AJAB

Asian J Agric & Biol. xxxx(x).
DOI: 10.35495/ajab.2024.059

Original Article

The chloroplast genome of *Diplazium polypodioides* and its comparison within the family Athyriaceae

Adeel Yunus[1], Nighat Sultana[1*], Alia Gul[2], Abdullah[3,4]
[1]Department of Biochemistry, Hazara University, Mansehra, Pakistan
[2]Department of Botany, Hazara University, Mansehra, Pakistan
[3]State Key Laboratory of Chinese Medicine Modernization, Tianjin University of Traditional Chinese Medicine, Tianjin, 301617, China
[4]Haihe Laboratory of Modern Chinese Medicine, Tianjin, 301617, China

Received:
March 26, 2024
Accepted:
September 17, 2024
Published Online:
October 23, 2024

**Corresponding author email:*
nighat.sultana@hu.edu.pk

Abstract

The plant family Athyriaceae Alston consists of 3 genera and 650 species dispersed in different regions of the world, including *Athyrium* Roth, *Diplazium* Sw., and *Deparia* Hook. & Grev. The family has diverse morphological characteristics, ranging from creeping rhizomes to ascending or erect scales at apices. In the present study, the chloroplast (cp) genome of *Diplazium polypodioides* Blume was *de novo* assembled and compared with twelve other reported genomes of the species of the family Athyriaceae. The cp genome of *D. polypodioides* was 152,009 bp and showed a quadripartite structure in which a large single copy (82,389 bp) and a small single copy (22,303 bp) were separated by a pair of long inverted repeats (IRa and IRb: 23,659 bp each). We identified 116 genes, including 4 rRNAs, 29 tRNAs, and 84 protein-coding genes, with 15 genes duplicated in inverted repeats. The cp genome sizes of the thirteen analyzed species ranged from 150,797 bp (*Diplazium striatum* Desv.) to 152,009 bp (*D. polypodioides*). Despite high variability in SSRs and oligonucleotide repeats, the species showed similarities in GC content, contraction and expansion of inverted repeats, codon usage, amino acid frequency, and substitutions. Transition substitutions were more common than transversion substitutions across all species. Phylogenetic analysis of 84 protein-coding genes revealed monophyletic relationships among the limited species of three genera, namely, *Diplazium*, *Athyrium*, and *Deparia*. In addition, *Diplazium* was more closely related to *Athyrium* than to *Deparia*. Our study provides preliminary insights into the evolutionary dynamics of the cp genome in Athyriaceae and clarifies its phylogenetic relationships.

Keywords: *Athyrium*, Athyriaceae, *Deparia*, *Diplazium polypodioides*, Pteridophytes, Phylogenetics

How to cite this:

Yunus A, Sultana N, Gul A and Abdullah. The chloroplast genome of *Diplazium polypodioides* and its comparison within the family Athyriaceae. Asian J. Agric. Biol. xxxx(x): 2024059. DOI: https://doi.org/10.35495/ajab.2024.059

Introduction

The chloroplast plays an important role in photosynthesis. The chloroplast (cp) genome has a mostly quadripartite structure in which large single copy (LSC) and small single copy (SSC) are separated by a pair of inverted repeats (IRs: IRa and IRb) (Palmer, 1985; Daniell et al., 2016; Abdullah et al., 2019; Jamal et al., 2021). Its size varies from 107 kbp (*Cathaya argyrophylla* Chun & Kuang) to 218 kbp (*Pelargonium* × *hortorum* L.H. Bailey), and it contains up to 120 genes (Daniell et al., 2016). In most angiosperms and pteridophytes, cp genome inheritance is maternal, whereas it is paternal in some gymnosperms (Neale and Sederoff, 1989; Daniell, 2007). The cp genome has evolved more slowly than the nuclear genome and lacks homologous recombination, such as in meiosis (Palmer, 1985). However, some regions are predisposed to mutations such as substitutions, insertions-deletions (InDels), repeats, and inversions (Abdullah et al., 2019; Xu and Wang, 2021). Additionally, intron loss, gene duplication, and pseudogenes within the cp genome have been reported, providing insights into its evolutionary dynamics (Abdullah et al., 2020a, Abdullah et al., 2021a, Rehman et al., 2021). Owing to its slow evolution and inheritance from a single parent, the cp genome is suitable for diverse studies ranging from population genetics to phylogenetics (Ahmed et al., 2020; Abdullah et al., 2021a; Li et al., 2021; Wang et al., 2021).

The plant family Athyriaceae includes 650 species in three genera, *Athyrium*, *Diplazium*, and *Deparia* (Schuettpelz et al., 2016; Wei et al., 2018). *Diplazium* is the largest genus, with 465 species worldwide in tropical and subtropical regions (POWO, 2024). The species belonging to *Diplazium* are traditionally used to treat diabetes, smallpox, asthma, hypertension, diarrhea, rheumatism, dysentery, constipation, headaches, and fever (Semwal et al., 2021). Species within this genus often share similar medicinal properties. *Diplazium polypodioides* Blume is native to southern Yunnan in China and extends across tropical Asia. This perennial fern thrives predominantly in wet tropical biomes, where it plays an essential role in the ecosystem (POWO, 2024). To the best of our knowledge, few studies have focused on the family Athyriaceae. However, a notable study on foliar epidermal micromorphology and its taxonomic implications has been conducted on selected species within this family, including *Diplazium* species from Pakistan (Shah et al. 2018). This research included *D. polypodioides*, highlighting the potential of foliar epidermal characteristics as valuable taxonomic markers for distinguishing species within Athyriaceae. Hence, to the best of our knowledge, there are no reports on the development of genomic resources of *D. polypodioides* from Pakistan.

In this study, we sequenced the chloroplast genome of *Diplazium polypodioides* to enhance the availability of genomic resources for this medicinally and biogeographically important genus and to analyze its genome structure. We compared the newly sequenced chloroplast genome of *D. polypodioides* with those of 12 previously reported species from the genera *Athyrium*, *Diplazium*, and *Deparia* within the family Athyriaceae. Through this comparative analysis, we identified key genomic features, as well as similarities and differences across these species, providing valuable insights into the molecular evolution of chloroplast genomes within the Athyriaceae family.

Material and Methods

Plant collection, DNA extraction and sequencing

In September 2022, samples of *Diplazium polypodioides* were collected from Mansehra District, Khyber Pakhtunkhwa, Pakistan, at coordinates 34°38'22.46"N, 73°28'8.01"E. The species identification was confirmed by Dr. Alia Gul, a botanist at Hazara University, Mansehra, under voucher number GHNO-05548. Leaves were dried using silica gel before DNA extraction, which was performed with the DNeasy Plant Mini Kit (Qiagen, Germany) following the manufacturer's protocol. The quality and quantity of the extracted DNA were assessed using 1% agarose gel electrophoresis and a NanoDrop spectrophotometer (Thermo Scientific, USA).

Chloroplast genome sequencing, *de novo* assembly, and annotation

The extracted DNA was sent to Novogene, Hong Kong, for whole-genome sequencing, where it was processed using the HiSeq 2500 platform. Paired-end sequencing was performed, generating 150 bp reads with a 350 bp insert size. The cp genome of *D. polypodioides* was assembled *de novo* using NOVOPlasty v.4.3.3 (Dierckxsens et al., 2017). The accuracy of the assembled genome was verified by mapping raw reads to the assembled genome using

Bowtie 2 (Langmead and Salzberg, 2012) embedded in Geneious Prime 2023.1.2, which also provided coverage depth information.

The cp genomes were annotated using GeSeq (Tillich et al., 2017), tRNAscan-SE v.2 (Lowe and Chan, 2016), and ARAGORN v.1.2.3.8 (Laslett and Canback, 2004) with default settings. The annotation of the protein-coding gene was verified by comparing it with other reported species in Geneious Prime 2023.1.2 (Table 1) and performing BLAST with other homologous genes in the National Center for Biotechnology Information (NCBI). Finally, the *de novo* assembled cp genome was deposited in NCBI under accession number PP146611. The five-column text file required for the submission of the annotated genome to NCBI was generated using GB2Sequin (Lehwark and Greiner, 2019). The above mentioned protocol has also been reported in other articles (Henriquez et al., 2020a,b; Abdullah et al., 2020a, 2021).

Comparative analyses and genomic features

The cp genomes of 12 species from three genera in Athyriaceae were obtained from NCBI and reannotated. After the annotations of genes, coding sequences, and introns were corrected, these cp genomes were subsequently compared with the cp genome of *D. polypodioides* (Table 1). The length of the complete genome, LSC, SSC, and IR were analyzed along with the gene content and arrangement via Geneious Prime 2023.1.2 (Kearse et al., 2012). IRscope was used to visualize the contraction and expansion of inverted repeats (IRs) (Amiryousefi et al., 2018). Mauve progressive alignment was used for collinear block analysis among all species for possible long inversions (Darling et al., 2004).

GC analysis, amino acid frequency, and codon usage

The GC content, amino acid frequency, and relative synonymous codon usage (RSCU) of the cp genomes of all the species were determined using Geneious Prime 2023.1.2. The RSCU value was drawn as a heatmap using TBtool (Chen et al., 2020).

Analysis of SSRs and oligonucleotide repeats

We analyzed the simple sequence repeats (SSRs) of whole genomes using the Perl script Microsatellite Identification Tool (MISA) (Beier et al., 2017). The parameters set for SSRs were 10 for mononucleotides, 5 for dinucleotides, 4 for trinucleotides, and 3 for tetra-, penta-, and hexanucleotides.

