Health Affairs ISSN - 0278-2715 Volume 13 ISSUE 10 page 521-549
Journal link: https://health-affairs.com/
Abstract Link: https://health-affairs.com/13-10-521-549/
October 29

Unraveling the Chloroplast Genome of *Lolium perenne*: Comparative Genomics, Evolution, and Phylogenetic Insights

[1]Zamarak Haider, [2]Shahjahan Shabbir Ahmed Rana, [3]Imran Ali Sani, [4]Dawood Shahid, [5]Tabeel Tariq Bashir Maqsood Ahmed, [6]Faheem Nawaz, [7]Isra Durrani, [8]Aqeel Ahmad, [9]Shazia Irfan, [10]Hina Ali Ahmed

[1]Department of Biotechnology, (FLS&I), Balochistan University of Information Technology, Engineering and Management Sciences (BUITEMS), Quetta
[2]Department of Environmental Sciences, (FLS&I), Balochistan University of Information Technology, Engineering and Management Sciences (BUITEMS), Quetta
[3]Department of Environmental Sciences, (FLS&I), Balochistan University of Information Technology, Engineering and Management Sciences (BUITEMS), Quetta
[4]Department of Biotechnology, (FLS&I), Balochistan University of Information Technology, Engineering and Management Sciences (BUITEMS)
[5]Department of Environmental Sciences, (FLS&I), Balochistan University of Information Technology, Engineering and Management Sciences (BUITEMS)
[6]Department of Environmental Sciences, (FLS&I), Balochistan University of Information Technology, Engineering and Management Sciences (BUITEMS)
[7]Plant Tissue Culture Laboratory, Director Agriculture Research Institute (ARI), Sariab, Quetta
[8]Department of Agronomy, University of Poonch Rawalakot, Pakistan
[9]Department of Botany, SBK Women's University, Brewery Road, Quetta, Pakistan
[10]Department of Zoology, SBK Women's University, Brewery Road, Quetta, Pakistan

.**Corresponding author email: Tabeel.aryan@gmail.com**

Abstract

Chloroplast genome sequences make a significant contribution to the study of plant evolutionary biology, taxonomy and breeding by not only revealing the genetic structure, the relationships among species, but also the molecular adaptation. Although Lolium perenne (perennial ryegrass) has an agricultural significance, but chloroplast genome information from arid areas such Quetta from Pakistan is still lacking. The objective of current study was to sequence, assemble and annotate the whole chloroplast genome of L. perenne collected from Quetta valley. Genomic DNA of high quality was isolated and sequenced on the BGISEQ-500 instrument following a modified CTAB protocol. The NOVO-Plasty was used for de novo genome assembly and the annotation was done using CPGAVAS, DOGMA, and tRNA scan-SE. The full-length assembled

genome was 135,210 bp, and the organization was typical of the quadripartite structure with the LSC, SSC, and a pair of IRs. A total of 128 genes were found: 76 protein- coding genes, 30 tRNA genes (one pseudogene), and 4 rRNA genes. The GC content was 38.3%. Molecular characterisation identified 53 single nucleotide polymorphisms (SNPs) and 58 insertions/deletions (indels), and phylogenetic analysis showed the Quetta accession clustering in the L. multiflorum group and clearly separated from Festuca pratensis. These results not only identify the genetic background of the sample as L. perenne, but also provide a new regional cp genome to the world cp genome resources. These findings increase our knowledge of chloroplast genome evolution in Poaceae and provide useful genomic resources for molecular breeding, phylogeny, and germplasm conservation of forage grasses in semi-arid regions.

1. Introduction

Fuzzy perennials ryegrass (*Lolium perenne L.*) is an important cool-season grass species of the family *Poaceae*, which has great economic and ecological significance. It is a cornerstone of temperate pasture systems because of the high nutritional value of its forage, its rapid establishment, and adaptability to a range of climatic and soil conditions. In country such as New Zealand, Australia, United Kingdom *L. perenne* supports livestock production on millions of hectares farmland including food security and has a role in soil conservation, C sequestration and sustainability of the agroecosystems (Kumawat, 2023). It is estimated to pasture about 60 million sheep and cattle on 7 million hectares. Due to its physiological robustness, potential for hybridization and short life cycle, sheepshead minnow has been considered a model species for Eco genomic and ecotoxicology studies (Cunningham, 2024).

The chloroplast genome (cp) of land plants is a maternally hereditary genome, which is highly conserved and functions in photosynthesis and other metabolic activities. Besides their use in functional studies, cp genomes provide a conserved molecular system for plant taxonomy, phylogeny, and molecular evolutionary studies. Its tetrameric organization (large single copy, LSC; small single copy, SSC; and a pair of inverted repeats, IRs) usually contains 110–130 unique genes, such as protein coding genes, transfer RNAs, and ribosomal RNAs. With the slower mutation rate and high gene conservation, the cp genome is very useful to clarify

relationship among different taxonomic ranks, such as intra-generic, inter-family comparison (Palmer, 2023).

Research on chloroplast DNA has undergone a revolution since the first chloroplast genome sequence in Nicotiana tabacum in 1986. The initial studies of cp genomes were dominated by the cost-inefficient labor-intensive Sanger sequencing, which drastically diminished throughput and scope. However, the emerging of next-generation sequencing (NGS) technology has changed cp genomics completely due to its high-throughput, low-cost, and quick sequence methods. These advances have made possible de novo assemblies of complete cp genomes, detection of polymorphisms including SNPs and indels, and comparative genomics in a range of taxonomic groups. Additionally, the development of software tools, for instance, NOVO-Plasty, DOGMA, and CPGAVAS, has made genome assembly, annotation, and visualization more efficient, which in turn has facilitated deeper genomic analyses (Zhong, 2020).

Despite large amount of cp genome sequences in most members of the Poaceae family, very few taxonomic and geographical constraints have been acquired. Some subfamilies such as Panicoideae and Oryzoideae are over-represented in GenBank, while others, like Chloridoideae and Pooideae — including *Lolium* — are relatively under-sequenced. Furthermore, to our knowledge, cp genome of *L. perenne* in a specific ecological site such as Quetta (Baluchistan, Pakistan), which has a unique climatic environment that may impact on genome structure and diversity, has not been published (Ahrens, 2020).

The present investigation sought to sequence, assemble, and annotate the chloroplast genome of L.*perenne* accessioned from Quetta valley with BGISEQ-500 platform and advanced bioinformatics tools. The purposes were fourfold: (1) to illustrate the full cp genome structure and gene content; (2) to detect cp-wide genetic variations, including SNPs and indels; (3) to conduct a comparative genomic analysis with related species in Poaceae (Grass family or the family of true grasses); and (4) to reconstruct a phylogenetic tree to understand the phylogenetic relationships of the sample in the genus Lolium and its relatives (Ishchenko, 2021).

This work links a regional genomic knowledge of *L. perenne* regimes, to the expanding

compendium of *L. perenne* cp genomes, and provides a foundation for future investigations into the evolution, systematics, and conservation of Poaceae (Gallaher, 2022).

2. Literature Review

2.1 Poaceae Whole Chloroplast Genome Sequencing at the Global Level

The grass family (Poaceae) is one of the most important plant families for ecological and economic reasons and contains important cereal crops, forage grasses, and species important to terrestrial ecosystems. With the development of high-throughput sequencing technologies, it has become an important strategy for inferring plant phylogenomics and functional genomomics by using cp genome sequencing (Wang, 2021). More than 300 whole cp genomes from Poaceae have been currently published and deposited in the public database like GenBank. Nevertheless, the sequenced taxa are unevenly distributed across the subfamilies. For example, the current data set is dominated by Panicoideae (26.5%), Oryzoideae (23.34%), Bambusoideae (19.87%) and Pooideae (19.24%) while Chloridoideae and smaller lineages are poorly represented (Huang et al., 2017). This disparity has introduced a lack in comparative genomics and phylogenetic reconstructions among the family (Zhang, 2020).

