International Journal of
Nanotechnology and Allied Sciences

Data Analysis 2026 | Volume 10 | Issue 1 | 97-111

Open Access

Article Information

Received: September 4, 2026

Accepted: September 27, 2026

Published: October 14, 2026

Authors' Contribution

AR, as the first and corresponding author, conceptualized the study, performed formal analysis, data curation, validation and wrote the original draft of the manuscript. HF: contributed to formal analysis and Visualization. UF: was responsible for data interpretation and formal analysis. FS: conducted the literature review and assisted in methodology and manuscript writing. MS, as the supervisor provided supervision, project administration, review and edited the final manuscript for publication.

Citation

Riaz, A., Fatima, H., Tabassum, U., Shakil, F., Sajid, M., 2026. Genome-Wide Characterization of the PPR Gene Family in *Cicer arietinum* L. and Its Evolutionary Diversification. Int. J. Nanotechnol. Allied Sci., 10(1): 97-111.

***Correspondence**

Adeeba Riaz
Email:
adeebariaz486@gmail.com

Possible submissions

Submit your article

Genome-Wide Characterization of the PPR Gene Family in *Cicer arietinum* L. and Its Evolutionary Diversification

Adeeba Riaz[1*], Huma Fatima[1], Urooj Tabassum[2], Faryal Shakil[3], Muhammad Sajid[1]

[1]Department of Biotechnology, University of Okara, Punjab, Pakistan.
[2]Department of Biochemistry, University of Agriculture, Faisalabad, Punjab, Pakistan.
[3]Department of Biotechnology, Shaheed Zulfikar Ali Bhutto Institute of Science and Technology (SZABIST), Karachi, Sindh, Pakistan.

Abstract:

Pentatricopeptide repeat (PPR) proteins make up a big group of proteins that are known to occur mainly in plants and take part in many processes of RNA metabolism in organelles, such as RNA editing, processing, translation, and stabilization. The activity of PPR proteins has been investigated but not their distribution and occurrence in the genome of chickpea (*Cicer arietinum L.*) What is why a genome-wide computational analysis was conducted in order to identify PPR gene family in *C. arietinum*. A total of 24 PPR genes were successfully identified, analyzed and described depending upon chromosomal localization, protein properties, exon-intron architecture and coding sequences. Conserved motifs were obtained using MEME Suite while evolutionary relationships were shown by phylogenetic analysis together with the information from gene structure analysis. In the result, important information about genomic and molecular organization, and evolutionary relationships and processes of PPR gene family in *C. arietinum* was received which can be used for functional characterization of PPR genes in chickpea.

Keywords: PPR gene family, *Cicer arietinum*, genome-wide analysis, phylogenetic analysis, conserved motifs, gene structure.

Scan QR code to visit this journal.

©2026 PSM Journals. This work at International Journal of Nanotechnology and Allied Sciences; ISSN (Online): 2523-9252, is an open-access article distributed under the terms and conditions of the Creative Commons Attribution-Non-commercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. To view a copy of this licence, visit https://creativecommons.org/licenses/by-nc-nd/4.0/.

INTRODUCTION

PPR proteins are one of the most extensive families of organelle-targeted RNA-binding proteins in plants (Chen *et al.,* 2018a). They have repeated sequence elements that are likely to participate in RNA binding and post-transcriptional regulation. These proteins are responsible for many key functions, including RNA editing, processing, stabilization, translation and organellar gene expression (Wang *et al.,* 2021). Plants are characterized by having large and diverse PPR gene families due to their important roles in mitochondria and chloroplasts (Gutmann *et al.,* 2020). The diversification of PPR genes is closely associated with the evolution and functional complexity of plant organelles. Their evolutionary conservation and substantial diversity in their structure, composition of motifs, localization in chromosomes and subcellular localization makes the PPR family of great interest from genome-wide perspective (Xing *et al.,* 2018).

Chickpea (*Cicer arietinum* L.) is an economically significant legume crop and a major source of protein food supply (Varshney *et al.,* 2021). Knowledge of PPR gene families in chickpea involved in organellar function and post-transcriptional regulation can be helpful in understanding molecular biology and adaptation of chickpea. Comprehensive studies of plant gene families were conducted; however, the studies into PPR gene family in *C. arietinum* in terms of its properties remain important (Hao *et al.,* 2021). The aim of the present study was to investigate the gene family of PPR in *C. arietinum* at the genome level. The analysis results for the identified PPR genes showed their sequences and physicochemical properties, exon-intron structure, predicted subcellular localization, conserved motifs, phylogenetic relationships, chromosomal distributions, and duplication events. This analysis provides information about the structural organization and evolutionary processes of the PPR gene family in chickpea.

PPR proteins are part of a larger functional group that consists of specific proteins binding to RNA. These proteins have distinct domains that are made by the repetition of motifs, those motifs having the same amino-acid sequence of length 35. Motif organization leads to two simpler subclasses: P-type and PLS-type. The P-type proteins contain only P-type repeats while PLS-type proteins comprise of the combination of P-type, L-type, S-type motifs and undertake the main task of transcript editing. The entire functional domain called E, which is represented in different versions (E-E-DYW), significantly adds to PPR diversity (Miglani *et al.,* 2021; Gutmann *et al.,* 2021; Wang and Tan, 2024). Among other regulatory proteins, the PPR protein family plays a critical role in mitochondrial and chloroplast gene expression processes in plants. They have various functions such as RNA editing, splicing, cleavage, stabilization, and translation (Gorchs Rovira and Smith, 2019; Wang *et al.,* 2021). PPR proteins located in chloroplasts have special significance in regulating photosynthesis and chloroplast related RNA transcripts (He *et al.,* 2019), whereas mitochondrial PPR proteins are mainly involved in regulating respiration- and energy metabolism-related RNA transcripts (Li *et al.,* 2025; Wang *et al.,* 2022). The editing ability of the PPR proteins allows them to alter the process of protein manufacturing and functioning in organelles; at the same time, PPR proteins involved in splicing and stabilization of RNA transcripts allow achieving precise transcript maturation (Oldenkott *et al.,* 2019; Small *et al.,* 2020).