Additionally, we used the REPuter program (Kurtz et al., 2001) to examine oligonucleotide repeats, including complementary, forward, reverse, and palindromic repeats. For this analysis, we selected parameters requiring a minimum repeat size of 30 base pairs (bp) with up to three mismatches.

Substitution and InDel analyses

The numbers of substitutions and insertions/deletions (Indels) in all regions of the cp genome were determined after performing pairwise alignments using Multiple Alignment Fast Fourier Transform (MAFFT v.1.5.0). Each genome and its regions were aligned to the reference genome *Athyrium anisopterum* Christ (NC_035738.1) (Katoh et al., 2013; He et al., 2022). The numbers and types of substitutions were identified using a Python script. For counting Indels, the pairwise alignments were analyzed using DnaSP v.5.10 (Rozas et al., 2017).

Phylogenetic analysis

All the protein-coding genes (84 unique genes) of the species were extracted and concatenated in Geneious Prime 2023.1.2. The concatenated sequences were aligned using MAFFT along the outgroup species *Matteuccia struthiopteris* (KY427353). The maximum likelihood tree was reconstructed using IQ-tree v.2 with ultrafast bootstrap approximation (Hoang et al., 2018; Minh et al., 2020) under the best-fit model GTR+F+I+R2 with default parameters.

Results

Organization and chloroplast genomic features of *D. polypodioides*

Using HiSeq 2500 sequencing, we obtained approximately 9.67 GB of data, which included 28.83 million reads for *D. polypodioides*. The de novo assembled cp genome revealed a coverage depth of 323X for the selected species. The cp genome of *D. polypodioides* was 152,009 bp long and included two inverted repeat regions (IRa and IRb), each 23,659 bp long. These inverted repeats were separated by a large single-copy region (LSC) of 82,389 bp and a small single-copy region (SSC) of 22,303 bp, forming a quadripartite structure (Figure 1, Table 1).

The cp genome size varied among the species, ranging from 150,179 bp in *D. dilatatum* to 152,009 bp in *D. polypodioides*. The LSC regions ranged

from 81,984 bp in *D. striatum* to 82,559 bp in *D. pycnosora*, the SSC regions ranged from 21,500 bp in *D. viridifrons* to 22,302 bp in *D. polypodioides*, and the IR regions ranged from 22,969 bp in *D. dushanense* to 23,659 bp in *D. polypodioides* (Table 1). These genomes contained 116 genes whose arrangement was highly similar across the genomes, without any long inversions (Figure 2).

The gene contents of *D. polypodioides* were similar to those of other species in the family, as shown in Table 2 and Figure 1. The cp genome contained 117 genes, including 84 protein-coding genes, 29 tRNA genes, and 4 ribosomal RNA genes (Table 1). Fifteen of these genes were duplicated in the IR regions, including 4 rRNA, 6 tRNA, and 4 protein-coding genes, except the truncated *ycf1*$^{\Psi}$ gene. Seventeen genes contained introns, comprising six tRNA genes and eleven protein-coding genes. The *rps12* gene was trans-spliced, with its 5′ region in the LSC region and its 3′ region in the IR region, resulting in the duplication of the 3′ region. Among the genes with introns, seventeen had one intron each, whereas two genes (*clp*P and *ycf*3) contained two introns (Table 2). Additionally, three protein-coding genes and three tRNA genes with introns were duplicated in the IR regions (Table 2).

Many RNA editing events have been observed in species of Athyriaceae, including *D. polypodioides*. In *D. polypodioides*, 71 RNA editing events were identified in the complete cp genome, with 54 U-to-C and 17 C-to-U conversions.

The cp genome of *D. polypodioides* was mostly AT-rich. A comparison of the GC contents among the species revealed a range from 43.0% in *D. bellum* to 44.6% in *A. anisopterum*. The GC contents in different regions varied among species, as shown in Table 1. Compared with the LSC (42.6% in *D. dushanense* to 43.6% in *A. anisopterum*) and SSC regions (39.7% in *D. bellum* to 42% in *D. pycnosora*), the IR regions presented the highest GC content, ranging from 45.1% in *D. bellum* to 45.9% in *D. lancea*. The high GC content in IRs was attributed to GC-rich genes, such as rRNAs (55.1% in *A. opacum* to 55.7% in *D. bellum*) and tRNAs (55.0% in *D. polypodioides* to 55.7% in *D. bellum*) (Table 1).

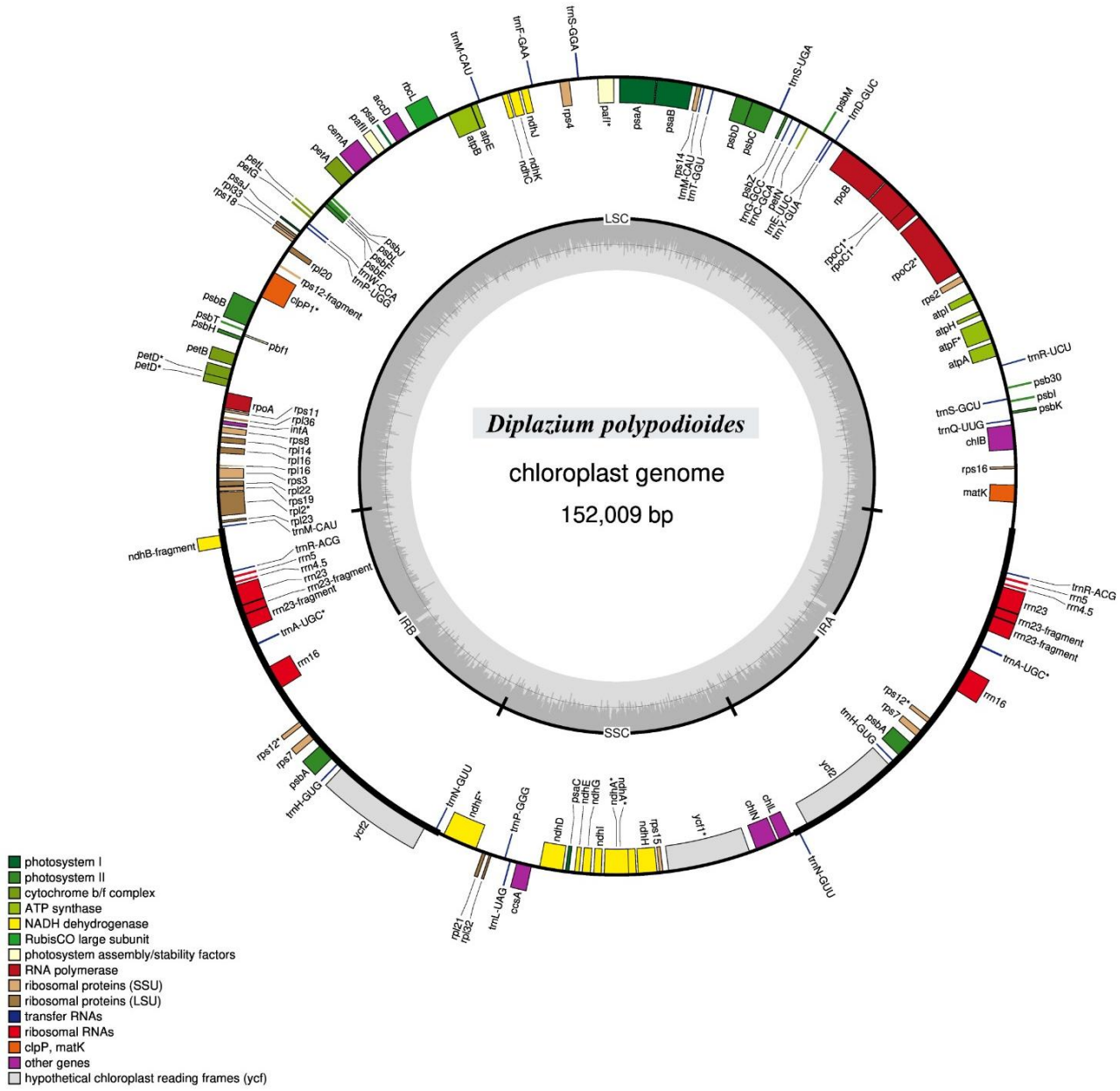

Figure-1. The chloroplast genome of *D. polypodioides* is depicted in a circular format as representative of all species. Genes transcribed in the counterclockwise direction are shown on the inner side of the circle, whereas those transcribed in the clockwise direction are on the outer side. The large single copy (LSC), small single copy (SSC), and inverted repeat (IRA and IRB) regions are shown in the circle. The GC content is indicated by the dashed gray color in the inner circle, whereas the AT content is highlighted in a lighter gray.

