI: 10.3390/horticulturae8020155.

Fay MF, Chase MW. 2009. Orchid biology: From Linnaeus via Darwin to the 21st century. Ann Bot 104: 359-364. DOI: 10.1093/aob/mcp190.

Galimberti A, Labra M, Sandionigi A, Bruno A, Mezzasalma V, De Mattia F. 2014. DNA barcoding for minor crops and food traceability. Adv Agric 2014: 832875. DOI: 10.1155/2014/831875.

Ho VT, Tran TKP, Vu TT, Widiarsih S. 2021. Comparison of *mat*K and *rbc*L DNA barcodes for genetic classification of jewel orchid accessions in Vietnam. J Genet Eng Biotechnol 19: 93. DOI: 10.1186/s43141-021-00188-1.

Huang S, Ge X, Cano A, Salazar BGM, Deng Y. 2020. Comparative analysis of chloroplast genomes for five *Dicliptera* species (Acanthaceae): Molecular structure, phylogenetic relationships, and adaptive evolution. Peer J 8: e8450. DOI: 10.7717/peerj.8450.

Kang Y, Deng Z, Zang R, Long W. 2017. DNA barcoding analysis and phylogenetic relationships of tree species in tropical cloud forests. Sci Rep 7: 12564. DOI:10.1038/s41598-017-13057-0.

Katoh K, Rozewicki J, Yamada KD. 2019. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform 20 (4): 1160-1166. DOI: 10.1093/bib/bbx108.

Kumar S, Stecher G, Li M, Knyaz C, and Tamura K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Mol Biol Evol 35: 1547-1549. DOI: 10.1093/molbev/msy096.

Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. 2001. REPuter: The manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res 29 (22): 4633-4642. DOI: 10.1093/nar/29.22.4633.

Li F, Liu Y, Wang J, Xin P, Zhang J, Zhao K, Zhang M, Yun H, Ma W. 2022. Comparative analysis of chloroplast genome structure and phylogenetic relationships among six taxa within the genus *Catalpa* (Bignoniaceae). Front Genet 13: 845619. DOI: 10.3389/fgene.2022.845619.

Li QL, Guo JZ, Yan N, Li CC. 2016. Complete chloroplast genome sequence of cultivated *Morus* L. species. Genet Mol Res 15 (4): gmr15048906. DOI: 10.4238/gmr15048906.

Liu H, Ye H, Zhang N, Ma J, Wang J, Hu G, Li M, Zhao P. 2022. Comparative analyses of chloroplast genomes provide comprehensive insights into the adaptive evolution of *Paphiopedilum* (Orchidaceae). Horticulturae 8: 391. DOI: 10.3390/horticulturae8050391.

Liu X, Chang E, Liu J, Juang Z. 2021. Comparative analysis of the complete chloroplast genomes of six white oaks with high ecological amplitude in China. J For Res 32: 2203-2218. DOI: 10.1007/s11676-020-01288-3.

Nishimaki T, Sato K. 2019. An extension of the Kimura two-parameter model to the natural evolutionary process. J Mol Evol 87 (1): 60-67. DOI: 10.1007/s00239-018-9885-1.

Park J, Xi H, Oh SH. 2020. Comparative chloroplast genomics and phylogenetic analysis of the *Viburnum dilatatum* complex (Adoxaceae) in Korea. Korean J Pl Taxon 50 (1): 8-16. DOI: 10.11110/kjpt.2020.50.1.8.

Raskoti BB, Ale R. 2021. DNA barcoding of medicinal orchids in Asia. Sci Rep 11: 23651. DOI: 10.1038/s41598-021-03025-0.

Roy SC, Moitra K, Sarker DD. 2016. Assessment of genetic diversity among four orchids based on ddRAD sequencing data for conservation purposes. Phyisol Mol Biol Plant 23 (1): 169-183. DOI: 10.1007/s12298-016-0401-z.

Sikdar S, Tiwari S, Thakur VV, Sapre S. 2018. An *in silico* approach for evaluation of *rbc*L and *mat*K loci for DNA barcoding of Fabaceae family. Intl Chem Stud 6 (6): 2446-2451.

Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R and Greiner S. 2017. GeSeq - versatile and accurate annotation of organelle genomes. Nucleic Acids Res 45: W6-W11.

Tran TKP, Pham MH, Trinh TH, Widiarsih S, Ho VT 2022. Investigation of the genetic diversity of jewel orchid in Vietnam using RAPD and ISSR markers. Biodiversitas 23 (9): 4816-4825. DOI: 10.13057/biodiv/d230950.

Winarto B, Samijan. 2018. Axillary shoots derived from shoot tips in in vitro mass propagation of *Anoectochilus formosanus* Hayata. J Agric Sci 2: 121-130. DOI: 10.15159/jas.18.11.

Xu J, Liu C, Song Y and Li M. 2021. Comparative analysis of the chloroplast genome for four *Pennisetum* species: Molecular structure and phylogenetic relationships. Front Genet 12: 687844. DOI: 10.3389/fgene.2021.687844.

Zhang JY, Liao M, Cheng YH, Feng Y, Ju WB, Deng HN, Li X, Plenkovic-Moraj A, Xu B. 2022. Comparative chloroplast genomics of seven endangered *Cypripedium* species and phylogenetic relationships of Orchidaceae. Front Plant Sci 13: 911702. DOI: 10.3389/fpls.2022.911702.

Zheng S, Poczai P, Hyvonen J, Tang J, Amiryousefi A. 2020. Chloroplot: An online program for the versatile plotting of organelle genomes. Font Genet 11:576124. DOI: 10.3389/fgene.2020.576124.

Zhou J, Xie TX, Ma SH, Chen MK, Zheng QD, Ai Y. 2019. The complete chloroplast genome sequence of *Goodyera foliosa* (Orchidaceae). Mitochondrial DNA Part B 4 (2): 3477-3478. DOI: 10.1080/23802359.2019.1674728.