The Pharma Innovation Journal 2022; 11(8): 2173-2180

www.ThePharmaJournal.com

The Pharma Innovation

**ISSN (E): 2277-7695**
**ISSN (P): 2349-8242**
**NAAS Rating: 5.23**
TPI 2022; 11(8): 2173-2180
© 2022 TPI
www.thepharmajournal.com
Received: 04-05-2022
Accepted: 30-07-2022

**Mathuranthagi TA**
Department of Plant Biotechnology, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu, India

**Kumar KK**
Department of Plant Biotechnology, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu, India

**Caroline Nirmala R**
Department of Plant Molecular Biology and Bioinformatics, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu, India

**Varanavasiappan S**
Department of Plant Biotechnology, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu, India

**Muthuswamy A**
Agricultural College and Research Institute, Tamil Nadu Agricultural University, Karur, Tamil Nadu, India

**Corresponding Author:**
**Kumar KK**
Department of Plant Biotechnology, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu, India

# Functional and phylogenetic analysis of prolamin proteins of foxtail millet using computational approaches

**Mathuranthagi TA, Kumar KK, Caroline Nirmala R, Varanavasiappan S and Muthuswamy A**

**DOI:** https://doi.org/10.22271/tpi.2022.v11.i8aa.15178

**Abstract**
Prolamin, the alcohol-soluble seed storage protein is the major seed protein in many cereals. Foxtail millet (*Setaria italica*) contains prolamin (Setarin) as major seed storage protein, which accounts > 50% of the total seed storage protein. To *in-silico* characterize the prolamin proteins of foxtail millet, sequence coding for 14.3 kDa, 20.4 kDa, 22 kDa, 22.4 kDa and 26 kDa prolamin of foxtail millet were retrieved from NCBI database. Protein physio-chemical characterization was done through Protparam and found that all the prolamin proteins were deficient in essential amino acids like cysteine, lysine, but contains relatively higher amount of essential amino acid, tryptophan. Isoelectric point of >7 indicated that all the proteins were basic in nature and low GRAVY value indicated that these proteins were hydrophobic in nature. Functional characterization by PRPSITE and Pfam revealed that all the proteins belong to the zein family of prolamins. Alpha helix was the dominant secondary structure found in all proteins. Multiple sequence alignment between foxtail millet prolamins show series of tandem repeats of glutamine (Q) residues. Phylogenetic analysis revealed that *Setaria viridis* were closely related with all the foxtail millet prolamins followed by other cereals. High level sequence homology was observed in signal peptide sequence of all the foxtail millet prolamins and signal peptide was prediction to cleave between 21 to 22 amino acids. The *in-silico* characterization of foxtail millet prolamins is able to assess the nutritional quality of prolamins, in addition to predicting the secondary structure and phylogenetic relationship with other cereal prolamins.

**Keywords:** Prolamin, *In-silico*, foxtail millet, multiple sequence alignment, phylogenetic analysis

## 1. Introduction
Plant seeds are the primary tissue taken out by humans. The most prevalent source of protein in the diet is seed of cereals and pulses. During seed developmental stages, plants store materials including starch, lipids, and proteins. Storage proteins build up in both vegetative tissues called as vegetative storage proteins (VSPs) and reproductive tissues (seeds) called as seed storage proteins (SSPs). SSPs are a group of proteins that build up in seeds at high levels during the last stages of seed development.