The growth and diversification of the PPR gene family result from gene duplications, sequence divergence, and functional specialization. Duplicated genes may either maintain their primitive functions or experience subfunctionalization and neofunctionalization (Qiao *et al.,* 2019; Feng *et al.,* 2023). However, owing to their modular structure, separate PPR motifs may undergo evolutionary processes that affect their RNA recognition specificity. Phylogenetic analysis, along with conserved motif and gene-structure analysis, sheds light on

the evolutionary pathways and diversification of PPR genes (Chen *et al.,* 2018b). Species-specific gene gains, losses and duplications may contribute to variations in the PPR family in different plant species (Gutmann *et al.,* 2020).

The structure of the genes, conserved motifs and the chromosomal locations can provide valuable information regarding the processes of evolution of the PPR genes. The comparison of exon-intron arrangement may give information about the degree of conservation of the structure but the analysis of the conserved motifs is concerned with pinpointing important for particular biological functions regions. The MEME Suite is commonly used for the search of conserved motifs in the proteins (Su *et al.,* 2019). In addition, the chromosomal distribution may reveal some patterns associated with gene family expansion, including tandem and segmental duplication (Qiao *et al.,* 2019). Taken together, by combining phylogeny, gene structure, conserved motifs and chromosomal distribution it is possible to create a complete picture of the evolution of PPR genes. The major findings of previous studies on PPR gene families across different plant species are summarized in table (1).

Table 1. Previous studies on PPR gene families in plants and their major findings.

Plant species	Study focus	Major findings	Relevance to present study
Arabidopsis thaliana	PPR gene identification and functional characterization	Large PPR family involved in organellar RNA processing and editing	Provides a model for PPR functional diversity
Rice (*Oryza sativa*)	Genome-wide PPR analysis	PPR genes associated with chloroplast and mitochondrial functions	Supports comparative analysis in crops
Maize (*Zea mays*)	PPR genes and organellar RNA regulation	PPR proteins involved in RNA editing, processing, and fertility	Demonstrates functional importance of PPR genes
Soybean (*Glycine max*)	Genome-wide PPR characterization	Identification of diverse PPR genes with variable structures and localization	Relevant legume comparison
Chickpea (*Cicer arietinum*)	PPR gene family characterization	Identification and characterization of PPR genes and their molecular features	Directly relevant to the present study

Source: Data Compiled by the author from review literature.

Leguminous crops, particularly chickpea (*Cicer arietinum* L.), depend on functional chloroplasts and mitochondria for photosynthesis, respiration and energy metabolism, with PPR proteins contributing to these organellar processes (Varshney *et al.,* 2021). The availability of the chickpea reference genome enables genome-wide characterization of PPR genes and their molecular features (Castro *et al.,* 2024). Such analysis can reveal their chromosomal distribution, gene structure, exon–intron organization and potential functions in chloroplasts and mitochondria (Chen *et al.,* 2018b).

The current study intends to conduct a wide-ranging genome characterization of the PPR gene family, as well as to investigate its structure, function and evolution in *Cicer arietinum* L. The specific aims are:

- To detect and characterize PPR genes from the genome of *Cicer arietinum*.
- To analyze the physicochemical properties of the identified PPR proteins, including CDS length, length of protein, molecular weight and pI.
- To analyze the exon-intron structure of the identified PPR gene family.

- To predict the cellular localization of the identified PPR proteins.
- To examine the phylogenetic relationships, chromosomal distribution and evolution of the PPR gene family *in Cicer arietinum* L.

MATERIALS AND METHODS

Wide-Scale Genomic Identification and Characterization of PPR Genes in *Cicer arietinum*

The PPR genes were identified and characterized using the reference genomic information available for *Cicer arietinum* L. Various bioinformatics resources have been utilized to extract the possible PPR sequences from chickpea genomic and protein sequences. Subsequently, both sequence and domain-oriented methods have been applied for candidate PPR gene filtering (Xing *et al.,* 2018). The candidates were then subjected to assessment against PPR family attributes while filtering out unusable or redundant sequences.

In total, 24 PPR genes have been selected for the analysis. The following information has been gathered for each gene: gene ID, chromosome number, strand orientation, CDS length, protein Size, molecular weight, theoretical pI (isoelectric point), structure of introns and exons (Die *et al.,* 2018). While CDS length and protein size were obtained from the sequences accordingly, the molecular weight and theoretical pI were calculated from the amino acid sequences using standard methods (Gasteiger *et al.,* 2005). This data set has been compiled for comparative analysis against selected PPR genes.

PPR Protein Subcellular Localization and Promoter Analysis

The subcellular localization of the identified PPR proteins was predicted through computational methods. Accordingly, amino acid sequences were analyzed to predict their localization into different cellular compartments. Predictions are made about the localization of the 24 proteins which were compared between the family members of the respective proteins. The special focus was paid to the mitochondrial and chloroplast localization due to the known association of PPR proteins with organellar RNA metabolism (Gorchs Rovira and Smith, 2019). The promoter regions of the identified genes in the PPR gene family were analyzed for the prediction of cis-acting elements.

Multiple Sequence Alignment and Phylogenetic Analysis of PPR Proteins

Multiple sequence alignment was performed for the amino acid sequences of the 24 PPR proteins to identify conserved and variable regions as well as perform phylogenetic analysis. The aligned sequences were then used to generate a phylogenetic tree to assess the evolutionary relationships between the identified PPR proteins (Tamura *et al.,* 2021; Kumar *et al.,* 2024). Phylogenetic clusters were compared with the gene structures and conserved motifs in order to further study the relationship between the PPR proteins.

Analysis of Gene Structure and Identification of Conserved Motifs

The exon-intron structure of the PPR genes discovered in this study was examined for genomic characterization. Examination of exon and intron structure was conducted across the 24 examined PPR genes. To ascertain the conserved motifs among the PPR genes, use of MEME (Multiple EM for Motif Elicitation) Suite was made. Sequences of protein of the identified PPR were made available to the MEME for motif identification, where a maximum of 10 motifs were taken into account (Subburaj *et al.,* 2020). The identified motifs were analyzed regarding their consensus sequence, width and distribution among PPRs.