Table-1. Comparison of different features of the chloroplast genome among different species of the family Athyriaceae

Characteristic		*Diplazium polypodioides* Blume	*Diplazium striatum* Desv.	*Diplazium dushanense* (Ching ex W.M.Chu & Z.R.He) R.Wei & X.C.Zhang	*Diplazium maximum* (D.Don) C.Chr.	*Diplazium bellum* (C.B.Clarke) Bir	*Diplazium dilatatum* Blume	*Athyrium anisopterum* Christ	*Athyrium sinense* (Baker) C.Chr.	*Athyrium sheareri* (Baker) Ching	*Athyrium opacum* (D.Don) Copel.	*Deparia lancea* (Thunb.) Fraser-Jenk.	*Deparia pycnosora* (Christ) M.Kato	*Deparia viridifrons* (Makino) M.Kato
Size (bp)		152,009	150,797	150,179	150,984	151,601	151,114	151,284	151,319	151,068	150,979	151,011	151,126	150,939
LSC length (bp)		82,389	81,984	82,412	82,293	82,483	82,411	82,514	82,450	82,508	82,343	82,258	82,559	82,547
SSC length (bp)		22,302	21,793	21,829	21,765	21,790	21,785	21,694	21,707	21,736	21,700	21,773	21,693	21,500
IR length (bp)		23,659	23,501	22,969	23,463	23,664	23,459	23,538	23,581	23,421	23,468	23,490	23,473	23,446
Number of unique genes		117	117	117	117	117	117	117	117	117	117	117	117	117
Protein-coding genes		84	84	84	84	84	84	84	84	84	84	84	84	84
tRNA genes		29	29	29	29	29	29	29	29	29	29	29	29	29
rRNA genes		4	4	4	4	4	4	4	4	4	4	4	4	4
Duplicated Genes in IR		15	15	15	15	15	15	15	15	15	15	15	15	15
GC content (%)	Total	43.7	43.9	43.2	43.9	43.0	43.9	44.6	43.9	43.5	43.7	43.9	44.1	43.9
	LSC	43.3	43.4	42.6	43.5	42.0	43.4	43.6	43.4	43.4	43.2	43.2	43.5	43.2
	SSC	41.5	41.4	40.4	41.8	39.7	41.8	41.8	41.4	41.2	41.4	41.7	42.0	41.7
	IR	45.3	45.7	45.5	45.7	45.1	45.8	45.7	45.5	45.5	45.7	45.9	45.8	45.7
	CDS	43.3	43.7	43.4	43.9	43.0	43.9	44.1	43.9	43.8	43.8	43.9	44.1	43.9
	rRNA	55.2	55.3	55.2	55.3	55.7	55.2	55.2	55.2	55.5	55.1	55.3	55.2	55.2
	tRNA	55.0	55.5	55.5	55.4	55.7	55.3	55.2	55.1	55.1	55.5	55.4	55.5	55.6
Accession Number		**PP146611**	**NC_035852.1**	**NC_035851.1**	**MN623359.1**	**NC_035849.1**	**NC_035850.1**	**NC_035738.1**	**NC_035839.1**	**NC_035836.1**	**NC_035841.1**	**NC_035844.1**	**NC_035845.1**	**NC_035846.1**

Table 2 Gene content and functional classification of the chloroplast genomes of *D. polypodioides*

The genes labeled in the table represent genes with two copies, * genes with one intron, ** genes with two introns

Category for genes	**Group of genes**	**Name of Gene**									**No of genes**
Photosynthesis related genes	Photosystem Ⅰ	*psa*A	*psa*B	*psa*I	*psa*J	*psa*C					5
	Photosystem Ⅱ	*psb*K	*psb*I	*psb*M	*psb*Z	*psb*C	*psb*D	*psb*J	*psb*L	*psb*F	16
		*psb*E	*psb*B	*psb*T	*psb*N	*psb*H	*psb*A [a]				
	Cytochrome b/f complex	*pet*N	*pet*A	*pet*L	*pet*G	*pet*B*	*pet*D*				6
	ATP synthase	*atp*A	*atp*F*	*atp*H	*atp*I	*atp*E	*atp*B				6
	NADPH dehydrogenase	*ndh*B	*ndh*J	*ndh*K	*ndh*C	*ndh*F	*ndh*D	*ndh*E	*ndh*G	*ndh*I	11
		*ndh*A*	*ndh*H								
	Rubisco & Cytochrome synthesis	*ccs*A	*rbc*L								2
Transcription and translation related genes	Transcription small units of ribosome	*rpo*C1*	*rpo*C2	*rpo*B	*rpo*A	*rps*2	*rps*4	*rps*14	*rps*12* [a]	*rps*16*	18
		*rps*18	*rps*8	*rps*15	*rps*19	*rps*7 [a]	*rps*3	*rps11*			
	Large subunit of ribosome	*rpl*33	*rpl*20	*rpl*36	*rpl*14	*rpl*16*	*rpl*22	*rpl*2*	*rpl*23	*rpl*21	10
		*rpl*32									
	Ribosomal tRNA	*rrn*4.5[a]	*rrn*5[a]	*rrn*16[a]	*rrn*23[a]						8
	Transfer RNA	*trn*Q-*UUG*	*trn*G-*UCC**	*trn*R-*UCU*	*trn*S-GCU[a]	*trn*D-GUC	*trn*Y-GUA	*trn*E-UUC* [a]	*trn*C-GCA	*trn*G-GCC	35
		*trn*S-UGA	*trn*T-GGT	*trn*M-CAU[a]	*trn*S-GGA	*trn*L-UAA*	*trn*F-GAA	*trn*V-UAC*	*trn*R-UCG	trnL-CAA	
		*trn*W-CAU	*trn*N-GUU[a]	*trn*I-CAU	*trn*T-UGU[a,*]	*trn*R-ACG[a]	*trn*P-UGU	*trn*H-GUU[a]	*trn*I-GAU[a,*]	*trn*A-UGC*,[a]	
		*trn*L-UAG	trnR-ACG								
Other genes	RNA processing and translation factor	*mat*K	*inf*A	*paf*L*							3
	Carbon, Fatty acids metabolism and Proteolysis	*cem*A	*clp*P*	*acc*D							3
	Hypothetical proteins	*ycf*12	*ycf*3*	*ycf*4	*ycf*2[a]	*ycf*1					5
	Other genes	*chl*B	*chl*L	*chl*N							3
Total											131

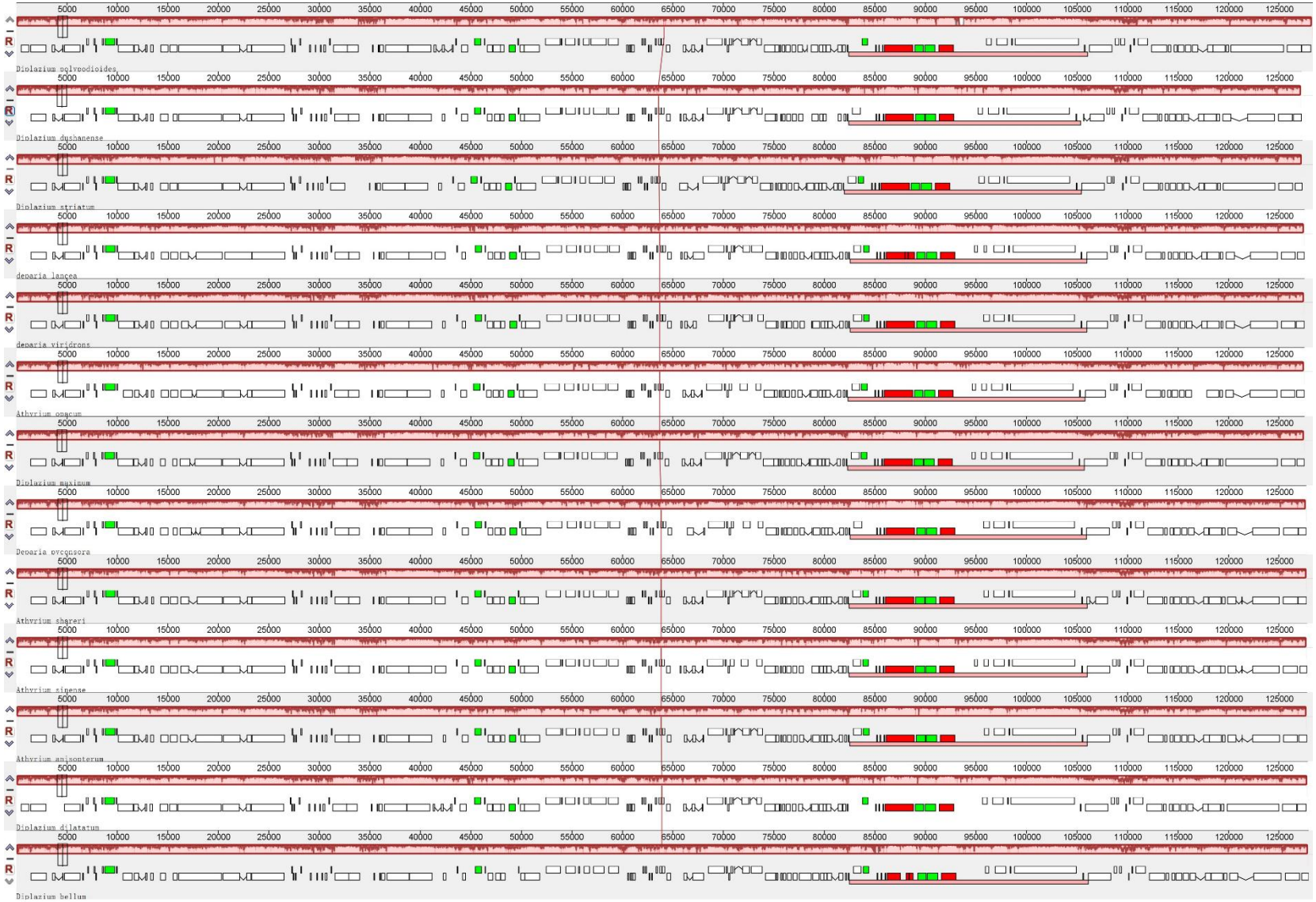

Figure-2. Mauve alignment of thirteen species of the family Athyriaceae revealed close resemblance in terms of gene content and arrangement. The small blocks represent genes. Red = rRNA; green = tRNA with introns; black = tRNA without introns; white = protein-coding genes.