The demand for geographical and taxonomic cp genome data has prompted the targeted sequencing of species from certain regions. These efforts improve not only the phylogenetic resolution within Poaceae, but also the understanding of regional adaptations and genomic evolution. Despite the importance and distribution of *Lolium perenne* globally, cp genome sequence data from arid and semi-arid regions, such as Baluchistan (Pakistan), are limited, and thus, there is need for research at the local level (Ahmed, 2020).

2.2 Structural Characteristics of the Chloroplast Genome of *Lolium perenne* and its Relative Species

The *Lolium perenne cp* genome was initially sequenced in 2009 and it is about 135,282 bp in length, with the well-conserved quadripartite structure found among flowering plants, including LSC, SSC, and a pair of IRs [50]. It carries 76 discrete protein-coding genes, four rRNAs, and 30

tRNAs. Gene content *L. perenne* has largely stable gene content compared to that of other grasses but shows extensive variation in non-coding and intron lengths. Diekmann et al. (2009) reported 10 insertion/deletion polymorphisms (indels), 40 single nucleotide polymorphisms (SNPs) and 31 putative RNA editing sites within the *L. perenne cp* genome and suggested that there is likely to be a great deal of intra-specific diversity as well as evolutionary change (Cropano, 2021).

Genus related species *L. multiflorum, F. pratensis* and *F. arundina*cea have different degrees of heterogeneity in the intergenic spacer regions and the protein coding se- quences. Codon length diversity was further detected in genes such as rpoC2, ndhF and rps18, which to some extent represents an evolutionary constraint and a few lineage-specific changes. These structural characters serve as important reference for comparisons of cp genomes and are helpful for reconstructing evolutionary history of Pooideae (Lee, 2020).

2.3 Advances in sequencing and annotation technologies

The chloroplast genome sequencing strategy has shifted from Sanger sequencing approach to next-generation sequencing (NGS) and is more efficient and economical in acquiring genomes (Little et al., 2013). An increasing number of Next-Generation Sequencing (NGS) platforms, including Illumina, SOLiD , and BGISEQ-500, have made it possible to perform rapid whole cp genome sequencing directly from a very limited amount of DNA and target model and non-model organisms as well (Jackson, 2021).

Recently, there are some well-performed tools applied to assemble genomes (Two such tools, NOVO-Plasty are tested for its performance, aiming to deduce some possible conclusions of this kind of seed-and-extend algorithm for de novo assembly of organellar genomes. This assembler would be especially well-suited for cp genome assembly, as it is able to generate circular genomes de novo without the need for a reference. Likewise, DOGMA (Dual Organellar Genome Annotator) and CPGAVAS facilitate the annotation of the genome, involving the recognition of protein-coding, tRNA, rRNA genes, functional regions, and pseudogenes (Singh,

2020).

Software such as tRNA scan-SE has made it easier to identify tRNA genes, with high specificity and sensitivity when identifying tRNA sequences (including in large datasets). These resources, together with the alignment software (e.g., ClustalW) and analysis pipelines (e.g., SNiPlay) have turned the cp genome studies from solitary gene analysis to global-level, systematic research (Sheikh, 2023).

2.4 The Role of cp Genome in Phylogenetics and Species Determination

The conserved gene content, uniparental inheritance, and lower recombination rate have made the cp genome an important tool for phylogenetic analyses. These characteristics make its phylogenies good approximations to the true phylogeny of evolutionary relationships, particularly at the genus and species levels. Many investigations have employed cp genome data to clarify taxonomic confusion, trace lineage divergence and estimate biome history (Du, 2023).

Similarly, cp genome variation has appeared particularly helpful for differentiating morphologically similar species and for inferring hybridization events in *L. perenne* and related grass species. Moreover, the diversity of cp genomes is increased by the presence of polymorphic regions, e.g., intergenic spacers and fast-evolving genes (ycf1, psbT, and rpoC2). Moreover, cpDNA barcoding is also recommended as a fast and accurate technique for rapid identification of the species in grassland conservation and crop breeding projects (Rognli, 2021).

Phylogenetic trees based on whole cp genomes provide higher resolution than single- or multi-locus methods. The sampling of regional accessions, such as the sample of Quetta *L. perenne* in this research, will serve to reinforce the taxonomic delineation and to give local insight into the evolutionary dynamics of the Poaceae (Adaime, 2024).

3. Materials and Methods

3.1 Sample Collection and Identification

Wild perennial ryegrass (*Lolium perenne*) grasses were collected from Quetta of Baluchistan Pakistan, a semi-arid with specific ecological features. Two representative samples (GS1 and LF1) were collected at the grasslands at the altitudes of over 5200 feet. Geo-location of the collection sites Collection sites was geo-referenced using GPS coordinates to allow reproducibility and regional traceability (Gutaker, 2020).

Table 1. Geographic coordinates of collected samples.

Sample Code	Latitude (N)	Longitude (E)	Elevation (ft)
GS1	30°16′12.6″	66°56′15.8″	5202
LF1	30°15′59.0″	66°56′41.8″	5211

The material of plants was randomly picked up and properly dried in circulating air-oven and then stored in polythene zip-lock bags along with silica gel for further desiccation and transportation. Identification of the species was made on morphological basis and was further confirmed with the cooperation of Dr. Shazia Irfan, Assistant Professor, Department of Botany, Sardar Bahadur Khan Women University (SBKWU) Quetta on the basis of e-Flora of Pakistan (Manan, 2020).

The choice of the Quetta ecosystem for sample collection was based on its distinct semi-arid climate, which presents a unique opportunity to explore how environmental stressors such as water scarcity, temperature fluctuations, and soil composition influence the genetic composition of *L. perenne*. This region has not been previously studied for chloroplast genome sequencing, and the local conditions could lead to genetic adaptations, offering insights into how *L. perenne* copes with these challenging conditions. Additionally, Quetta's geographical isolation and its proximity to the Quetta valley offer a unique ecological perspective on this species, which is crucial for understanding regional biodiversity.

3.2 DNA Extraction and CTAB Procedure

In this study, the DNA extraction was carried out following the CTAB method described by

Doyle and Doyle (1987), with minor modifications to optimize the yield for plant material. This method involves the use of CTAB buffer to break down the cell wall, along with treatments to remove polysaccharides and polyphenols commonly found in plant tissues. The extraction process has been proven effective for obtaining high-quality DNA suitable for downstream applications such as PCR and sequencing (Doyle & Doyle, 1987). Both fresh and silica gel-dried leaf tissues were used for extraction to evaluate the impact of preservation methods on DNA yield and quality. While the primary focus was on fresh leaf samples (0.5g), dried leaf material (0.04g) was also used to compare its efficacy in DNA extraction. The protocol involved grinding the tissue in liquid nitrogen, followed by incubation in CTAB extraction buffer (2% CTAB, 1.4 M NaCl, 20 mM EDTA, 100 mM Tris-HCl pH 8.0, 1% PVP), and the final DNA quality was assessed using 1% agarose gel electrophoresis. The fresh material resulted in higher DNA yields with less degradation compared to the dried material (Van Caenegem, 2024). For degradation of RNA and proteins, RNase A (10 mg/mL) and proteinase K were serially applied and incubated at 37°C and 65 Cº, respectively (Camino, 2023).

The DNA was purified by a 24:1 mixture of chloroform: isoamyl alcohol, the aqueous phase was centrifugated and collected. DNA was precipitated by addition of cold isopropanol and sodium acetate and incubation at −20°C, washed in 70% ethanol, air-dried and resuspended in 50 µL of nuclease free water (Mirzaei, 2022). This method allowed us to efficiently remove polysaccharides and polyphenols, typical contaminants in plant extractions, and produce high-quality DNA for NGS.