Gene Duplication and Chromosomal Distribution of PPR Genes

Chromosomal information regarding the identified PPR genes was obtained from the

chickpea genome annotation. The corresponding chromosome for each gene was determined using the genomic coordinates. For the purpose of examining the chromosomal distribution, the 24 PPR genes were mapped onto chromosomes 1-8 (Chen *et al.,* 2018b). To evaluate possible duplication events, a comparison of both the genomic locations and sequence relationships of these genes was performed. Genes having closer sequence relationships and proximity in genomic locations were analyzed for duplication events. These observations were considered as possible causes of the diversification of the PPR gene family without inferring which specific mechanism of duplication could be responsible according to insufficient biological evidence (Zhou *et al.,* 2025).

RESULTS

Genome-Wide Identification and Characterization of PPR Genes in *Cicer arietinum*

A total of 24 PPR genes were located in the genome of the *C. arietinum* and were situated on 8 chromosomes, specifically, starting from chromosome 1 to chromosome 8. Characteristics of the genes that were located vary greatly, structurally and in terms of chemicals. The lengths of the coding sequences were between 1275 and 3213 bp and the corresponding protein lengths ranged between 424 and 1071 amino acids. LOC101502837 coded for the shortest protein, while LOC101504413 had the largest protein in terms of amino acids coded.

The theoretical pI of the proteins discovered was anywhere in the range between 4.83 and 9.7 thus leading to some differences in the physicochemical properties of the proteins predicted. The molecular mass differed too and often matched the differences in the length of proteins. The structure of exons and introns exhibited large variations too, as LOC101507871 did not contain any introns and consisted of only one exon, while for LOC101513383 and LOC101500993 there were 3 and 5 introns and 4 and 6 exons.

It states that there exist 24 PPR genes in *C. arietinum* that are located in all 8 chromosomes, but they show uneven distribution across chromosomes. The genes are found across chromosomes 1-8, with chromosome 5 having the most number of the genes and chromosome 2 having only 1 of them. The details of the PPR genes are presented in table (2), including their characteristics, physicochemical properties, genomic position and the gene structures.

Table 2. Sequence, physicochemical, chromosomal and exon–intron characteristics of the identified PPR genes and their encoded proteins in *Cicer arietinum*.

No.	Gene Name	Chr.	Strand	CDS (bp)	Protein Length (AA)	Protein M. Weight (kDa)	pI	No. of Intron/Exon
1	LOC101513383	4	Forward (+)	2394	798	87.78	6.50	3 introns / 4 exons
2	LOC101511241	5	Reverse (−)	1662	554	60.94	6.50	4 introns / 5 exons
3	LOC101494995	8	Forward (+)	1701	567	63.50	8.50	3 introns / 4 exons
4	LOC101510683	7	Reverse (−)	1377	458	52.26	8.31	3 introns / 4 exons
5	LOC101505926	1	Forward (+)	2094	697	77.84	6.28	4 introns / 5 exons
6	LOC101505319	5	Forward (+)	1887	628	82.11	7.09	4 introns / 5 exons
7	LOC101502837	3	Forward (+)	1275	424	54.52	7.45	4 introns / 5 exons
8	LOC101500993	6	Forward (+)	2631	876	114.09	6.72	5 introns / 6 exons
9	LOC101500923	8	Forward (+)	1818	605	68.00	9.00	4 introns / 5 exons
`10	LOC101500276	7	Forward (+)	2016	671	75.00	8.70	5 introns / 6 exons

11	LOC101498454	7	Forward (+)	2079	692	77.00	8.80	4 introns / 5 exons
12	LOC101494932	8	Forward (+)	1701	567	63.50	8.50	3 introns / 4 exons
13	LOC101494495	5	Forward (+)	1851	617	68.78	8.67	3 introns / 4 exons
14	LOC101493320	5	Forward (+)	2358	786	87.96	5.97	5 introns / 6 exons
15	LOC101492150	2	Forward (+)	2283	761	84.35	6.33	1 intron / 2 exons
16	LOC101508684	6	Forward (+)	1575	525	60.72	6.74	1 intron / 2 exons
17	LOC101507871	8	Forward (+)	2163	721	81.39	8.93	0 introns / 1 exon
18	LOC101504413	1	Forward (+)	3213	1071	121.63	6.77	4–5 introns / 5–6 exons
19	LOC101495760	3	Forward (+)	1662	554	61.0	8.8	1–4 introns / 2–5 exons
20	LOC101499257	4	Forward (+)	1440	480	54.2	9.7	4–5 introns / 5–6 exons
21	LOC101510641	1	Forward (+)	1839	612	68.04	7.55	1 intron / 2 exons
22	LOC101509850	5	Forward (+)	2049	682	75.48	9.20	3 introns / 4 exons
23	LOC101498959	5	Forward (+)	2568	856	255.47	4.91	2 introns / 3 exons
24	LOC101493436	6	Forward (+)	1434	478	383.08	4.83	2 introns / 3 exons

PPR Protein Subcellular Localization and Cis-Acting Regulatory Element Analysis

According to the subcellular localization study, different types of predictions were recorded for 24 PPR proteins. All the proteins were expected to be found in different regions of the cell such as the nucleus, cytoplasm, mitochondria, vacuole, cytoskeleton, chloroplast, endoplasmic reticulum, Golgi apparatus and extracellular space. Among all studied proteins, there were PPR proteins predicted to localize in mitochondria or chloroplast with many others having significantly higher localization in the nucleus, vacuole and cytoplasm among other cellular compartments. The subcellular localization predictions for all 24 PPR proteins can be found in table (3).

Cis-Acting Regulatory Element Analysis

Analysis of the promoters of the identified PPR genes showed that they contain different putative cis-acting regulatory elements related with plant development, hormone responsiveness, light responsiveness and environmental and stress responses. It was found that besides diversity in their types, the regulatory elements differ from each other in their distribution form from gene to gene.