Contraction and expansion of inverted repeats

Contractions and expansions in the IR regions can alter genome size and gene arrangement, leading to genetic diversity among species, genera, and families. In the family Athyriaceae, the patterns of contraction and expansion at the LSC/IR and SSC/IR junctions were highly similar across the selected species, with a few exceptions (Figure 3). The locations and sizes of essential genes, such as *rpl*2, *ndh*F, *trn*N, and *mat*K, were consistent across the cp genomes of the selected species. Notably, the *ndh*F gene, located at the IRB and SSC junction (JSB), exhibited expansions of 23-36 bp related to photosynthesis in all species except *D. dushanense*, *A. sheareri*, and *A. opacum*. In *D. dushanense*, a significant portion of the *ndh*F gene was deleted from the IR and SSC regions, leaving only a 262-base pair fragment in the middle of the SSC region. In *A. sheareri* and *A. opacum*, the *ndh*F gene was absent from the IR regions and was found only in the SSC region. The *ndh*B gene was present in the IRB region across all selected species, with a portion also located in the LSC region of *D. dilatatum*. The *chl*L gene was also positioned at the JSA, with expansions into the IRA ranging from 3-45 bp in the selected species.

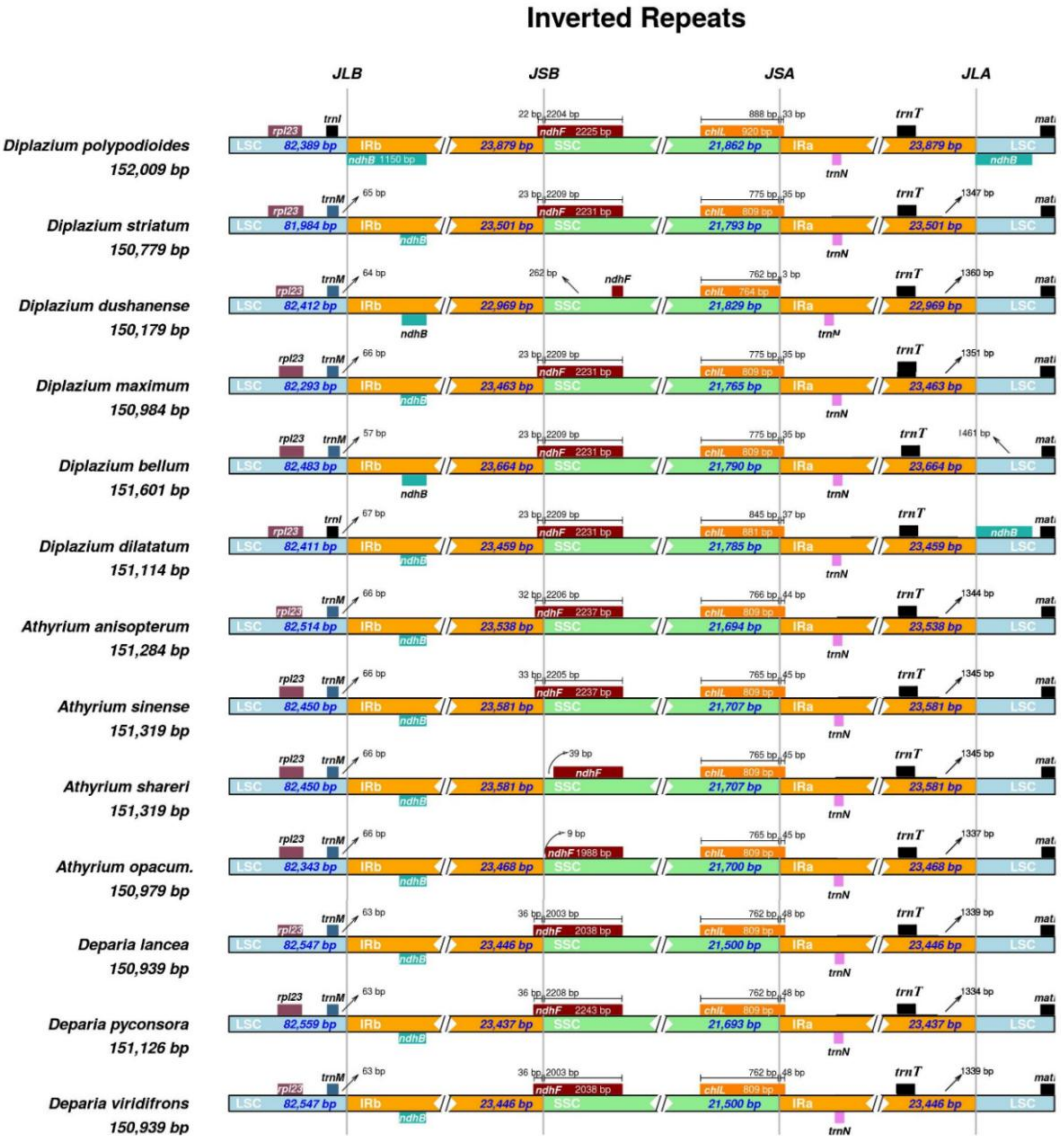

Figure-3. Inverted repeats contraction and expansion at the junction sites of the cp genome. JLB: LSC/IRb, JSB: IRb/SSC; JSA: SSC/IRa; JLA: IRa/LSC. The arrows indicate the distance of genes from the junction site, as shown for *trn*M at JLB and for *trn*T at JLA. The scale bar above some genes shows the base pairs that each gene contains in specified regions of the genome.

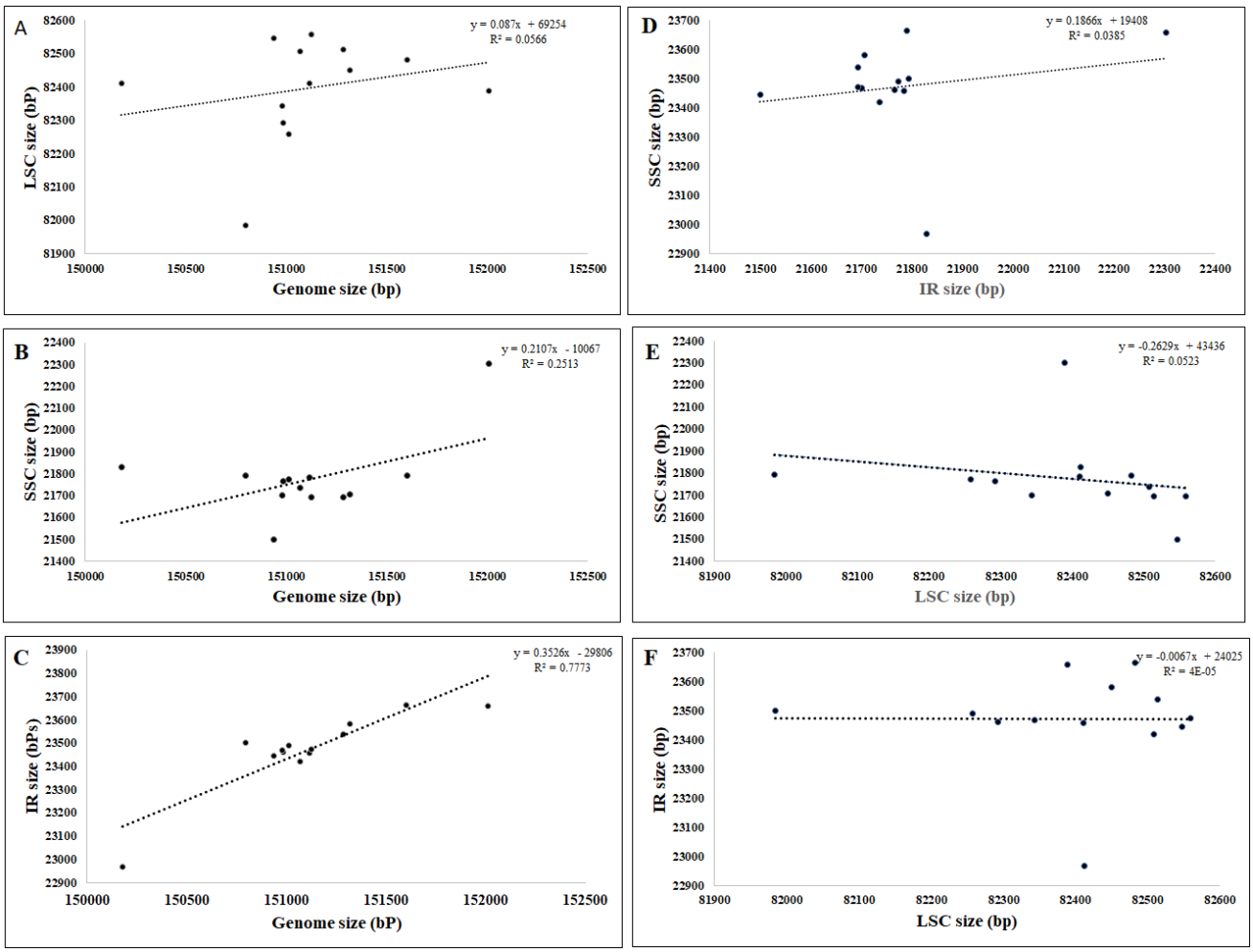

Figure-4. Correlation between complete genome size and different regions of the genome. A = Genome size vs. LSC, B = Genome size vs. SSC, C = Genome size vs. IR, D = IR vs. SSC, E = LSC vs. SSC, F = LSC vs. IR.