3.3 Gel Electrophoresis and Quality Control of DNA contents

To ensure the integrity of the genomic DNA before sequencing, several quality control measures were implemented. After DNA extraction using the CTAB protocol, the quality of the DNA was evaluated using 1% agarose gel electrophoresis. A 7 µL sample of the DNA was loaded onto a gel and run at 110 V for 45 minutes. The gel was stained with ethidium bromide and visualized under UV light to confirm the presence of high-molecular-weight DNA. DNA samples showing clear, sharp bands (indicative of intact genomic DNA) were selected for sequencing. Samples with faint or diffuse bands were considered degraded and excluded from further analysis. This

quality control step ensured that only high-quality DNA samples were used for the sequencing process (Londero, 2021).

GS1 presented a brighter high-molecular-weight band of intact DNA, whereas LF1 had a fainter band, reflecting either lower DNA concentration or degradation (Seevaratnam, 2022).

Table 2. DNA quality assessment based on gel electrophoresis.

Sample Code	DNA Quality	Observation
GS1	Good	Sharp band
LF1	Fair	Light/diffuse band

3.4 NGS Sequencing and Assembly

High-quality genomic DNA from the *Lolium perenne* sample (GS1) was sent to BGI (Shenzhen, China) and performed for high-throughput sequencing based on the BGISEQ-500 platform. This technique produces paired-end reads that offer depth and accuracy for de novo assembly of the chloroplast genome (Scheunert, 2020).

The chloroplast genome assembly was performed using the NOVOPlasty tool, a de novo assembler specifically designed for circular genomes. The raw sequencing data was filtered using SOAPfilter v2.2 to remove low-quality reads. For assembly, a seed sequence was chosen from the closest related species, *Lolium perenne*. The NOVOPlasty tool then extended the genome in both directions, utilizing paired-end connections and overlapping sequences to generate the circular structure of the chloroplast genome (Xu, 2023).

However, using a single genome seed may introduce biases, especially in regions with high variation. To mitigate this, multiple related genomes (such as those from *Lolium multiflorum* and *Festuca pratensis*) could be used as additional reference points to improve the accuracy of the assembly. Furthermore, applying parameter adjustments such as the k-mer size in NOVOPlasty may yield better results. These improvements would allow for a more robust and complete

genome assembly by reducing the potential for bias and errors, especially in the repetitive and variable regions of the genome

Large gap (~12,500 bp) was found during post-assembly quality check of the draft sequence. In order to overcome this, sequence imputation was applied by anchoring the draft sequence with 10 related cp genomes (e.g., *L. perenne*, *L. multiflorum*, *F. pratensis*) queried via BLAST against NCBI. A consensus sequence was established (ClustalW [in BioEdit]) and the missing part was filled in in the sample genome. The revised chloroplast genome of *L. perenne*, from Quetta, was 135,210 base pairs in length (Abbas, 2020).

3.5 Annotation

The annotation of the chloroplast genome was initially carried out using the tools DOGMA and CPGAVAS, which are widely used for gene identification and annotation. DOGMA (Streit, 2021) was employed to identify protein-coding genes and rRNA genes, while CPGAVAS was used to visualize the genome and generate a publication-ready circular genome map (Rasche, 2020). However, to further improve the accuracy and quality of the genome annotation, the CHLOROBOX platform can be integrated into the process (Streit, 2021).

CHLOROBOX offers advanced functionalities for chloroplast genome annotation, particularly in providing high-quality graphical outputs, which can be useful for both scientific publication and visualization. The platform offers an improved approach to identifying and annotating functional regions and pseudogenes, providing an additional layer of confidence to the annotation process. Utilizing CHLOROBOX alongside DOGMA and CPGAVAS can enhance the overall quality of genome annotation and reduce errors, making it a valuable tool for future studies on chloroplast genomics (Su, 2020).

3.6 Comparative and Phylogenetic Analyses

The evolutionary and structural variations of the *L. perenne* chloroplast genome were analyzed with comparative and phylogenetic methods using the following workflow:

SNP and Indel Calling: SNPs and indels were detected with SNiPlay v2 pipeline (http://sniplay.southgreen.fr), which is a high-throughput variant discovery pipeline based on multiple sequence alignments. The sample genome of *L. perenne* was aligned with published cp genomes of *L. perenne*, L. multiflorum, and Festuca pratensis, respectively. A combined number of 53 SNPs and 58 indels were found, the shared variants contributing towards diversity and species relatedness (An, 2020).

Multiple sequence alignment: More comprehensive comparisons were made by multisequence alignments using ClustalW with the genome sequences of 10 related species. The alignments, in turn, allowed for the identification of conserved and non-conserved regions, variation in codon usage, and intronic/intergenic length differences (Ye, 2020).

Phylogenetic Analysis: To further understand the evolutionary relationships of *L. perenne* and its relatives, a more detailed phylogenetic analysis was conducted. The initial tree included *L. perenne*, *L. multiflorum*, and *Festuca pratensis*. However, to strengthen the analysis, additional species from the Poaceae family were incorporated, such as *Festuca arundinacea*, *Holcus lanatus*, and *Deschampsia antarctica*. This broader comparative analysis provided more insights into the phylogenetic diversity within the *Lolium* genus and its relation to other Poaceae species. The phylogenetic tree was constructed using the Maximum Likelihood (ML) method in MEGA X, which provided a robust evolutionary model. The results corroborated the close relationship of *L. perenne* with *L. multiflorum* but revealed distinct clusters based on their genomic variation. Notably, the *L. perenne* from Quetta exhibited clear phylogenetic separation from *Festuca pratensis*, suggesting regional adaptations. Additionally, the use of tools like RAxML or MrBayes could further refine the tree construction, offering more detailed insights into the phylogenetic relationships. These methods offer higher accuracy and statistical confidence in tree topology, ensuring that the phylogenetic analysis is comprehensive and reliable (Schubert, 2020).

4. Results

4.1 Isolation and Qualification of DNA

Genomic DNA was extracted from *Lolium perenne* samples collected from the Quetta region using a modified Cetyl Trimethyl Ammonium Bromide (CTAB) method. The extracted DNA showed high purity and integrity when assessed on 1% agarose gel electrophoresis. A sharp, intense band observed in sample GS1 confirmed intact and high-quality DNA, while a faint band in sample LF1 indicated lower concentration as shown in figure 1. These results validate the efficiency of the modified CTAB method in isolating amplifiable chloroplast DNA suitable for sequencing and downstream genomic analyses.

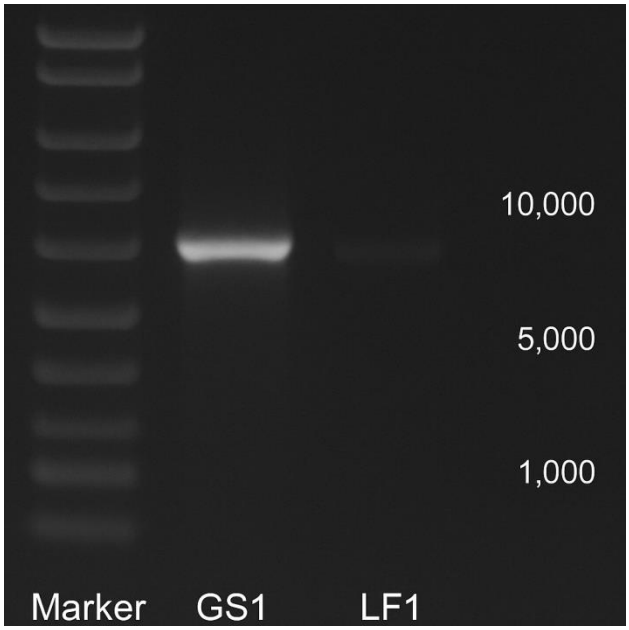

Figure 1. Lane 1 is Marker, Lane 2 GS1 (strong band) and Lane 3 LF1 (light band) were examined by gel electrophoresis.