Table 3. Predicted subcellular localization profiles of the identified PPR proteins in *Cicer arietinum.*

No.	Gene Name	Nucl	Cyto	Mito	Vacu	Cysk	Chlo	E. R	Golgi	Extra
1	LOC101513383		1		2		1		1	3
2	LOC101511241			2			9	1		9
3	LOC101494995	4		2			7			1
4	LOC101510683	3		2			5			1
5	LOC101505926				1				2	
6	LOC101505319		1		1				1	3
7	LOC101502837			3	1		1	1		8
8	LOC101500993		1		1			1	2	
9	LOC101500923		1		2				1	3
10	LOC101500276				1			1	2	
11	LOC101498454		1		1				2	3
12	LOC101494932				1				2	
13	LOC101494495		1		1			1	2	

14	LOC101493320	6	2	5						1
15	LOC101492150	6	1		5			1		
16	LOC101508684		1		4		1			4
17	LOC101507871			4		1				1
18	LOC101504413	3	1						1	
19	LOC101495760			2	1		2	2		7
20	LOC101499257			3	1		1	1		8
21	LOC101510641		1		3				1	1
22	LOC101509850		1		2				1	3
23	LOC101498959	5		2			6			1
24	LOC101493436		1	3				1		1

PPR Proteins in *Cicer arietinum*: Phylogenetic, Gene Structure and Conserved Motif Analysis

The 24 identified PPR proteins were put into a multiple sequence alignment in order to reveal conserved and divergent regions of these proteins. The phylogenetic analysis clustered the PPR proteins according to their relationships, as noted that related proteins were clustered closer in terms of sequence similarity compared to the other branches. The relationship of phylogeny to gene structure and conserved motif patterns showed the common and different features of PPR proteins. The phylogeonomy and domain organization of the identified PPR proteins are shown in figure (1).

Gene Structure Analysis

Alignment of sequences of PPR proteins allowed determining conserved and varied regions of the proteins. Phylogenetic analyses indicated that the proteins clustered based on sequence similarity, with closely related sequences being more similar than those found on more distant branches. Combining phylogenetic images with the analysis of gene structural organization of the proteins and their conserved motifs helped define common and varying features of the PPR proteins. The phylogeny and the domain structure of the identified PPR proteins are shown in figure (2).

Conserved Motif Analysis

MEME Suite analysis gave rise to the discovery of ten conserved motifs in the 24 PPR proteins examined in this study. The motifs ranged from 44 to 50 in length, whereby motifs 1, 2, 4, 7 and 8 had a length of 50 amino acids in contrast to motif 3 that was 48, motif 6 and 9 had a length of 44 amino acids and motifs 5 and 10 were 46 amino acids long. The figure below presents the distribution of the studied conserved motifs in the found PPR proteins.

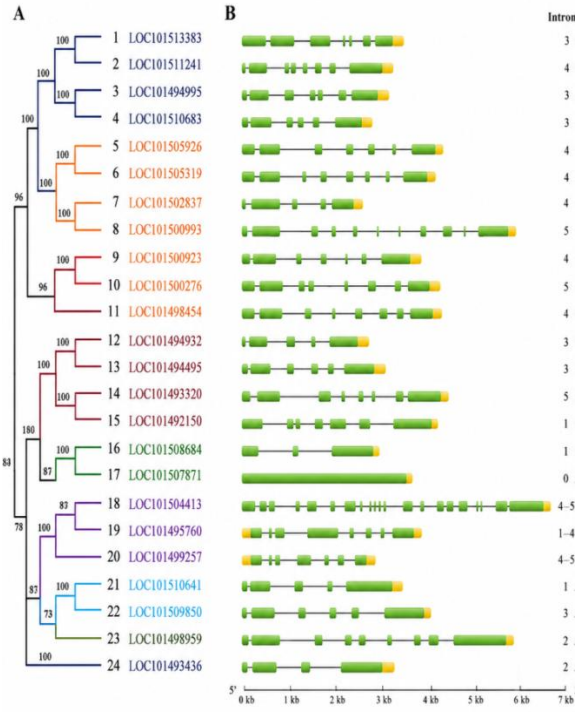


Fig. 1. Gene structure organization of PPR genes in *Cicer arietinum*.

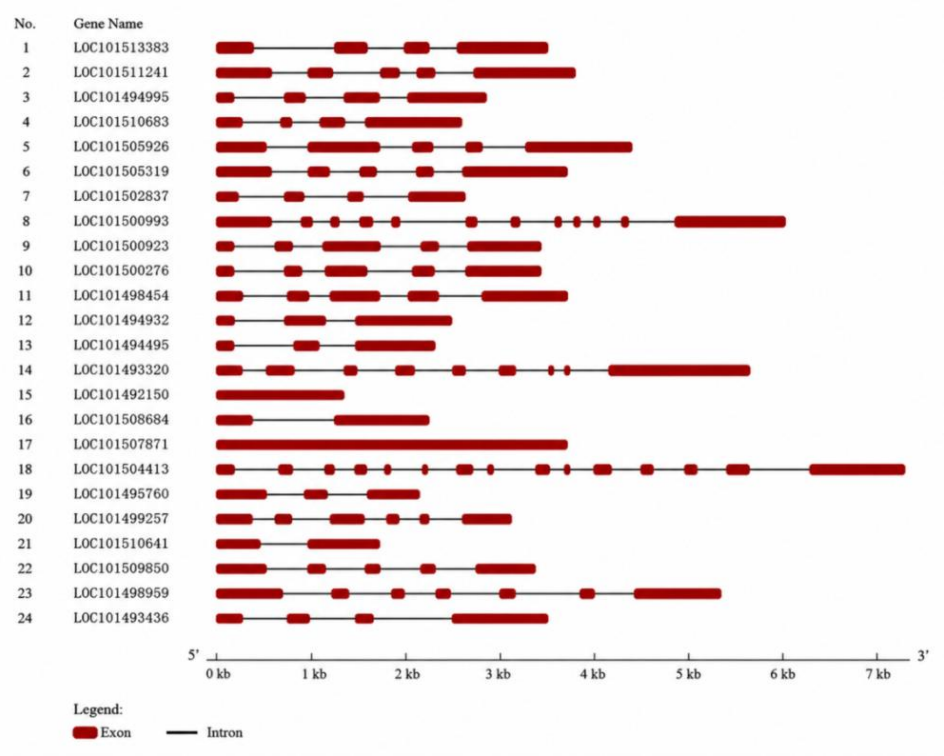


Fig. 2. Phylogenetic relationships and protein domain organization of PPR proteins in *Cicer arietinum*.

In figure (3), the distribution and position of conserved motifs in the PPR protein sequences have been illustrated. While some proteins contain several conserved motifs, the types, order of appearance of each conserved motif differs in various proteins of the PPR family. Hence, there is a conservation of certain sequence regions of PPR proteins and variations in the order of motifs present in its protein sequences. Table (4) below provides the respective motifs and their distribution across the studied PPR protein family.