Correlation between genomic size and its three main regions

The correlation between genomic size (bp) and the sizes of specific regions of the cp genome, including the large single copy (LSC), small single copy (SSC), and inverted repeats (IRs), was analyzed, as shown in Figure 4. A positive correlation was observed between the overall genome size and the sizes of all these regions. Furthermore, a positive correlation was also found between the IR and SSC regions. Conversely, a negative correlation was detected between the sizes of the LSC and SSC regions, as well as between the IR and SSC regions across all selected species.

Codon Usage and Amino acid frequency

Relative synonymous codon usage (RSCU) values were used to analyze codon usage patterns. Moreover, the percentage frequency of amino acids was used to identify the predominant amino acids in selected species of the Athyriaceae family. The results indicated that the most common codons encode different amino acids, predominantly ending with A or T at the 3′ position, rather than C or G (Figure 5).

Leucine emerged as the most abundant amino acid across all the selected species and was encoded primarily by the TTA and TTG codons, followed by serine. Conversely, cystine was the least prevalent amino acid, as illustrated in Figure 5.

Analyses of simple sequence repeats and oligonucleotide repeats

Simple sequence repeats (SSRs) and oligonucleotide repeats were analyzed in the cp genomes of selected species from the Athyriaceae family. SSRs were categorized based on motif type: mononucleotide, dinucleotide, trinucleotide, tetranucleotide,

pentanucleotide, and hexanucleotide, as illustrated in Figure 6(A-D). Across all the species, the number of repeats ranged from 51-101, with mononucleotide SSRs being the most prevalent. Similarly, dinucleotide, trinucleotide, and tetranucleotide repeats appeared consistently across all species, albeit in varying quantities. In contrast, pentanucleotide and hexanucleotide repeats were present in some species but absent in others, as shown in Figure 6(A). Most SSRs were in the LSC region, followed by the IR region, with the fewest in the SSC region, as depicted in Figure 6(B). The SSR count ranged from 51 in *A. anisopterum* to 101 in *D. bellum.*

In addition, oligonucleotide repeats were analyzed via the REPuter Bibi server. *A. sinense* exhibited the minimum number of repeats (7), whereas *D. polypodioides* had the greatest number of repeats (387). The identified oligonucleotide repeats included palindromic, forward, reverse, and complementary types, which were located primarily in the intergenic spacer (IGS) regions of all the selected genomes, as shown in Figure 6C. The IR region contained the greatest number of repeats among the cp regions, followed by the LSC and SSC regions. The fewest repeats were shared between LSC/SSC, IR/SSC, and LSC/IR, as illustrated in Figure 6(D).

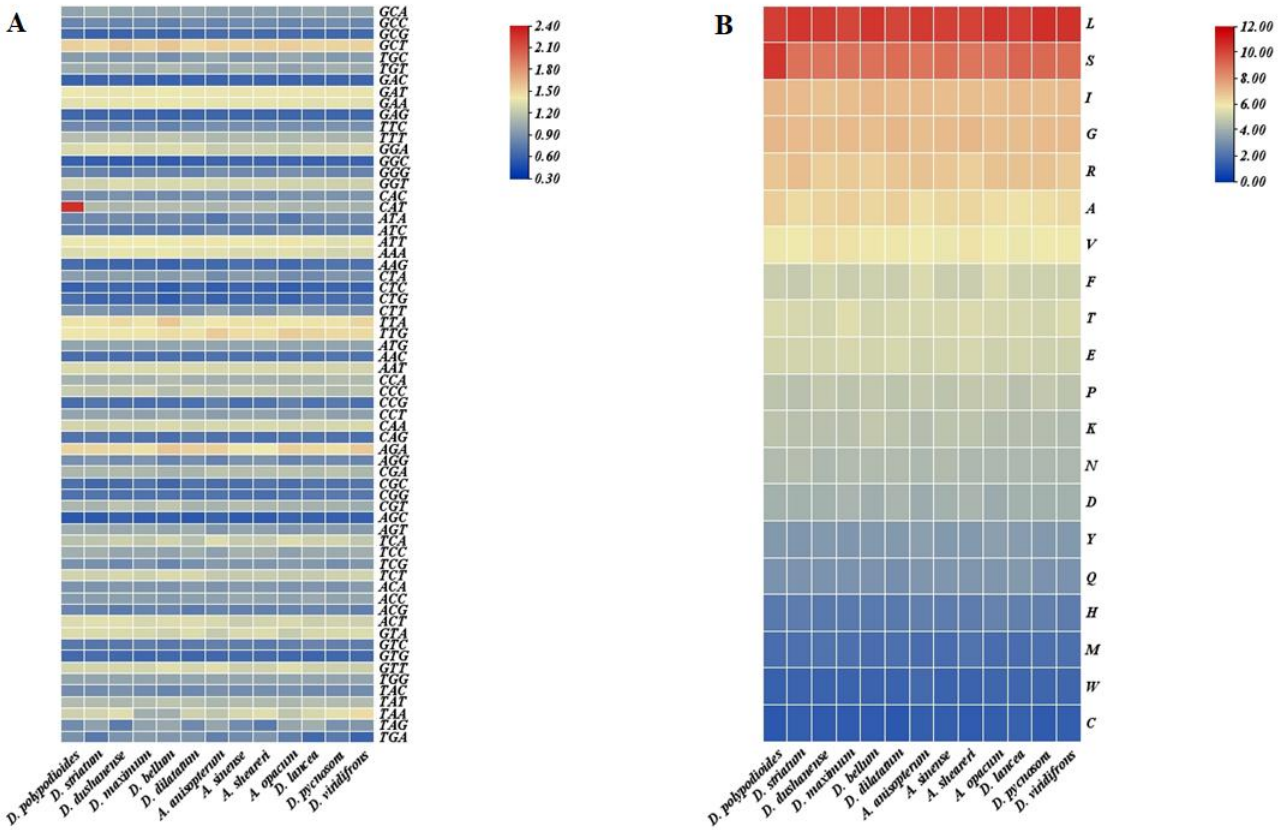

Figure-5. Amino acid frequency and codon usage among the selected species of the Athyriaceae family. A= Codon usage (RSCU values) B = Amino acid frequency

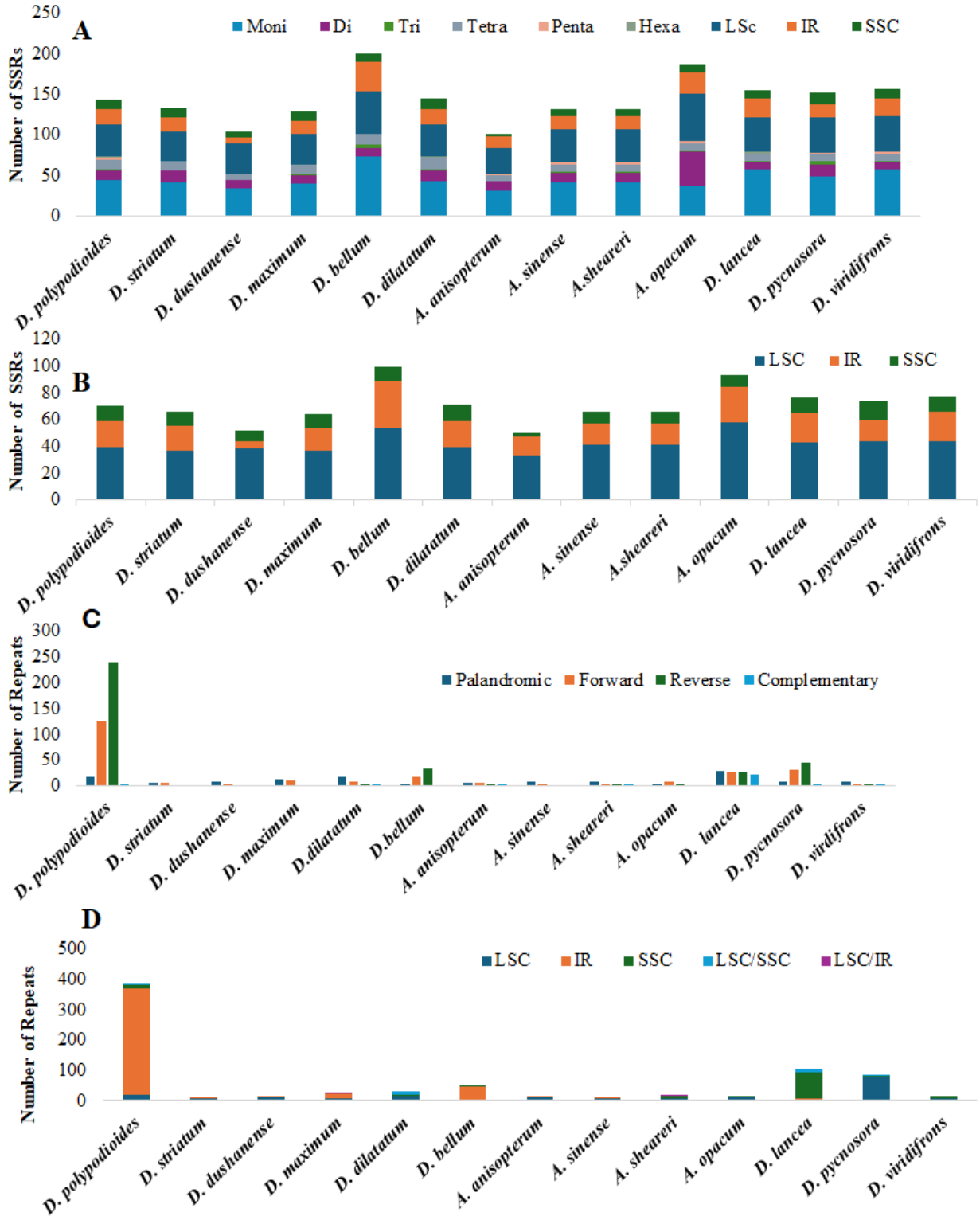

Figure-6. Simple Sequence Repeats and Oligonucleotide Repeats in selected species of the family Athyriaceae. A= Types of SSRs B= Regionwise SSRs; C= Types of oligonucleotides; D = Regionwise oligonucleotides.