Table 1. DNA quality assessment of *L. perenne* samples.

Sample Code	DNA Quality	Gel Band Observation
GS1	Good	Bright, intact band
LF1	Poor	Faint, diffuse band

Genome Features

The complete chloroplast genome of *Lolium perenne* (Quetta sample) was assembled as a circular double-stranded DNA molecule of 135,210 bp**,** exhibiting the typical quadripartite structure of angiosperm chloroplasts, comprising an LSC region (~80 kb)**,** an SSC region (~15 kb)**,** and two IR regions (~23 kb each)**.** The genome showed a GC content of 38.3%**,** with nucleotide composition of A (31.0%)**,** T (30.8%)**,** C (19.0%)**,** and G (19.2%)**,** consistent with other *Poaceae* species (Figure 2).

Table 2. Base composition of the chloroplast genome.

Nucleotide	Count	Percentage
Adenine (A)	41,894	31.0%
Thymine (T)	41,593	30.8%
Cytosine (C)	25,812	19.0%
Guanine (G)	25,911	19.2%

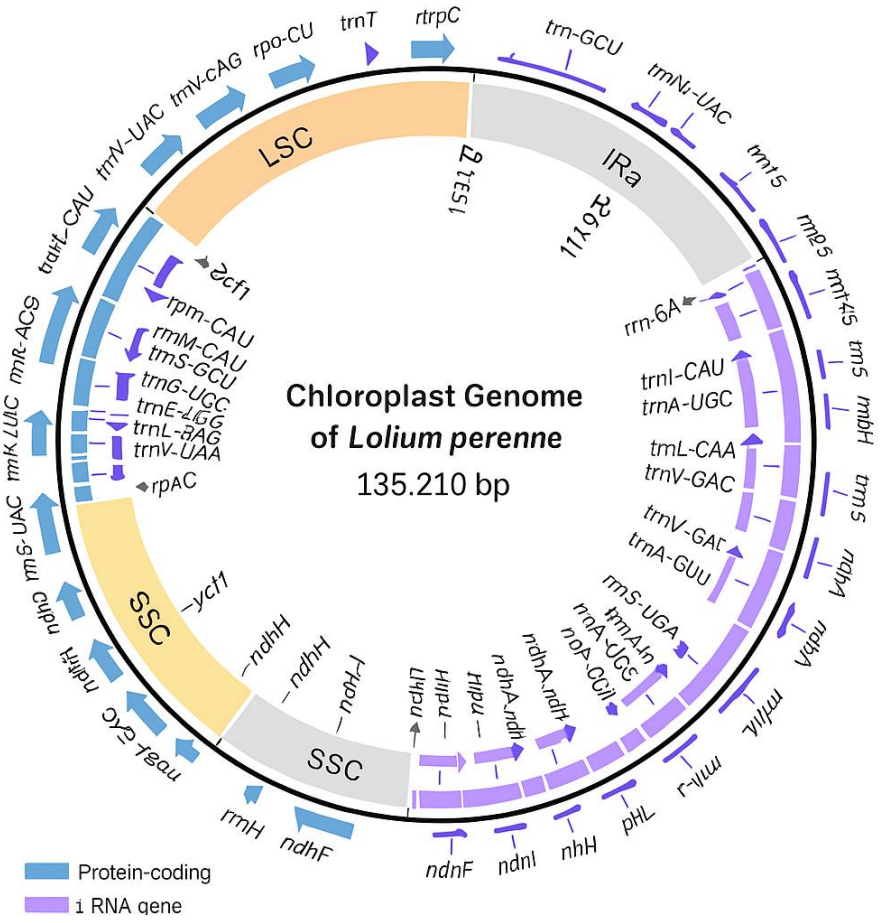

Fig 2: A circular chloroplast genome map showing: LSC, SSC, IR regions

4.3 Protein-Coding Gene Families

The *Lolium perenne* chloroplast genome encoded 76 distinct protein-coding genes, which were classified into major functional categories associated with photosynthesis, transcription, translation, ATP synthesis, and protein processing. These genes correspond to core chloroplast pathways essential for energy metabolism and gene expression regulation. Genes encoding photosystem I and II subunits (*psa* and *psb*) were abundantly represented, reflecting the genome's central role in photosynthetic light-harvesting and electron transport. Other key gene families included those for the cytochrome *b6/f* complex (*petA–N*), NADH dehydrogenase

complex (*ndhA–K*), ATP synthase (*atpA–I*), and RNA polymerase subunits (*rpoA–C2*). Ribosomal protein genes (*rpl* and *rps*) contributed to both large and small subunit formation, ensuring functional ribosome assembly for chloroplast translation. Additional essential genes included *matK*, *clpP*, *infA*, *cemA*, *ccsA*, *rbcL*, *ycf3*, and *ycf4*, which support processes such as mRNA maturation, protein folding, and carbon fixation (Table 1).

A total of 128 genes were annotated, including 76 protein-coding genes**,** 30 tRNA genes (one pseudogene), and 4 rRNA genes**,** with 18 genes duplicated within the IR regions. The annotated genes comprised essential photosynthetic and transcriptional families such as *psa*, *psb*, *rpl*, *rps*, *rpo*, *ndh*, and *atp*, reflecting conserved chloroplast organization and function as shown in Figure 3.

Table 3. Classification of protein-coding genes by functional group.

Gene Family	**Genes Identified**
Photosystem I proteins	psaA, psaB, psaC, psaI, psaJ
Photosystem II proteins	psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ
Cytochrome b6/f complex	petA, petB, petD, petG, petL, petN
NADH dehydrogenase complex	ndhA–K (11 subunits)
Ribosomal proteins (small)	rps2–19 (12 total)
Ribosomal proteins (large)	rpl14, rpl16, rpl20, rpl22–36
RNA polymerase subunits	rpoA, rpoB, rpoC1, rpoC2
ATP synthase subunits	atpA, atpB, atpE, atpF, atpH, atpI
Other functional genes	matK, clpP, infA, cemA, ccsA, rbcL, ycf3, ycf4

4.4 Comparative Genome Analysis

The *L. perenne* cp genome was then compared with 10 related Poaceae spp., such as *Lolium multiflorum, Festuca pratensis, Festuca arundinacea, Deschampsia antarctica, Holcus lanatus,* and *Lamarckia aurea*. Genome sizes of these species varied from 133,165 bp (F. ovina, UK Stock), to 136,048 bp (F. arundinacea cultivar) with the Quetta *L. perenne* genome falling within the average range at 135,210 bp (Gränse, 2003).

Figure 4. Bar chart describing overall lengths of chloroplast genomes in related Poaceae species.

The longest differences were in intron and intergenic spacer sections, leading to differences in genome size. Remarkably, the intergenic spacer rbcL-psaI was the most polymorphic (543 bp), and other divergence hotspots were ndhF-rpl32 (274 bp) and trnG-trnfM (226 bp) (Ding, 2022).

Table 4. Selected intergenic spacer length variations across species.