Phylogenetic and Gene Duplication Analysis of the PPR Gene Family

A phylogenetic analysis was carried out to determine the evolutionary connections among the 24 PPR proteins that were found in *Cicer arietinum*. The phylogenetic tree revealed the cluster arrangement of proteins according to their sequence similarities. Proteins grouped together in the same cluster exhibited closely related evolution, while proteins coming from different clusters differed more in their evolution. This clustering indicates that there is the diversification of the family of PPR proteins, while conserved relations exist among some members of the family, as evidenced in figure (4).

The evolutionary relationship depicted is indicative of possible gene duplication events that may occur in the PPR gene family. The PPR proteins that are evolutionarily related but belong to the same phylogenetic clusters may suggest a possible duplication event followed by sequence divergence. Gene duplication could also explain the emergence and expansion of the PPR family in *C. arietinum*. Still, phylogenetic relationships alone cannot ascertain the type of duplication and the time of occurrence of such events and thus need to be correlated to the sequence similarities and genomic information.

More in-depth pairwise sequence similarity testing involving the 24 PPR proteins was done and represented as a heat map in figure (5). These results indicate a certain degree of sequence similarity among the PPR proteins with both highly similar and differentially sophisticated pairs. Generally, greater similarity levels can be attributed to the close phylogenetic relationship established earlier in the study and may provide an insight into the quantitative dimension of the PPR sequence similarities.

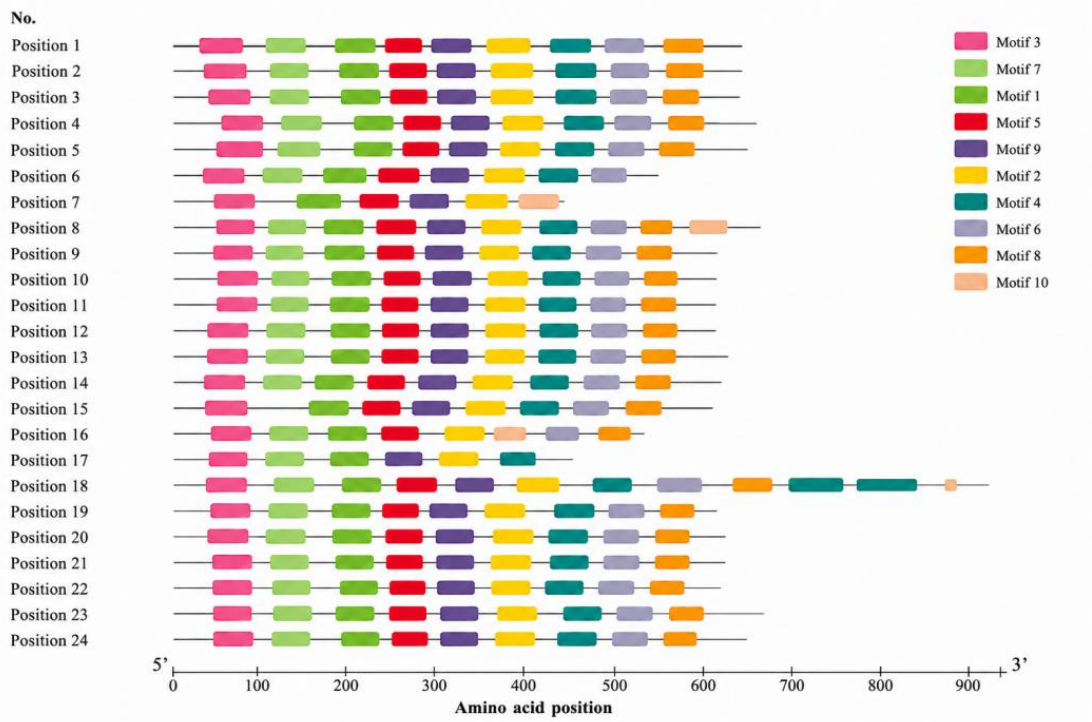


Fig. 3. Conserved motifs of PPR proteins in *Cicer arietinum* shown schematically using information obtained from MEME Suite**.**

Table 4. Conserved motifs identified in the PPR proteins of Cicer arietinum using the MEME suite, with representative motif sequences and corresponding motif widths.

Motif No	Most suitable matches	Width
1	QRVFNETPEKTQFSWTGMISAYARHGDYESVMGLYKEMERESCSPDSITF	50
2	HPDIVSWNTILSGFDKSVDALNFACFMRLNGIVFDPVTYTTALSFCWDDD	50
3	KCDVVRDAKEVFWVMSNRNVVSWTTMISIDEDDAVSLFNAMRIDGVYPND	48
4	DVGHRIFDTMVKKHSIEPTPEHYSIMVDMLGRVGRLNEAEELMHRIPGGP	50
5	QESKKVFEELNYQGTISWNSLISGYAQNGLYKEAFLTFLSAIKEIKPNQY	46
6	IQMDPASSGPYVLMANLYAAKGNWEKVAQVRKGMRGRGVKKEVGFSWVDV	44
7	CEVFIGNALVTMYSRLEELDDARRVFDEMPIRDLVSWNAMLSGYAQEGEC	50
8	QTSFKTCSKLLHSLKHEHEHHLFDNMPHLNVSSINRSMLNFLHQNLPFQA	50
9	NMVRQGMLLDHVSFTGAVSACGYMRNLELGKQIHGLAQKVGYGTHVAVCN	44
10	FKACNGEFNLGSQIHGFAVSTGFVSRIAVSNSLMKMYCKADKFKCALRVF	46

Motif analysis showed recurring sequence patterns among the analyzed proteins. The occurrence and distribution of individual motifs varied among members of the family.