In addition to serving as a supply of nitrogen and sulphur for sprouting seedlings, seed storage proteins (SSPs) also serve as a source of protein for both human and animals. People in developing nations essentially rely solely on seed protein to meet their full protein needs rather than from animal sources which are expensive (Mandal and Mandal 2000) [15]. Seed storage proteins are divided into classes depending on their solubility i.e., albumins in water, globulins in diluted saline, prolamins in alcohol, and glutelins in dilute acid or alkali (Osborne 1924) [17]. Prolamins are the primary endosperm storage proteins of all cereal grains, except oats and rice. Prolamins range greatly in molecular mass from about 10 kDa to approximately 100 kDa. They are distinguished by their high glutamine and proline content and have poor lysine and tryptophan content (Shewry *et al*., 1999) [18]. Distinct cereals have different names for prolamins, including gliadin in wheat, hordein in barley, secalin in rye, zein in maize, Kafirin in sorghum, avenin in oats and setarin in foxtail millet. Prolamins have a low solubility in water. However, prolamin sub-families were categorised based on variations in water solubility and capacity for disulfide bonds. For example, there are four types that the zeins and Kafirins are divided into: α, β, γ and δ (Holding 2014) [4]. A S-rich (sulphur rich) Barley's B-hordeins contain a significant amount of glutamine-proline and a relatively high amount of cysteine, whereas the S-poor (sulphur poor) C-hordeins are deficient in cysteine and contain very little methionine and lysine (Shewry *et al*., 1981) [19].

Similar to this, the level of lysine, arginine, histidine, and tryptophan is low in α, β and γ-gliadins of wheat, α, β and γ-avenins of oats and α, β, γ and δ zeins of maize (Arvinder *et al.*, 2009) [1]. Low lysine and tryptophan concentration degrades the nutritional quality of the prolamin proteins. A study reported that N-terminal amino acid sequences of six prolamin gene encoded by alleles present in *Pro1* and *Pro2* of foxtail millet were classified into three groups which were homologous to prolamins of pearl millet, *Echinochloa crus-galli* and α-zein of maize, sorghum (Nakayama *et al*., 1999) [16]. Minor millets are renowned for their high-quality proteins and nutritional benefits.

*Setaria italica* (L) is commonly known as foxtail/ Italian millet. It is one of the oldest domesticated diploids (2n = 2x = 18), second most produced millet after pearl millet and cultivated in 26 countries in the world. Prolamin is the major seed storage protein of foxtail millet. Essential amino acid composition and phylogenetic analysis of foxtail millet prolamin proteins with the prolamin proteins of other cereals were studied by many scientists. Phylogenetic analysis revealed that seed storage proteins of foxtail millet were closely related to SSPs of sorghum and maize (Brutnell *et al*., 2010) [2]. The motif "NPAAFWQQQQLL" was uniquely repeated in the foxtail millet high tryptophan prolamin protein and foxtail millet 10 kDa prolamin protein contains 9.4% phenylalanine content and 2.5% tryptophan content (Gaur *et al*., 2018) [3]. Hence *in-silico* identification and characterization of different prolamin proteins of foxtail millet have been done in the current study.

## 2. Materials and Method

### 2.1 Retrieval of protein sequences

Sequence of different prolamins in foxtail millet (*Setaria italica*) were retrieved in FASTA format from NCBI protein database (http://www.ncbi.nlm.nih.gov/) [5]. They are KQL15786.1 – 14.3 kDa, KQK95359.1 – 20.4 kDa, KQK95356 – 22 kDa, KQL15787.1 – 22.4 kDa and KQK95355.1 – 26 kDa.

### 2.2 Physio-chemical characterization

The ProtParam server from Expasy (https://web.expasy.org/protparam) [6] was used to characterise the physio-chemical properties of protein sequences and determine the amino acid composition of the different prolamin proteins of foxtail millet. The isoelectric point (pI), the instability index (II), the aliphatic index (AI), and the grand average hydropathy (GRAVY) were calculated.

### 2.3 Functional characterization

Functional characterization of prolamin proteins was done using Expasy PRPSITE server (https://PRPSITE.expasy.org) [7]. Protein family, domain and functional site data were computed using the PRPSITE server. Pfam analysis (http://pfam.xfam.org/search/sequence) [8] was carried out to characterise the chosen prolamin proteins in relation to the particular protein family.

### 2.4 Secondary structure prediction

Secondary structural characteristics of different prolamin proteins of foxtail millet was calculated using SOPMA (Self Optimized Prediction Method with Alignment) method with their default parameters like Similarity threshold 8 and Window width 17 (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa%20_sopma.html) [9]. It uses the amino acid sequence to determine the location of the beta-strands and alpha-helix.