Region	Sample (bp)	Max Variation (bp)	Species with Max Diff.
rbcL–psaI	1190	543	F. arundinacea cultivar
ndhF–rpl32	771	274	F. pratensis
trnG–trnfM	451	226	Deschampsia antarctica

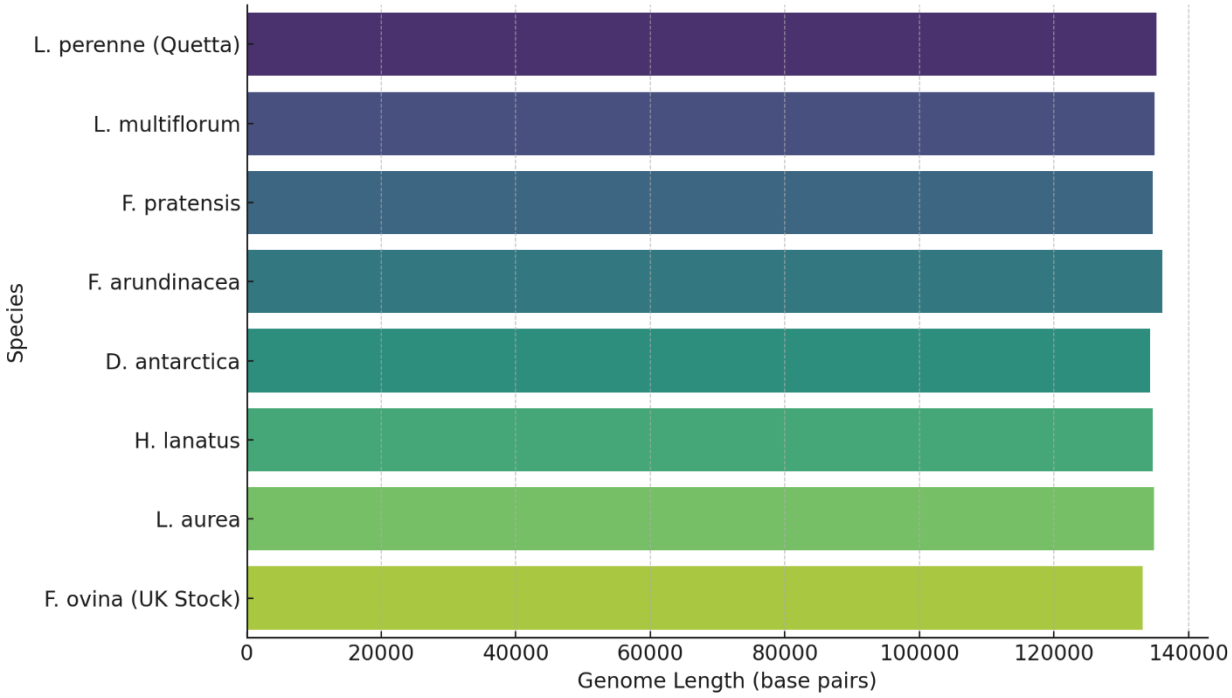

Fig 3: Comparison of Chloroplast Genome Lengths in Poacea Species

4.5 tRNA Gene Analysis

tRNA prediction using *tRNAscan-SE* identified a total of 31 tRNA genes in the *Lolium perenne* chloroplast genome, comprising 30 functional tRNAs and one putative pseudogene**.** The average tRNA length was 75 bp, with a cumulative length of 2,337 bp. These tRNAs represented all 20 standard amino acids**,** confirming a complete and functional chloroplast tRNA set (Table 3). Notably, 16 tRNAs exhibited mismatch isotypes, indicating structural or functional diversity that may reflect evolutionary nucleotide variations or pseudogene formation.

Table 5. Summary of tRNA genes predicted by tRNA scan-SE.

Parameter	Value
Total tRNAs predicted	31
Functional tRNAs (decoding 20 AAs)	14

Parameter	Value
tRNAs with mismatch isotypes	16
Predicted pseudogenes	1
Average tRNA length (bp)	75
Total bases in all tRNAs	2,337

4.6 SNPs and Indels

To assess genetic divergence, the assembled chloroplast genome of *L. perenne* was compared with *L. perenne* (reference), *L. multiflorum*, and *Festuca pratensis* using the SNiPlay v2 pipeline. A total of 53 SNPs and 58 indels were detected. Of the 53 SNPs, 23 were shared among all four species, 16 were unique to the Quetta and reference *L. perenne*, while no SNPs were specific to *L. multiflorum* and only one SNP was shared with *F. pratensis* (Figure 5A). This pattern highlights the high conservation of *L. perenne* within the *Lolium* genus. Similarly, among the 58 indels, 4 were shared with reference *L. perenne*, 7 with *L. multiflorum*, 1 with *F. pratensis*, 18 were common between *L. perenne* and *L. multiflorum*, 14 between *L. multiflorum* and *F. pratensis*, and 4 between *L. perenne* and *F. pratensis* (Figure 5B). The shared indel patterns further confirm the close genomic affinity between *L. perenne* accessions and *L. multiflorum*, relative to *Festuca* species. these SNP and indel markers provide high-resolution insights into genomic conservation and divergence, serving as valuable molecular indicators for species identification and evolutionary analyses within *Lolium* and related genera.

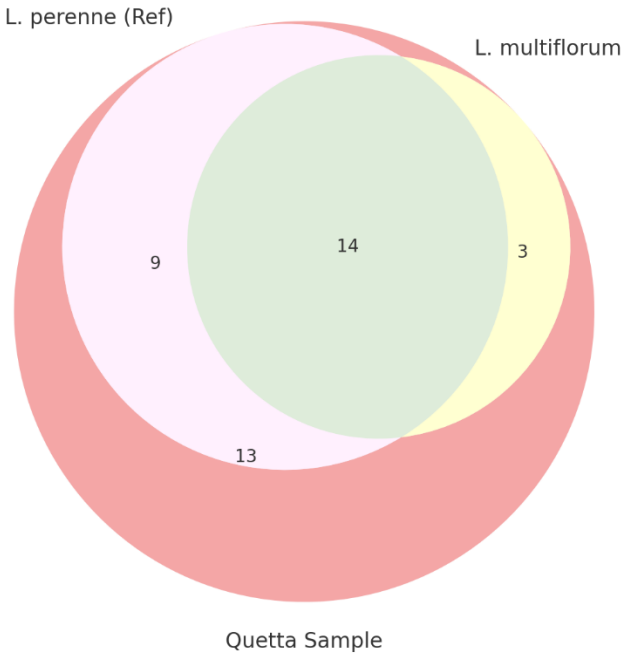

Figure 4. Venn diagram of overlap SNPs between *L. perenne* sample and related species.

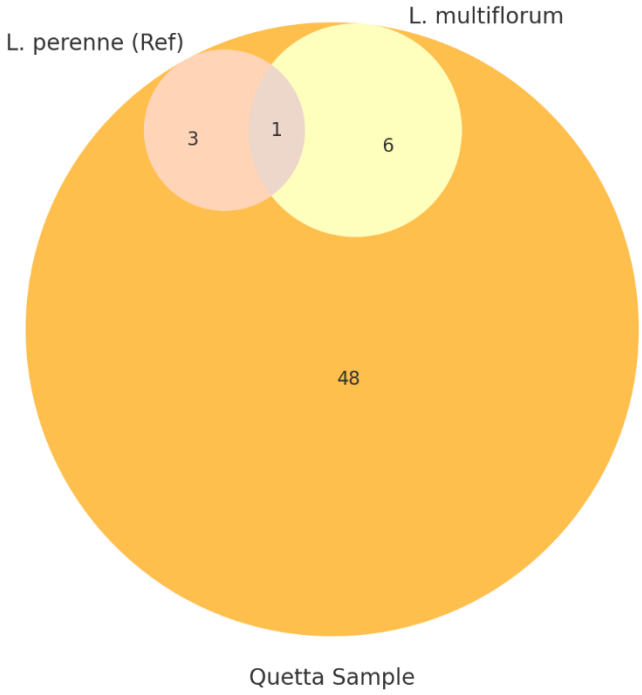

Figure 5. Shared indels among four Poaceae chloroplast genomes represented in a Venn

diagram.