Analysis of substitutions and Indels

Substitutions and InDels were analyzed across the complete cp genome, three main regions (LSC, SSC, and IR), and coding regions of various species. Transversion events predominated over transition events across all species, with a Ts/Tv ratio ranging from 0.12 to 0.17. *D. viridifrons* exhibited the greatest number of substitutions (9568), whereas *A. sinense* showed the lowest number of substitutions (1963). Within the genera, the species presented very similar Ts/Tv ratios. Substitutional hotspots were predominantly located in the LSC regions, followed by the SSC regions, and least common in the IR regions. With respect to SNPs in coding regions, *D. lancea* had the highest count at 4195, followed by *D. pycnosora* at 3934, whereas *A. sinense* had the lowest count at 813. Similarly, InDels in coding regions were most prevalent in *D. maximum* and *D. polypodioides*, with *A. sinense* showing the fewest InDels.

Phylogenetic analysis

Thirteen species from three genera of the Athyriaceae family were selected for phylogenetic analysis via IQ-tree 2.0, with the GTR+F+I+R2 model applied to the coding sequences of 84 protein-coding genes. The multiple sequence alignments covered 56,990 nucleotide sites after the gaps were removed. Among the nucleotide sites, 5,422 sites were parsimony informative, and 1650 exhibited distinct site patterns. The analysis revealed the monophyly of three genera, namely, *Diplazium*, *Deparia*, and *Athyrium* (Figure 8). *Diplazium* species showed close resemblance to *Athyrium* instead of *Deparia*. The species within the genus *Diplazium* are divided into two primary subclades. In the first subclade, *D. dilatatum* and *D. maximum* form a sister group rooted by *D. polypodioides*, which is further rooted by *D. striatum* (Figure 8). In the second subclade, *D. bellum* and *D. dushanense* are grouped as sister species.

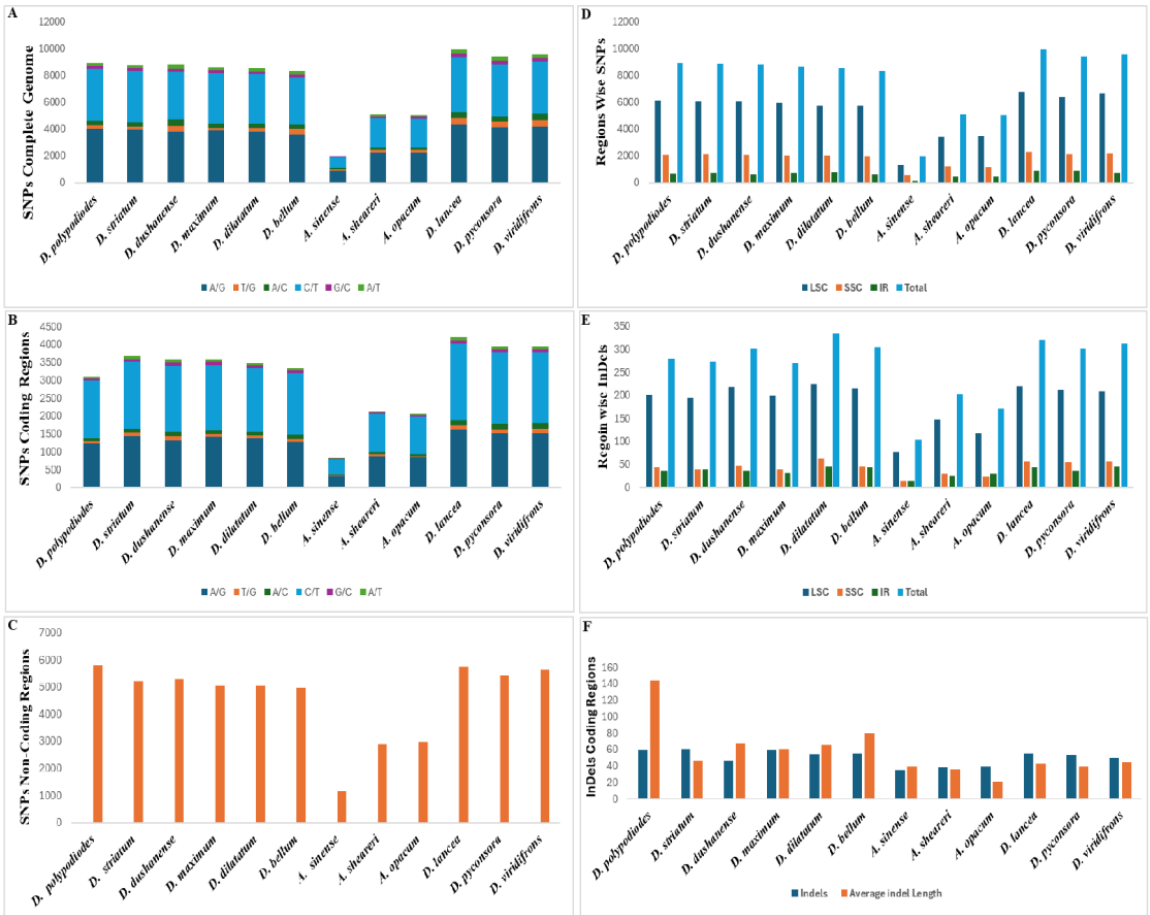

Figure-7. Substitution and Indel mutations in the chloroplast genomes of 13 species. A = SNPs in the complete genome; B = SNPs in coding regions; C = SNPs in noncoding regions; D = SNPs in different regions of the genome; E = Indels in different regions of the genome; F = InDels in coding regions.

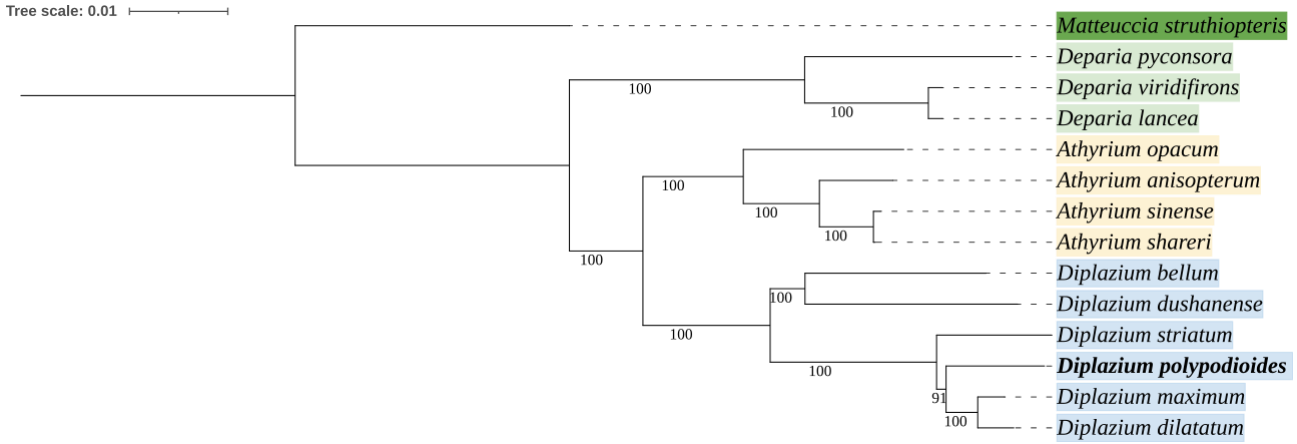

Figure-8. Phylogenetic relationships among 13 species of the family Athyriaceae on the basis of the coding sequences of 84 protein-coding genes.

Discussion

The cp genome is a valuable molecule for phylogenetic studies because of its slow evolution and single-parent inheritance. This study provides insight into evolutionary changes that establish genomic diversity among closely related species. This study compares the cp genome of *D. polypodioides* (newly sequenced) with those of twelve previously reported species of the family Athyriaceae. One major issue in data comparison among family members is error in genome annotation (Amiryousefi et al., 2018; Abdullah et al., 2020b, 2021a; Li et al., 2024). After annotation correction, these genomes were highly similar. Here, we provide insights into the cp structure, IR contraction and expansion, SSR and oligonucleotide repeats, substitution events and their types.

The typical size of fern cp genomes ranges from 131-168 kb (Loiseau et al., 2020), and the size of *D. polypodioides* falls within this range. The cp genome of *D. polypodioides* includes four junctions: LSC, SSC, and IR. Genome size variation among species is due mainly to contractions and expansions in the IR and SSC regions (Li et al., 2024).