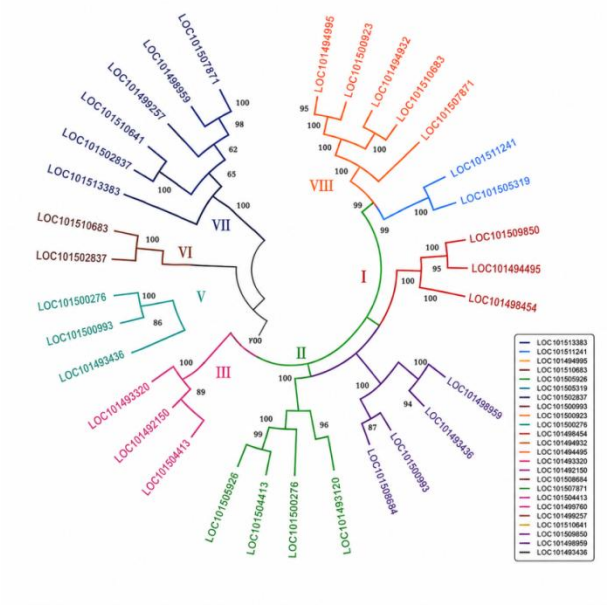


Fig. 4. Phylogenetic tree showing the evolutionary relationships among the 24 PPR proteins identified in *Cicer arietinum*.

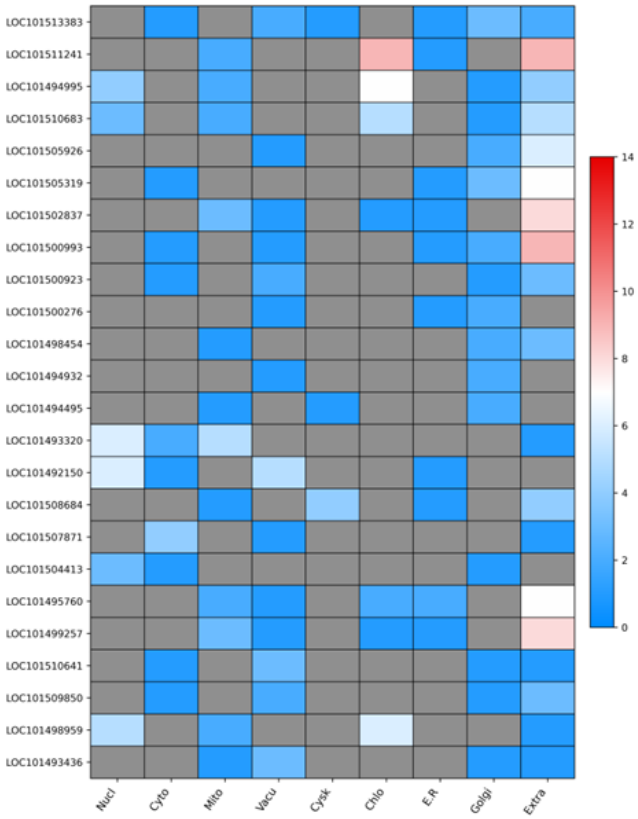


Fig. 5. Pairwise sequence similarity heatmap showing the sequence relationships among the 24 PPR proteins identified in *Cicer arietinum* L.

In conclusion, it can be said that the combined analysis of phylogeny and sequence similarity reveals that *C. arietinum* PPR proteins demonstrate conserved evolutionary relationships as well as considerable diversification of their sequences. The closely related PPR proteins identified through phylogenetic and sequence similarity analyses may reflect potential duplication relationships followed by sequence divergence. However, confirmation of specific duplication events requires additional evidence, including genomic location, syntenic relationships and analyses capable of distinguishing between tandem and segmental duplication. However, the proof of some duplication events will require some more evidence such as genomic location of the proteins, evidence from syntenic relationships and criteria that will allow one to distinguish between tandem and segmental duplication types.

DISCUSSION

The investigation provides a comprehensive computational analysis of twenty-four PPR genes found in *Cicer arietinum* L. It was found that the PPR genes showed considerable variability in various aspects of their composition such as length of CDSs, properties of proteins, structure of exons and introns, predicted sites of localization, presence of conserved motifs and analysis of phylogeny. The changes concerning CDS length and length of the proteins indicate high diversity of the PPR genes under study. Particularly, the length of the proteins varied from 424 amino acids to 1071 that is, the longest protein is more than twice as long as the shortest one. Such kinds of changes are likely to be caused by differences in the number of PPR repeats and other functional domains of the proteins (Chen *et al.,* 2018b). While differences in predicted molecular weight were mainly attributed to differences in the length of proteins, the observed difference may arise from different amino acid composition and sequence characteristics of the proteins analyzed.

The theoretical pI values are indicative of the overall charge of the proteins which ranged from acidic to basic. This variation can affect the interactions of the concerned proteins with other proteins or protein RNA combinations. Despite the fact that calculations made based on the theoretical pI values lead to obtaining estimates, these values cannot provide us with the information connected to whether the PPR proteins function differently or not. Analysis of exons and introns revealed that genes of the PPR family have architecture that is different. LOC101507871 gene that has no introns and other genes that have different number of introns show that PPR genes do not follow any uniform organization with respect to their genes published. Different structures of the exons - intron gene architecture have helped to trace evolutionary pathways of the gene families with regard to their intron gains and losses, gene duplications, or sequence rearrangements (Qiao *et al.,* 2019). But the method used in this particular study does not allow us to see the true reason behind those differences.

The predicted subcellular localization analysis provides insight into the possible cellular distribution of PPR proteins. A number of PPR proteins appeared to be localized to mitochondria or chloroplasts, consistent with previous studies of organelle-associated PPR proteins in plants (Wang *et al.,* 2021). This predicted localization is relevant from the biological perspective because plant PPR proteins are known to have a role in several processes of organelles RNA, such as RNA editing, processing, stabilization, splicing and translation. Hence, PPR proteins localized to mitochondria and chloroplasts could help discover more about their operation in the mechanism of regulation of organelles RNA. On the other hand, these predictions should be treated cautiously because localization score for the computer analysis does not mean localized proteins and requires experimental confirmation.

The promoter analysis yielded different putative cis-acting regulatory elements involved in the processes of plant development, response to

hormones, response to light and environmental signal perception. This suggests that each gene encoding the PPR proteins has its own regulatory activity mechanisms. This diversity may lead to differential expression of the PPR genes. Although the presence of a cis-element in the promoter suggests that this element is functional it does not mean it is so. Therefore, experimental validation of promoter activity should be carried out. The phylogenetic analysis showed an evolutionary relationship between the 24 PPR proteins. The analysis results showed that evolutionarily close proteins were clustered in one branch, while divergently evolved proteins were placed in different branches. Integrating the phylogenetic and motif analysis approaches makes evidence of family diversification much stronger than evidence based solely on the sequence of proteins.