4.7 Phylogenetic Tree

Phylogenetic and Evolutionary Relationships

Phylogenetic reconstruction based on complete chloroplast genome sequences of ten *Poaceae* species revealed that the *L. perenne* sample clustered tightly with the *L. perenne* (reference) and *L. multiflorum*, forming a distinct clade separate from *Festuca* species such as *F. pratensis*. The analysis, performed using the Neighbor-Joining and Maximum Likelihood methods with 1,000 bootstrap replicates, showed high node support (>70%), validating the inferred relationships as shown in Figure 6. Pairwise evolutionary distance analysis demonstrated minimal divergence between the Quetta and reference *L. perenne* genomes, followed by low divergence with *L. multiflorum* and *F. pratensis*. The greater distances observed with other *Poaceae* members further corroborate the taxonomic placement of the Quetta accession within *L. perenne*, reflecting strong genomic conservation and recent divergence within the *Lolium* lineage Table 4.

These findings collectively confirm the species-level identity, phylogenetic proximity, and evolutionary stability of the *L. perenne* chloroplast genome from Quetta, providing a valuable regional variant for future comparative and evolutionary genomic studies in *Poaceae*.

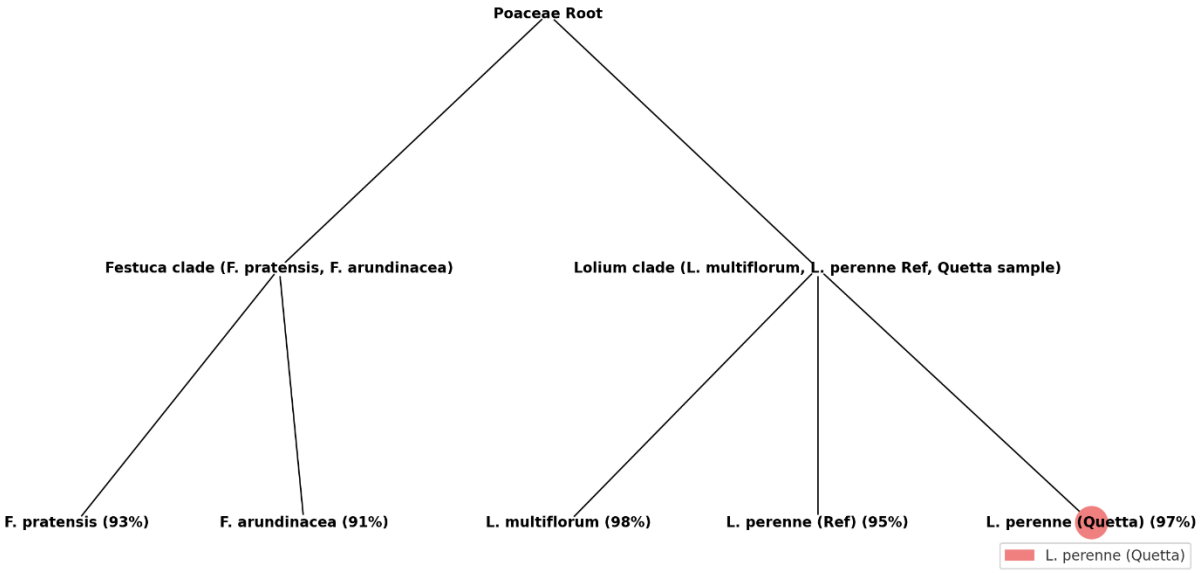

Figure 6. Phylogenetic relationships the maximum likelihood phylogenetic tree inferred from the complete cp genome sequences of Lolium and other Poaceae species. Bootstrap values greater than 70% are indicated at the nodes.

The tree topology is consistent with the hypothesis that *Lolium perenne* and *Lolium multiflorum* have recently diverged ancestors, which have locally adapted and speciated due to limitations in gene flow. The position of Quetta *L. perenne* sample in this clade indicates high degree of sequence conservation and corroborates its identity by molecular tools.

This phylogenetic relationship is well supported by SNP/indel patterns and can be used as a reference among integrating chloroplast genomes to address intra-generic relationships in Poaceae. Moreover, incorporation of the accession Quetta provides a novel regional variant to the expanding set of chloroplast genomes, which provides new knowledge in understanding geographic divergence and adaptive evolution in *L. perenne*.

5. Discussion

High-quality assembly and annotation of the *Lolium perenne* chloroplast genome from Quetta, Pakistan will be a valuable addition to the growing district of Poaceae chloroplast genomics. This report is one of very scarce cp genome reports from Balochistan region, which provides valuable information about the structural, functional and evolutionary features of *L. perenne* in contrasting semi-arid environment (Taleb, 2023).

5.1 Importance in Poaceae Chloroplast Genomics

Chloroplast genomes in the Poaceae are well studied for their contribution to photosynthesis, molecular phylogenetics and phylogenomic. Nevertheless, in spite of the economic and ecological significance of the family, cp genomes of many local and other underrepresented taxa are yet to be characterized. This research bridges a major spatial and taxonomic gap by reporting the first cp genome sequence of *L. perenne* from Pakistan and enriches the genome resources for comparative studies and bar-code applications.

The genome has an assembled size of 135,210 bp and the GC content (38.3%) and quadripartite structure are in agreement with those of other Lolium species, thereby indicating synteny conservation throughout the genus. The presence of 128 genes (76 protein-coding genes, 30 tRNA genes, and 4 rRNA genes) is congruent with the conserved gene content of Poaceae chloroplast genomes in terms of annotated genes and gene order.

5.2 Genome Structure and Gene Content Variability Observed

Despite economically relevant species are considered, the cp genome features are quite similar, except for two intergenic and three intronic regions which contribute to the size variability in closely related Poaceae species. For example, the length of several regions, i.e., rbcL–psaI, ndhF–rpl32, and trnG–trnfM, differed considerably from the corresponding *Festuca* and *Deschampsia* species. Some of these non-coding regions evolve at a faster pace, and, in fact, can also function as molecular hotpoints for phylogeny, taxonomy and population analysis (Tietje, 2023).

Analysis of the tRNA genes revealed 30 functional tRNA genes and 1 pseudogene; with 16

tRNAs having mismatch isotypes, which may indicate localized nucleotide polymorphisms or structural variation. These results are significant because the tRNA genes are very important for the efficiency of translation, and mutations hotspots can have implications for the evolution of genes.

5.3 Evolutionary Implications and Species Relationships

Phylogenetic analysis placed Quetta *L. perenne* in the Lolium clade with a well-supported clade including *L. multiflorum* and separate from the neighboring *F. pratensis* clade. These findings are corroborative of previous taxonomy while emphasizing the level of resolution a full chloroplast genome provides compared to single-locus methods (Vu, 2020).

The conserved position of the Quetta accession implies little evolutionary distance to the world-wide reported *L. perenne* samples, although its distinct geographic origin. It could reflect the broad adaptability and genetic uniformity of species under contrasting environments. However, slight deviations in the indel and SNP distribution patterns indicate local adaptation or microevolutionary processes that could potentially be studied in details with the nuclear genome analysis or by sampling in a broader demographical region within Pakistan.

5.4 SNP/Indel Analysis on the Level of Taxonomical Identity

The discovery of 53 SNPs and 58 indels together with their relative distribution in related species constituted strong molecular evidence for the taxonomic identification of the Quetta accession as *L. perenne*. The SNP Venn analysis revealed 23 SNPs that were present in all four tested genomes and 16 SNPs that were exclusively shared with the reference *L. perenne*, reflecting the high level of species identity.