The key evolutionary events in the cp genome include intron gain and loss, IR contraction and expansion, SSR, oligonucleotide repeat, substitution, and RNA editing. The cp genomes of all the species closely resemble each other in structure, as they are similar in terms of genes content and their arrangement (Henriquez et al., 2020a; Abdullah et al., 2021a; Kwon et al., 2023). The contraction and expansion of the cp genomes of pteridophytes and angiosperms have been noted previously (Jamal et al., 2021; Rehman et al., 2021). This study revealed *chl*L and *ndh*F at the SSC/IRa and IRb/SSC boundaries, respectively, covering some IR regions in the selected species. This type of travel of genes from single copies to the IR or vice versa leads to a rate of heterotachy (Lockhart et al., 2006; Zhu et al., 2016; Abdullah et al., 2020a). c

SSRs are important molecular genetic markers for population studies. An unusual pattern in SSRs was observed in the selected species of Athyriaceae. All six types of nucleotides were detected, with mono- and dinucleotides being predominant. Hexa-nucleotides were present in some species but absent in others. Mononucleotides (594) were the most abundant, followed by dinucleotides (177). Previous studies reported heterogeneous divergence in SSR markers of higher plants (Wang et al., 2018). Most SSRs detected in this study were GC rich, which aligns with previous reports in ferns (Gao et al., 2018). However, *D. bellum* and *D. dushanense* contained more AT-rich SSRs, similar to some angiosperms (Fan et al., 2019; Mehmood et al., 2020).

Oligonucleotide repeats, which are widespread in the cp genome, generate mutations, including substitutions and InDels (Abdullah et al., 2020c; Abdullah et al., 2021b). These repeats are critical in substitution and InDel generation. The identified oligonucleotide repeats in this study were forward, reverse, palindromic, and complementary in all selected species of Athyriaceae, mainly within the IGS regions, as previously reported in angiosperms (Abdullah et al., 2019; Xu and Wang, 2021). *Diplazium polypodioides* had the maximum number of oligonucleotide repeats (370), surpassing

previously reported numbers in *D. fragrans* (204) (Gao et al., 2018).

Substitutions and InDels were abundant in the LSC region of the genome and noncoding regions of all the studied species, and *A. anisopterum* was used as a reference. These findings align with previous studies reporting more InDels and substitutions in the LSC and SSC regions than in the IR regions (Abdullah et al., 2019; Liu et al., 2020; He et al., 2022). Most substitutions and InDels were in the complete genome of the genus *Deparia*, such as in *D. lancea* and *D. pycnosora*, which presented 9911 and 9352 substitution events, respectively. Among the noncoding regions, the greatest number of substitutions was observed in *D. polypodioides* (5805), followed by the genus *Deparia*, with the lowest number in *A. sinense* (1150). Similarly, *D. lancea* (320) and *D. pycnosora* (302) had the most InDels. The Ts/Tv ratio was >1, which is consistent with the cp genomes of angiosperms and gymnosperms (Kim et al., 2015; He et al., 2022).

Here, our phylogenetic analysis of the limited species set revealed that the genera *Diplazium*, *Athyrium*, and *Deparia* form monophyletic groups. Moreover, the internal branches of the phylogeny displayed strong bootstrap support, indicating high confidence in the inferred relationships. However, earlier studies reported that *Athyrium* is not monophyletic, and while the backbone of *Diplazium*'s phylogeny was resolved with robust support, significant incongruence across different datasets and weakly to moderately supported internal relationships were observed (Wei et al., 2018; Wei and Zhang, 2020).

For *Athyrium*, two possible explanations for the observed monophyly in our results may be considered. First, the high resolution provided by the cp genome, compared with the eight loci used in earlier studies (Wei et al., 2018), could have resolved relationships more clearly. Second, our phylogeny included a limited number of species, and the absence of previously identified paraphyletic species might have contributed to the apparent monophyly. Thus, the current findings are likely constrained by sampling, and future studies using a larger, more comprehensive dataset are essential to validate these results.

Our sample size of *Diplazium* is relatively small compared with that of Wei and Zhang (2020), limiting the scope of our conclusions. We concur with their recommendation that future phylogenetic analyses should incorporate broader taxon sampling and include nuclear genome markers, utilizing high-throughput sequencing approaches. Such comprehensive studies would be critical to resolving the phylogeny of *Diplazium* more conclusively.

Conclusion

Our study provides insights into the first-ever sequenced cp genome of *D. polypodioides*. The family members closely resembled each other in terms of gene content, with minor differences in gene size and arrangement. Amino acid frequency, codon usage, and GC content were consistent across the famiy. Phylogenetic analysis revealed that members of *Diplazium* are more closely related to *Athyrium* than to *Deparia* and confirmed the monophyletic nature of the genus *Diplazium*. Sequencing a wider range of species in the Athyriaceae family could aid in reconstructing a high-resolution phylogenetic tree and offer deeper insights into systematics. Identifying SSRs and oligonucleotides will be valuable for developing novel molecular markers for molecular identification, assessing diversity, and understanding population genetics and phylogeny.

Acknowledgment

The authors acknowledge the financial support provided by the International Foundation of Science [Project No: GRANT 1-3-F-6135-1] and the National Research Program for the University by the Higher Education Commission [NRPU-14667]. The authors also acknowledge Alpha Genomics, Islamabad, for their technical and instrumental support.

Disclaimer: None.

Conflict of Interest: None.

Source of Funding: This study was financially supported by the International Foundation of Science [Project No: GRANT 1-3-F-6135-1] and the National Research Program for University by the Higher Education Commission [NRPU-14667].

Contribution of Authors

Yunus A: Performed all the experimental work, collected and analyzed the data and wrote first draft of the manuscript

Sultana N: Supervised the entire project, developed the study design, and edited and revised the manuscript

Gul A: Co-supervised the study and supported

research from funded projects
Abdullah: Performed all the experimental work, collected and analyzed the data and wrote first draft of the manuscript

References

Abdullah, Shahzadi I, Mehmood F, Ali Z, Malik MS, Waseem S, Mirza B, Ahmed I and Waheed MT, 2019. Comparative analyses of chloroplast genomes among three *Firmiana* species: Identification of mutational hotspots and phylogenetic relationship with other species of Malvaceae. Plant Gene. 19: 100199. https://doi.org/10.1016/j.plgene.2019.100199.

Abdullah, Henriquez CL, Mehmood F. Carlsen MM., Islam M, Waheed MT, Poczai P, Croat TB and Ahmed I, 2020a. Complete chloroplast genomes of *Anthurium huixtlense* and *Pothos scandens* (Pothoideae, Araceae): unique inverted repeat expansion and contraction affect rate of evolution. J. Mol. Evol. 88: 562–674. https://doi.org/10.1101/2020.03.11.987859

Abdullah, Waseem S, Mirza B, Ahmed I and Waheed, M.T., 2020b. Comparative analyses of chloroplast genomes of *Theobroma cacao* and *Theobroma grandiflorum*. Biologia (Bratisl). 75, 761–771. https://doi.org/10.2478/s11756-019-00388-8

Abdullah, Mehmood F, Shahzadi I. Waseem S, Mirza B, Ahmed I and Waheed MT, 2020c. Chloroplast genome of *Hibiscus rosa-sinensis* (Malvaceae): comparative analyses and identification of mutational hotspots. Genom. 112: 581–591. https://doi.org/10.1016/j.ygeno.2019.04.010

Abdullah, Mehmood F, Rahim A, Heidari P, Ahmed I and Poczai P, 2021a. Comparative plastome analysis of *Blumea*, with implications for genome evolution and phylogeny of Asteroideae. Ecol. Evol. 11: 7810–7826. https://doi.org/10.1002/ece3.7614

Abdullah, Mehmood F, Shahzadi I, Ali Z, Islam M, Naeem M, Mirza B, Lockhart P, Ahmed I, and Waheed MT, 2021b. Correlations among oligonucleotide repeats, nucleotide substitutions and insertion- deletion mutations in chloroplast genomes of plant family Malvaceae. J. Syst. Evol, 59: 388–402. https://doi.org/10.1111/jse.12585

Ahmed I, Lockhart PJ, Agoo EMG, Naing KW, Nguyen DV, Medhi DK, and Matthews PJ, 2020. Evolutionary origins of taro (*Colocasia esculenta*) in Southeast Asia. Ecol. Evol. 1–14. https://doi.org/10.1002/ece3.6958

Amiryousefi A, Hyvönen J and Poczai P, 2018. The chloroplast genome sequence of bittersweet (*Solanum dulcamara*): Plastid genome structure evolution in Solanaceae. PLoS One. 13: 1–23. https://doi.org/10.1371/journal.pone.0196069

Amiryousefi A, Hyvönen J, Poczai P, 2018. IRscope: an online program to visualize the junction sites of chloroplast genomes. Bioinformatics 34: 3030–3031. https://doi.org/10.1093/bioinformatics/bty220

Beier S, Thiel T, Münch T, Scholz U and Mascher M, 2017. MISA-web: A web server for microsatellite prediction. Bioinform. 33: 2583–2585. https://doi.org/10.1093/bioinformatics/btx198

Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, and Xia R, 2020. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. Mol. Plant. 13: 1194–1202. https://doi.org/10.1016/j.molp.2020.06.009

Daniell H, 2007. Transgene containment by maternal inheritance: Effective or elusive? Proc. Natl. Acad. Sci. https://doi.org/10.1073/pnas.0702219104