### 2.5 Multiple sequence alignment

Multiple sequence alignment was done by using Clustal omega to generate alignment between all the prolamin proteins of foxtail millet and to determine the conserved domains.

### 2.6 Phylogenetic analysis of the sequences

The phylogenetic analysis was done to study the similarity and distinctiveness between proteins of closely related species. Homologous sequences for foxtail millet were identified by performing BLASTp against protein sequences in NCBI database. Multiple sequence alignment for the individual prolamin proteins with other closely related prolamin proteins was done using MUSCLE (Multiple Sequence Comparison by Log-Expectation) tool to identify conserved domains. Phylogenetic analysis was done using MEGAX tool (https://www.megasoftware.net/dload_mac_beta) [10] using neighborhood - joining method with 100 bootstrap replications by Poisson model.

### 2.7 Signal peptide prediction

All living things are regulated by signal peptides (SPs), which are short amino acid sequences. SPs are short N-terminal amino acid sequences that target proteins to the secretory (Sec) pathway for translocation across the endoplasmic reticulum in eukaryotes and plasma (inner) membrane in prokaryotes. Signal peptides of all the proteins were predicted by using Signal P 5.0 software (https://services.healthtech.dtu.dk/service.php?SignalP-5.0) [11].

## 3. Results and Discussion

### 3.1 Physio-chemical characterization:

Amino acid composition of all prolamin sequences of foxtail millet were computed using Expasy Protparam tool (Table 1). Prolamin proteins studied here were rich in certain amino acids like Ala, Gln, Leu, and Pro as reported earlier (Kohama *et al*., 1999) [12]. The prolamin proteins were deficient in sulphur containing amino acids like cysteine and methionine and were deficient in essential amino acid like lysine, but contain higher proportion of tryptophan.

**Table 1:** Amino acid composition of prolamin proteins of foxtail millet

| Amino Acids | KQL15786.1 (14.3 kDa) | KQK95359.1 (20.4 kDa) | KQK95356 (22 kDa) | KQL15787.1 (22.4 kDa) | KQK95355.1 (26 kDa) |
|---|---|---|---|---|---|
| Ala (A) | 18.7% | 20.6% | 13.3% | 17.6% | 15% |
| Arg (R) | 0.7% | 0.5% | 1.5% | 2.0% | 2.5% |
| Asn (N) | 4.5% | 4.6% | 5.4% | 5.4% | 5.4% |
| Asp (D) | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| Cys (C) | 0.0% | 0.5% | 1.0% | 0.0% | 1.2% |
| Gln (Q) | 13.4% | 17.0% | 17.2% | 18.1% | 17.1% |
| Glu (E) | 0.7% | 0.0% | 1.0% | 0.5% | 0.4% |
| Gly (G) | 3.0% | 1.0% | 1.0% | 2.0% | 2.1% |
| His (H) | 2.2% | 0.5% | 2.0% | 1.0% | 1.7% |
| Ile (I) | 6.0% | 3.6% | 3.0% | 4.9% | 4.2% |
| Leu (L) | 14.2% | 19.6% | 17.2% | 13.7% | 17.1% |
| Lys (K) | 0.7% | 0.5% | 0.5% | 0.5% | 0.4% |
| Met (M) | 2.2% | 1.0% | 3.4% | 1.0% | 2.5% |
| Phe (F) | 6.7% | 4.6% | 3.4% | 4.9% | 3.8% |
| Pro (P) | 9.7% | 8.2% | 8.9% | 7.8% | 8.3% |
| Ser (S) | 6.7% | 7.7% | 6.4% | 6.9% | 6.2% |
| Thr (T) | 3.0% | 3.6% | 4.9% | 3.4% | 2.9% |
| Trp (W) | 1.5% | 0.5% | 2.0% | 2.5% | 2.1% |
| Tyr (Y) | 1.5% | 1.5% | 1.5% | 3.4% | 2.1% |
| Val (V) | 4.5% | 4.1% | 6.4% | 4.4% | 5.0% |