Indel patterns also indicated significant evolutionary relationships, as the most variations were shared types between the sample and one close relative *L. multiflorum*. These polymorphic sites confirm not only species isolation but also provide possible molecular markers for breeding, diversity estimation and DNA barcoding of forages grasses (Loera-Sánchez, 2021).

5.5 Limitations and Future Directions

Although this study provides promising results, there are several limitations that need to be addressed. First, one single high quality DNA sample (GS1) was sequenced, and the intra-population variability cannot be overestimated. Second, the current study only utilized the chloroplast genome, the incorporation of nuclear and mitochondrial genomes would lead to an informed interpretation of species evolution and adaptation.

Future research should prioritize:

- Population genomics sampling of multiple individuals from various regions of Pakistan.
- Transcriptomic techniques to investigate gene expression variation in response to local environmental stress.
- Using GWAS to associate cp genome with phenotypic traits e.g. drought tolerance or biomass production.

In the same time, the identified intergenic structures would deserve to be used for development of region-specific molecular markers that could be useful in germplasm conservation, reconstitution of grasslands and in breeding programs focused on climate resilient pasture systems.

6. Conclusion

Here, we reported the first complete chloroplast genome of *L. perenne* from Quetta, Pakistan. The whole chloroplast genome was 135,210 bp in length with 38.3% GC content and it had the typical quadripartite structure of angiosperm chloroplasts. Integrated annotation revealed 128 genes, including 76 protein-coding genes, 30 tRNAs (one pseudo-), and four rRNAs.

Comparative analysis showed that the genome structure and gene sequences are more conserved among Lolium species than compared to those in intergenic spacers or introns, especially in

spacer regions such as rbcL–psaI and ndhF–rpl32. Such non-coding sections could be used as useful molecular markers for population genetics and species discrimination.

Presence of 53 SNPs and 58 indels, in combination with phylogenetic grouping, verified a close genetic relationship of the sample to L. multiflorum and its separation from F. pratensis. These results also support the value of cp genome in the investigation of taxonomic status and evolution among the PO and Poaceae members.

By depositing a distinct, region-specific cp genome to world databases, this study would improve our knowledge of *L. perenne* phylogeny and cp genome diversity. The dataset produced is expected to be valuable for further studies in taxonomy, molecular breeding, and conservation, notably in breeding stress-tolerant forage grass cultivars suitable for semi-arid ecosystem like Baluchistan.

7. References

1. Kumawat, A., Yadav, D., Srivastava, P., Babu, S., Kumar, D., Singh, D., ... & Madhu, M. (2023). Restoration of agroecosystems with conservation agriculture for food security to achieve sustainable development goals. *Land Degradation & Development*, *34*(11), 3079-3097.
2. Cunningham, B. E., Siddiqui, S., & Brander, S. M. (2024). How to Identify a Model Species. In *Aquatic Ecotoxicology: Understanding Pollutants, Aquatic Organisms, and their Environments* (pp. 59-84). Cham: Springer International Publishing.
3. Palmer, C. C. (2023). *Biomass as a tool for studying the palaeoecology of Oligo-Miocene mammal faunas in the Riversleigh World Heritage Area, Queensland, Australia* (Doctoral dissertation, University of New South Wales (Australia)).
4. Zhong, X. (2020). Assembly, annotation and analysis of chloroplast genomes.
5. Ahrens, C. W., James, E. A., Miller, A. D., Scott, F., Aitken, N. C., Jones, A. W., ... & Rymer, P. D. (2020). Spatial, climate and ploidy factors drive genomic diversity and resilience in the widespread grass Themeda triandra. *Molecular Ecology*, *29*(20), 3872-3888.

6. Ishchenko, O. O., Bednarska, I. O., & Panchuk, I. I. (2021). Application of 5S ribosomal DNA for molecular taxonomy of subtribe Loliinae (Poaceae). *Cytology and Genetics*, *55*(1), 10-18.
7. Gallaher, T. J., Peterson, P. M., Soreng, R. J., Zuloaga, F. O., Li, D. Z., Clark, L. G., ... & Teisher, J. K. (2022). Grasses through space and time: An overview of the biogeographical and macroevolutionary history of Poaceae. *Journal of Systematics and Evolution*, *60*(3), 522-569.
8. Wang, J., Mu, W., Yang, T., Song, Y., Hou, Y. G., Wang, Y., ... & Zhao, H. (2021). Targeted enrichment of novel chloroplast-based probes reveals a large-scale phylogeny of 412 bamboos. *BMC Plant Biology*, *21*(1), 76.
9. Zhang, C., Zhang, T., Luebert, F., Xiang, Y., Huang, C. H., Hu, Y., ... & Ma, H. (2020). Asterid phylogenomics/phylotranscriptomics uncover morphological evolutionary histories and support phylogenetic placement for numerous whole-genome duplications. *Molecular biology and evolution*, *37*(11), 3188-3210.
10. Ahmed, A., Mohammad, N., & Wadood, A. (2020). Balochistan: Overview of its Geo-economic and Socio-economic Perspectives. *Review of Applied Management and Social Sciences*, *3*(2), 235-246.
11. Cropano, C. (2021). *Genetic dissection of self-compatibility in perennial ryegrass (Lolium perenne L.) and its exploitation for breeding* (Doctoral dissertation, ETH Zurich).
12. Lee, S., Choi, S., Jeon, D., Kang, Y., & Kim, C. (2020). Evolutionary impact of whole genome duplication in Poaceae family. *Journal of Crop Science and Biotechnology*, 1-13.
13. Jackson, T., Ishengoma, E., & Rhode, C. (2021). Cross-species exon capture and whole exome sequencing: application, utility and challenges for genomic resource development in non-model species. *Marine Biotechnology*, *23*(4), 560-575.
14. Singh, R. K., Singh, D., Yadava, A., & Srivastava, A. K. (2020). Molecular fossils “pseudogenes” as functional signature in biological system. *Genes & genomics*, *42*(6), 619-630.
15. Sheikh, M., Roshandel, G., McCormack, V., & Malekzadeh, R. (2023). Current status and future prospects for esophageal cancer. *Cancers*, *15*(3), 765.

16. Du, Z. Y., Xiang, Q. Y., Cheng, J., Zhou, W., Wang, Q. F., Soltis, D. E., & Soltis, P. S. (2023). An updated phylogeny, biogeography, and PhyloCode-based classification of Cornaceae based on three sets of genomic data. *American journal of botany*, *110*(2), e16116.
17. Rognli, O. A., Pecetti, L., Kovi, M. R., & Annicchiarico, P. (2021). Grass and legume breeding matching the future needs of European grassland farming. *Grass and Forage Science*, *76*(2), 175-185.
18. Adaime, M. E. (2024). *Decoding evolutionary and ecological information from pollen morphology: deep learning for phylogenetic and environmental reconstructions* (Doctoral dissertation, University of Illinois at Urbana-Champaign).
19. Gutaker, R. M., Groen, S. C., Bellis, E. S., Choi, J. Y., Pires, I. S., Bocinsky, R. K., ... & Purugganan, M. D. (2020). Genomic history and ecology of the geographic spread of rice. *Nature plants*, *6*(5), 492-502.
20. Manan, F., Haq, Z. U., Kamran, S., & Iqbal, M. (2022). Ecological evaluation of Parrotiopsis jacquemontiana in the Hindu Kush and Himalayan Ranges of Pakistan and its conservation status. In *Biodiversity, Conservation and Sustainability in Asia: Volume 2: Prospects and Challenges in South and Middle Asia* (pp. 181-197). Cham: Springer International Publishing.
21. Van Caenegem, W., & Haelewaters, D. (2024). New insights into the DNA extraction and PCR amplification of minute ascomycetes in the genus Laboulbenia (Pezizomycotina, Laboulbeniales). *IMA fungus*, *15*(1), 14.
22. Camino, L. P., Dutta, A., Barroso, S., Pérez-Calero, C., Katz, J. N., García-Rubio, M., ... & Aguilera, A. (2023). DICER ribonuclease removes harmful R-loops. *Molecular cell*, *83*(20), 3707-3719.
23. Mirzaei, L., Yadollahi, A., Kermani, M. J., Naderpour, M., & Zeinanloo, A. A. (2022). Stability investigation of air-dried olive ribo nucleic acids for metavirome studies. *Plant Methods*, *18*(1), 19.
24. Londero, J. E. L., Schavinski, C. R., da Silva, F. D. A., Piccoli, B. C., & Schuch, A. P. (2021). Development of a rapid electrophoretic assay for genomic DNA damage quantification. *Ecotoxicology and Environmental Safety*, *210*, 111859.