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

Darling ACE, Mau B, Blattner FR. and Perna NT, 2004. Mauve: Multiple Alignment of Conserved Genomic Sequence with Rearrangements. Genome Res. 14: 1394–1403. https://doi.org/10.1101/gr.2289704

Dierckxsens N, Mardulyn P and Smits G, 2017. NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data. Nucleic Acids Res, 45(4): e18-e18. https://doi.org/10.1093/nar/gkw955

Fan R, Ma W, Liu S, and Huang Q, 2019. Integrated analysis of four new sequenced fern Chloroplast Genomes: Genome structure and comparative analysis. doi: 10.21203/rs.2.18299/v1

Gao R, Wang W, Huang Q, Fan R, Wang X, Feng P, Zhao G, Bian S, Ren H, and Chang Y, 2018. Complete chloroplast genome sequence of *Dryopteris fragrans* (L.) Schott and the repeat structures against the thermal environment. Sci.

rep. 8(1): 16635. https://doi.org/10.1038/s41598-018-35061-8

He X, Dong S, Gao C, Wang Q, Zhou M, and Cheng R, 2022. The complete chloroplast genome of *Carpesium abrotanoides* L. (Asteraceae): structural organization, comparative analysis, mutational hotspots, and phylogenetic implications within the tribe Inuleae. Biologia. 77(7): 1861-1876. https://doi.org/10.1007/s11756-022-01038-2

Henriquez CL, Abdullah, Ahmed I, Carlsen MM, Zuluaga A, Croat TB, and McKain MR, 2020a. Molecular evolution of chloroplast genomes in Monsteroideae (Araceae). Planta 251: 72. https://doi.org/10.1007/s00425-020-03365-7

Henriquez CL, Abdullah, Ahmed I, Carlsen MM, Zuluaga A, Croat TB, and McKain MR, 2020b. Evolutionary dynamics of chloroplast genomes in subfamily Aroideae (Araceae). Genomics 212(3): 2349-2360. https://doi.org/10.1007/s00425-020-03365-7

Hoang DT, Chernomor O, Von Haeseler A, Minh BQ and Vinh LS, 2018. UFBoot2: improving the ultrafast bootstrap approximation. Mol. Biol. Evol. 35(2): 518-522. https://doi.org/10.1093/molbev/msx281

Jamal A, Wen J, Ma ZY, Ahmed I, Abdullah, Chen LQ, Nei ZL, and Liu XQ, 2021. Comparative chloroplast genome analyses of the winter blooming eastern Asian endemic genus *Chimonanthus* (Calycanthaceae), with implications for its phylogeny and diversification. Front. Genet. https://doi.org/10.3389/FGENE.2021.709996

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

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, and Thierer T, 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinform. 28(12): 1647-1649. doi: 10.1093/bioinformatics/bts199

Kim K, Lee SC, Lee J, Lee HO, Joh HJ, Kim NH, Park HS, and Yang TJ, 2015. Comprehensive survey of genetic diversity in chloroplast genomes and 45S nrDNAs within *Panax ginseng* species. PloS one. 10(6): p.e0117159. https://doi.org/10.1371/journal.pone.0117159

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

Kwon SH, Kwon HY, Choi YI, and Shin H, 2023. Comprehensive Analysis of Chloroplast Genome of *Hibiscus sinosyriacus*: Evolutionary Studies in Related Species and Genera. Forests, 14: 2221. https://doi.org/10.3390/F14112221/S1

Langmead B, and Salzberg SL, 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods, 9: 357–359. https://doi.org/10.1038/nmeth.1923

Laslett D, and Canback B, 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res, 32, 11–16. doi: 10.1093/nar/gkh152

Lehwark P, and Greiner S, 2019. GB2sequin - A file converter preparing custom GenBank files for database submission. Genom. 111: 759–761. https://doi.org/10.1016/J.YGENO.2018.05.003

Li C, Cai C, Tao Y, Sun Z, Jiang M, Chen L, and Li J, 2021. Variation and evolution of the whole chloroplast genomes of *Fragaria* spp. (Rosaceae). Front. Plant Sci, 12: 754209. https://doi.org/10.3389/fpls.2021.754209

Li H, Zheng Q, Wang Y, Abdullah and Tian X., 2024. Unveiling the chloroplast genomes of *Capsella bursa-pastoris*, *Descurainia sophia*, and *Lepidium apetalum* (Brassicaceae). J. Asia-Pacific Biodivers, 17(2): 245-249. https://doi.org/10.1016/J.JAPB.2024.01.010

Liu Q, Li X, Li M, Xu W., Schwarzacher T. and Heslop-Harrison JS, 2020. Comparative chloroplast genome analyses of *Avena*: insights into evolutionary dynamics and phylogeny. BMC Plant Biol. 20: 1-20. https://doi.org/10.1186/s12870-020-02621-y

Lockhart P, Novis P, Milligan BG, Riden J, Rambaut A, and Larkum, T, 2006. Heterotachy and Tree Building: A Case Study with Plastids and Eubacteria. Mol. Biol. Evol. 23: 40–45. https://doi.org/10.1093/molbev/msj005

Loiseau O, Weigand A, Noben S, Rolland J, Silvestro D, Kessler M, Lehnert M, and Salamin N, 2020. Slowly but surely: gradual diversification and phenotypic evolution in the hyperdiverse tree fern family Cyatheaceae. Ann. Bot. 125(1): 93-103. https://doi.org/10.1093/aob/mcz145

Lowe TM, and Chan PP, 2016. tRNAscan-SE On-line: Integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res, 44: W54–W57. doi: 10.1093/nar/gkw413

Mehmood F, Abdullah, Shahzadi, I, Ahmed I, Waheed MT, and Mirza B, 2020. Characterization of *Withania somnifera* chloroplast genome and its comparison with other selected species of Solanaceae. Genom. 112: 1522–1530. doi:10.1016/j.ygeno.2019.08.024.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, Lanfear R, and Teeling E, 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol. Biol. Evol. 37: 1530–1534. https://doi.org/10.1093/MOLBEV/MSAA015

Neale DB, and Sederoff RR, 1989. Paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in *Loblolly pine*. Theor. Appl. Genet. 77: 212–216. https://doi.org/10.1007/BF00266189

Palmer JD, 1985. Comparative organization of chloroplast genomes. Annu. Rev. Genet. 19: 325–354. https://doi.org/10.1146/annurev.ge.19.120185.001545

POWO, 2024. "Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew. http://www.plantsoftheworldonline.org

Rehman U, Sultana N, Abdullah, Jamal A, Muzaffar M, and Poczai P, 2021. Comparative Chloroplast Genomics in Phyllanthaceae Species. Divers. 13: 403. https://doi.org/10.3390/d13090403

Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, and Sánchez-Gracia A, 2017. DnaSP 6: DNA sequence polymorphism analysis of large datasets. Mol. Biol. Evol. 34: 3299–3302. https://doi.org/10.1093/molbev/msx248

Semwal P, Painuli S, Painuli KM, Antika G, Tumer TB, Thapliyal A, Setzer WN, Martorell M, Alshehri MM, Taheri Y, Daştan SD, Ayatollahi SA, Petkoska AT, Sharifi-Rad J, Cho WC, 2021. *Diplazium esculentum* (Retz.) Sw.: Ethnomedicinal, Phytochemical, and Pharmacological Overview of the Himalayan Ferns. Oxidative Medicine and Cellular Longevity 2021: 1917890. https://doi.org/10.1155/2021/1917890

Schuettpelz E, Schneider H, Smith AR, Hovenkamp P, Prado I, and Rouhan G, 2016. A community–derived classification for extant lycophytes and ferns. J. Syst. Evol. 54 (6): 563–603. https://doi.org/10.1111/jse.12229

Shah SN, Ahmad M, Zafar M, Razzaq A, Malik K, Rashid N, Ullah F, Iqbal M, Zaman W, 2018. Foliar epidermal micromorphology and its taxonomic implications in some selected species of Athyriaceae. Microscopy Research and Technique 81: 902–913. https://doi.org/10.1002/JEMT.23055

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. https://doi.org/10.1093/nar/gkx391

Wang JH, Moore MJ, Wang H, Zhu ZX and Wang HF, 2021. Plastome evolution and phylogenetic relationships among Malvaceae subfamilies. Gene 765: 145103. https://doi.org/10.1016/J.GENE.2020.145103

Wang W, Chen S, and Zhang X, 2018. Whole-genome comparison reveals heterogeneous divergence and mutation hotspots in chloroplast genome of *Eucommia ulmoides* Oliver. Int. J. Mol. Sci. 19(4): 1037. doi: 10.3390/ijms19041037

Wei A and Zhang CZ, 2020. Phylogeny of *Diplazium* (Athyriaceae) revisited: Resolving the backbone relationships based on plastid genomes and phylogenetic tree space analysis. Mol. Phylogenetics Evol. 43:106699. https://doi.org/10.1016/j.ympev.2019.106699

Wei R, Ebihara A, Zhu YM, Zhao CF, Hennequin S, and Zhang XC, 2018. A total-evidence phylogeny of the lady fern genus *Athyrium Roth* (Athyriaceae) with a new infrageneric classification. Mol. Phylogenetics Evol. 119: 25-36. https://doi.org/10.1016/j.ympev.2017.10.019

Xu X, and Wang D, 2021. Comparative chloroplast genomics of Corydalis species (Papaveraceae): evolutionary perspectives on their unusual large-scale rearrangements. Front Plant Sci. 11: p.600354. https://doi.org/10.3389/fpls.2020.600354

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