Ten conserved motifs were identified across the PPR protein dataset, although their presence, order, and distribution varied among family members with a length of 44 to 50 amino acids. The presence of repeated sequence motifs means that some parts of PPR proteins have been conserved during evolution. However, there is also variation in how motifs occur and are distributed, which signifies variation in the sequences among the family members. Such conservation and divergence correlate with the modularity of PPR proteins, as those proteins consist of repeating units of sequences that ensure the function of recognition of RNA and perform different RNA-related functions in organelles organelles (Yan *et al.,* 2019). Nevertheless, it is not correct to assign any biochemical functions to the MEME motifs just because it found signs of conservation. MEME is used to find shared sequence patterns only and therefore the role of motifs in the functional analysis should be studied along with information derived from the analysis of Pfam/SMART domains, phylogenetic data, structural data and results of experimental studies.

The presence of PPR genes across all 8 chickpea chromosomes indicates their widespread existence in the genome. Differences in chromosomal presence may be due to duplication of PPR genes, chromosomal rearrangements, segmental duplications, or various retention of genes through the process of evolution. Duplication of genes is a key event in the processes when gene families evolve (Qiao *et al.,* 2019). After duplication, the gene copies may perform the same function, differ in expression, or both (Feng *et al.,* 2023). The repetitive and modular structure of PPR proteins may be among the contributing factors in the evolution of their capability of recognizing RNA. The analysis of the gene structure in combination with the molecular phylogeny and the patterns of the motifs can certainly help to reveal the story of evolution of the gene family. As for the study conducted, genes of the same family showed similarities in exon-intron organization yet differed in the motifs, which indicates the evolutionary possibilities of the family members.

The PPR gene group analyzed reveals both structural conservation and molecular variation in *Cicer arietinum*. The occurrence of conserved motifs, as well as phylogenetic relationships, is indicative of conservation of some features in the sequence and related functions of some of the members of the PPR family group. Differences in protein size, theoretical pI, exon-intron pattern, potential subcellular location and motif composition demonstrate a diversification among individual members of the family group (Xiao *et al.,* 2018). The characterization of the PPR gene family in chickpea should have further biological and agricultural relevance as mitochondria and chloroplasts are responsible for cellular processes such as respiration, photosynthesis, energy metabolism, development and responses to environmental conditions (Ali *et al.,* 2024). Therefore, the PPR genes described in this work may be of further interest in functional studies to clarify their roles in organellation process and plant development.

Nonetheless, it is important to note that the study's findings rely on computational analysis. To that end, predicted subcellular localization,

cis-elements, conserved motifs and potential duplication relationships are best understood as hypotheses rather than established biological processes. Future studies that utilize transcriptomic analysis, tissue-specific expression profiling, stress-responsive expression profiling, experimental methods for subcellular localization, RNA-binding tests, gene-editing technologies and functional complementation may provide valuable insight into the specific functions of PPR genes in chickpea.

CONCLUSION

The 24 PPR genes that were analyzed were distributed throughout chromosomes 1 through 8 and exhibited significant variability in their CDS lengths, proteins sizes, molecular weights, isoelectric points, exon-intron structures, subcellular localizations, conserved motifs and phylogenetic relations. They were characterized by their CDS lengths and the corresponding proteins which varied between 1,275 to 3,213 bp and 424 -1,071 amino acids, respectively. The theoretical pI values of PPR proteins were from 4.83 to 9.70 indicating that their physicochemical properties varied significantly. The analysis of gene structure shows that there are some variations in exon-intron structures thus indicating that comparative gene structure analysis has been helpful in identifying novel members of the PPR family; it is worth noting that the exon-intron structure varies from genes with no introns and a single exon to those with many exons and introns. Besides, the predicted subcellular localization reveals some differences among PPR family members in that some PPR proteins were predicted to be located in mitochondria and chloroplasts thus reflecting the known function of PPR proteins in organelle RNA metabolism. According to MEME analysis, there are 10 conserved motifs varying from 44 to 50 amino acids. The fact that shared motifs are present in many family members indicates that certain features have remained unchanged in terms of structure or function, whereas variations in the composition of certain motifs demonstrate that the PPR family has become functionally more diverse.

The established phylogenetic relationships based on sequences of the PPR proteins show that their phylogenetic grouping is accompanied by similarities of their exon/intron organization and motifs distribution, so it can be concluded that PPR genes have been conserved and evolved differently in the case of chickpeas. Apart from this, the presence of PPR family genes in all eight chromosomes proves that they occupy a wide area of the genome of C. arietinum. To summarize, the research carries out a thorough study of the genomic structure, molecular characteristics, preserved trait, and common evolutionary relations of PPR genes in chickpeas.

ACKNOWLEDGEMENT

The author wishes to acknowledge the authorities of the Department of Biotechnology University of Okara, Pakistan for their support.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

GENERATIVE AI STATEMENT

The author(s) declare that no Generative AI was used in the creation of this manuscript.

PUBLISHER'S NOTE

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. The publisher does not guarantee or

endorse any product that may be reviewed in this article or any claim made by its manufacturer.

REFERENCES

Ali, N.A., Song, W., Huang, J., Wu, D., Zhao, X., 2024. Recent advances and biotechnological applications of RNA metabolism in plant chloroplasts and mitochondria. Crit. Rev. Biotechnol., 44(8):1552-73.

Castro, P., Carmona, A.F., Pérez-Rial, A., Millán, T., Rubio, J., Gil, J., Die, J.V., 2024. Finding consensus on the reference genomes: A chickpea case study. Legume Sci., 6.

Chen, G., Zou, Y., Hu, J., Ding, Y., 2018a. Genome-wide analysis of the rice PPR gene family and their expression profiles under different stress treatments. BMC Genom., 19(1): 720.

Chen, L., Li, Y., Li, C., Shi, Y., Song, Y., Zhang, D., Li, Y., Wang, T., 2018b. Genome-wide analysis of the pentatricopeptide repeat gene family in different maize genomes and its important role in kernel development. BMC Plant Biol., 18(1): 366.

Die, J.V., Gil, J., Millán, T., 2018. Genome-wide identification of the auxin response factor gene family in Cicer arietinum. BMC Genom., 19(1): 143.