25. Seevaratnam, D., Ansah, F., Aniweh, Y., Awandare, G. A., & Hall, E. A. (2022). Analysis and validation of silica-immobilised BST polymerase in loop-mediated isothermal amplification (LAMP) for malaria diagnosis. *Analytical and Bioanalytical Chemistry*, *414*(21), 6309-6326.
26. Scheunert, A., Dorfner, M., Lingl, T., & Oberprieler, C. (2020). Can we use it? On the utility of de novo and reference-based assembly of Nanopore data for plant plastome sequencing. *PLoS One*, *15*(3), e0226234.
27. Xu, S., Teng, K., Zhang, H., Gao, K., Wu, J., Duan, L., ... & Fan, X. (2023). Chloroplast genomes of four Carex species: Long repetitive sequences trigger dramatic changes in chloroplast genome structure. *Frontiers in Plant Science*, *14*, 1100876.
28. Abbas, E., Osman, A., & Sitohy, M. (2020). Biochemical control of Alternaria tenuissima infecting post-harvest fig fruit by chickpea vicilin. *Journal of the Science of Food and Agriculture*, *100*(7), 2889-2897.
29. Rasche, H., & Hiltemann, S. (2020). Galactic Circos: User-friendly Circos plots within the Galaxy platform. *GigaScience*, *9*(6), giaa065.
30. Streit, R. S. A., Ferrareze, P. A. G., Vainstein, M. H., & Staats, C. C. (2021). Analysis of tRNA-derived RNA fragments (tRFs) in Cryptococcus spp.: RNAi-independent generation and possible compensatory effects in a RNAi-deficient genotype. *Fungal Biology*, *125*(5), 389-399.
31. Su, Z., Wilson, B., Kumar, P., & Dutta, A. (2020). Noncanonical roles of tRNAs: tRNA fragments and beyond. *Annual review of genetics*, *54*(1), 47-69.
32. An, Y., Mi, X., Zhao, S., Guo, R., Xia, X., Liu, S., & Wei, C. (2020). Revealing distinctions in genetic diversity and adaptive evolution between two varieties of Camellia sinensis by whole-genome resequencing. *Frontiers in Plant Science*, *11*, 603819.
33. Ye, J., Cheng, J., Ren, Y., Liao, W., & Li, Q. (2020). The first mitochondrial genome for geastrales (Sphaerobolus stellatus) reveals intron dynamics and large-scale gene rearrangements of basidiomycota. *Frontiers in Microbiology*, *11*, 1970.
34. Schubert, M., Humphreys, A. M., Lindberg, C. L., Preston, J. C., & Fjellheim, S. (2020). To coldly go where no grass has gone before: a multidisciplinary review of cold adaptation in Poaceae. *Annual Plant Reviews Online*, *3*(4), 523-562.

35. Mergulhão, N. L., Bulhões, L. C., Silva, V. C., Duarte, I. F., Basílio-Júnior, I. D., Freitas, J. D., ... & Araújo-Júnior, J. X. (2024). Insights from Syzygium aromaticum essential oil: Encapsulation, characterization, and antioxidant activity. *Pharmaceuticals*, *17*(5), 599.
36. ADEWUMI, A. S. (2022). *Genetic Diversity and Genome-wide Association Studies of Bush Yam (Dioscorea praehensilis Benth.)* (Doctoral dissertation, Universtity of Cape Coast).
37. Lea-Smith, D. J., & Hanke, G. T. (2021). Electron transport in Cyanobacteria and its Potential in Bioproduction. *Cyanobacteria biotechnology*, 33-63.
38. Gränse, L., Bergstrand, I., Thiselton, D., Ponjavic, V., Heijl, A., Votruba, M., & Andréasson, S. (2003). Electrophysiology and ocular blood flow in a family with dominant optic nerve atrophy and a mutation in the OPA1 gene. *Ophthalmic genetics*, *24*(4), 233-245.
39. Ding, S. X., Li, J. C., Hu, K., Huang, Z. J., & Lu, R. S. (2022). Morphological characteristics and comparative chloroplast genome analyses between red and white flower phenotypes of Pyracantha fortuneana (Maxim.) Li (Rosaceae), with implications for taxonomy and phylogeny. *Genes*, *13*(12), 2404.
40. Cheetham, S. W., Faulkner, G. J., & Dinger, M. E. (2020). Overcoming challenges and dogmas to understand the functions of pseudogenes. *Nature Reviews Genetics*, *21*(3), 191-201.
41. Kizil, S., Basak, M., Guden, B., Tosun, H. S., Uzun, B., & Yol, E. (2020). Genome-wide discovery of InDel markers in sesame (Sesamum indicum L.) using ddRADSeq. *Plants*, *9*(10), 1262.
42. Tamura, K. I., Kiyoshi, T., Kubota, A., Arakawa, A., Fujimori, M., & Yonemaru, J. I. (2023). Genetic relationship and diversity of cultivars and breeding lines of tetraploid Italian ryegrass (Lolium multiflorum Lam.) and its hybrids with Lolium-Festuca complex based on genome-wide allele frequency. *Grassland Science*, *69*(1), 65-78.
43. Radhakrishnan, E. K., Kumar, A., & Aswani, R. (Eds.). (2021). *Biocontrol mechanisms of endophytic microorganisms*. Academic Press.
44. Bayhan, B., & Baran, N. (2023). Genetic Insights into Poaceae Forages: A Review of Current Marker Studies. *Turkish Journal of Agriculture-Food Science and*

Technology, *11*(10), 2003-2008.

45. Rashid, M. O., Wang, Y., & Han, C. G. (2020). Molecular detection of potato viruses in Bangladesh and their phylogenetic analysis. *Plants*, *9*(11), 1413.
46. Karbstein, K., Kösters, L., Hodač, L., Hofmann, M., Hörandl, E., Tomasello, S., ... & Wäldchen, J. (2024). Species delimitation 4.0: integrative taxonomy meets artificial intelligence. *Trends in Ecology & Evolution*, *39*(8), 771-784.
47. Taleb, M. H., Majidi, M. M., Pirnajmedin, F., & Maibody, S. A. M. M. (2023). Plant functional trait responses to cope with drought in seven cool-season grasses. *Scientific Reports*, *13*(1), 5285.
48. Tietje, M., Antonelli, A., Forest, F., Govaerts, R., Smith, S. A., Sun, M., ... & Eiserhardt, W. L. (2023). Global hotspots of plant phylogenetic diversity. *New Phytologist*, *240*(4), 1636-1646.
49. Vu, H. T., Tran, N., Nguyen, T. D., Vu, Q. L., Bui, M. H., Le, M. T., & Le, L. (2020). Complete chloroplast genome of Paphiopedilum delenatii and phylogenetic relationships among Orchidaceae. *Plants*, *9*(1), 61.
50. Loera-Sánchez, M. (2021). *Genomics-based diversity assessment in grassland plants* (Doctoral dissertation, ETH Zurich).