Feng, L.-Y., Lin, P.-F., Xu, R., Kang, H., Gao, L.-Z., 2023. Comparative genomic analysis of Asian cultivated rice and its wild progenitor (*Oryza rufipogon*) has revealed evolutionary innovation of the pentatricopeptide repeat gene family through gene duplication. Int. J. Mol. Sci., 24.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005. Protein identification and analysis tools on the ExPASy server. In: Walker, J.M. (Ed.), The Proteomics Protocols Handbook, 571-607. Humana Press.

Gorchs Rovira, A., Smith, A.G., 2019. PPR proteins: Orchestrators of organelle RNA metabolism. Physiol. Plant., 166(1): 451-459.

Gutmann, B., Royan, S., Schallenberg-Rüdinger, M., Lenz, H., Castleden, I., McDowell, R., Vacher, M., Tonti-Filippini, J., Bond, C.S., Knoop, V., Small, I., 2020. The expansion and diversification of pentatricopeptide repeat RNA editing factors in plants. Mol. Plant., 13(2): 215-230.

Gutmann, B., Millman, M.R.C., Sanglard, L.V.P., Small, I., Colas des Francs-Small, C., 2021. The pentatricopeptide repeat protein MEF100 is required for the editing of four mitochondrial editing sites in Arabidopsis. Cells., 10.

He, P., Wu, S., Jiang, Y., Zhang, L., Tang, M., Xiao, G., Yu, J., 2019. GhYGL1d, a pentatricopeptide repeat protein, is required for chloroplast development in cotton. BMC Plant Biol., 19.

Hao, Y.Y., Zhao, X.Q., Huang, F.D., Li, C.S., 2021. The role of PPR proteins in posttranscriptional regulation of organelle components in plants. Yi Chuan., 43(11): 1050-1065.

Kumar, S., Stecher, G., Suleski, M., Sanderford, M., Sharma, S., Tamura, K., 2024. MEGA12: Molecular Evolutionary Genetic Analysis version 12 for adaptive and green computing. Mol. Biol. Evol., 41(12): msae263.

Li, J., Zhang, L., Li, C., Chen, W., Wang, T., Tan, L., Qiu, Y., Song, S., Li, B., Li, L., 2025. The pentatricopeptide repeat protein OsPPR674 regulates rice growth and drought sensitivity by modulating RNA editing of the mitochondrial transcript ccmC. Int. J. Mol. Sci., 26.

Miglani, G.S., Kaur, M., Manchanda, P., 2021. Pentatricopeptide repeat proteins: I. Involvement in RNA editing in plants. J. Crop Improv., 35.

Oldenkott, B., Yang, Y., Lesch, E., Knoop, V., Schallenberg-Rüdinger, M., 2019. Plant-type pentatricopeptide repeat proteins with a DYW domain drive C-to-U RNA editing in Escherichia coli. Commun. Biol., 2: 85.

Qiao, X., Li, Q., Yin, H., Qi, K., Li, L., Wang, R., Zhang, S., Paterson, A.H., 2019. Gene duplication and evolution in recurring

polyploidization-diploidization cycles in plants. Genome Biol., 20(1): 38.

Small, I., Schallenberg-Rüdinger, M., Takenaka, M., Mireau, H., Ostersetzer-Biran, O., 2020. Plant organellar RNA editing: What 30 years of research has revealed. Plant J., 101(5): 1040-1056.

Su, H.-G., Li, B., Song, X., Ma, J., Chen, J., Zhou, Y., Chen, M., Min, D., Xu, Z.-S., Ma, Y.-Z., 2019. Genome-wide analysis of the DYW subgroup PPR gene family and identification of GmPPR4 responses to drought stress. Int. J. Mol. Sci., 20.

Subburaj, S., Tu, L., Lee, K., Park, G.-S., Lee, H., Chun, J.-P., Lim, Y.-P., Park, M., McGregor, C., Lee, G.-J., 2020. A genome-wide analysis of the pentatricopeptide repeat (PPR) gene family and PPR-derived markers for flesh color in watermelon (Citrullus lanatus). Genes, 11.

Tamura, K., Stecher, G., Kumar, S., 2021. MEGA11: Molecular Evolutionary Genetics Analysis version 11. Mol. Biol. Evol., 38(7): 3022-3027.

Varshney, R.K., Roorkiwal, M., Sun, S., Bajaj, P., Chitikineni, A., Thudi, M., et al., 2021. A chickpea genetic variation map based on the sequencing of 3,366 genomes. Nat., 599(7886): 622-627.

Wang, X., An, Y., Xu, P., Xiao, J., 2021. Functioning of PPR proteins in organelle RNA metabolism and chloroplast biogenesis. Front. Plant Sci., 12: 627501.

Wang, Y., Li, H., Huang, Z.-Q., Ma, B., Yang, Y.-Z., Xiu, Z., Wang, L., Tan, B.-C., 2022. Maize PPR-E proteins mediate RNA C-to-U editing in mitochondria by recruiting the trans deaminase PCW1. Plant Cell, 34.

Wang, Y., Tan, B.-C., 2024. Pentatricopeptide repeat proteins in plants: Cellular functions, action mechanisms, and potential applications. Plant Commun., 5.

Xiao, H., Zhang, Q., Qin, X., Xu, Y., Ni, C., Huang, J., Zhu, L., Zhong, F., Liu, W., Yao, G., Zhu, Y., Hu, J., 2018. Rice PPS1 encodes a DYW motif-containing pentatricopeptide repeat protein required for five consecutive RNA-editing sites of nad3 in mitochondria. New Phytol., 220(3): 878-892.

Xing, H., Fu, X., Yang, C., Tang, X., Guo, L., Li, C., Xu, C., Luo, K., 2018. Genome-wide investigation of pentatricopeptide repeat gene family in poplar and their expression analysis in response to biotic and abiotic stresses. Sci. Rep., 8.

Yan, J., Yao, Y., Hong, S., Yang, Y., Shen, C., Zhang, Q., Zhang, D., Zou, T., Yin, P., 2019. Delineation of pentatricopeptide repeat codes for target RNA prediction. Nucleic Acids Res., 47(7): 3728-3738.

Zhou, Z., Sun, J., Wen, M., Liu, A., Yu, A., 2025. Genome-wide analysis of pentatricopeptide repeat genes in castor and their potential functions in seed development. Ind. Crops Prod., 227.