The physio-chemical characterization was computed by Expasy's ProtParam (Table 2). The total number of amino acid residues for five prolamin proteins studied ranges from 134 to 240 with different molecular weight. The isoelectric point computed for these proteins were greater than 7 indicated that they are basic in nature. Aliphatic index of all the prolamin proteins ranges from 103.09 to 123.04. The Instability Index value for all the prolamin proteins ranges from 50.46 to 69.27, which indicates that all the proteins were unstable. Positive GRAVY indices of the prolamin proteins under investigation suggested that they were hydrophobic in nature. Proteins with GRAVY score above 0 are more likely to be hydrophobic proteins (Magdeldin *et al*., 2012) [14].

**Table 2:** Parameters computed using Expasy's ProtParam tool

| S No | Protein Name | No of amino acids | MW (kDa) | PI | II | AI | Gravy |
|---|---|---|---|---|---|---|---|
| 1 | KQL15786.1 | 134 | 14.3 | 8.31 | 50.46 | 110.22 | 0.500 |
| 2 | KQK95359.1 | 194 | 20.4 | 9.10 | 69.27 | 123.04 | 0.548 |
| 3 | KQK95356 | 203 | 22.4 | 8.81 | 65.99 | 110.64 | 0.238 |
| 4 | KQL15787.1 | 204 | 22.4 | 9.77 | 59.97 | 103.09 | 0.143 |
| 5 | KQK95355.1 | 240 | 26.4 | 9.81 | 63.99 | 112.38 | 0.258 |

Where, MW: Molecular weight; pI: Isoelectric Point; II: Instability Index; AI: Aliphatic Index; GRAVY: Grand Average Hydropathicity

### 3.2 Functional characterization

For functional characterization of prolamin proteins, the prediction of pattern, profile, and Pfam analysis were done (Table 3). Pattern and profile of the proteins were analysed using Expasy's PRPSITE. PRPSITE is a database of annotations for motif descriptors used to identify protein families and domains (Sigrist *et al.*, 2002) [20]. Patterns were found in all the proteins which includes protein kinase C phosphorylation site and N-myristoylation site in KQK95355.1 and KQK15787.1, and N-myristoylation site alone in other three proteins. Profile was found in four proteins out of five. All the four prolamin proteins (KQK95356, KQK95355.1, KQL15787.1 and KQK95359.1) have glutamine-rich region profile. Pfam analysis was done to five prolamin proteins and found that they belong to the zein family of prolamin.

**Table 3:** Prediction of patterns and profile by using PRPSITE and Pfam analysis

| S. No | Protein ID | Name of protein (as available on NCBI) | Pfam Analysis | Patterns by PRPSITE | Profile by PRPSITE |
|---|---|---|---|---|---|
| 1 | KQL15786.1 (14.3 kDa) | Hypothetical protein SETIT_024170mg | Zein | N-myristoylation site (27 - 32) | - |
| 2 | KQK95359.1 (20.4 kDa) | Hypothetical protein SETIT_028022mg | Zein | N-myristoylation site (47 - 52) | Glutamine-rich region (66 - 186) |
| 3 | KQK95356 (22 kDa) | Hypothetical protein SETIT_027546mg | Zein | N-myristoylation site (164 - 169) | Glutamine-rich region (57 - 180) |
| 4 | KQL15787.1 (22.4 kDa) | Hypothetical protein SETIT_023272mg | Zein | Protein kinase C phosphorylation site (54 - 56) N-myristoylation site (27 - 32) | Glutamine-rich region (74 - 178) |
| 5 | KQK95355.1 (26 kDa) | Hypothetical protein SETIT_026796mg | Zein | Protein kinase C phosphorylation site (50 - 52) N-myristoylation site (144 - 149) | Glutamine-rich region (70 - 214) |

### 3.3 Secondary structure prediction

Secondary structure of all the prolamin proteins were predicted by using SOPMA. Random coils and alpha helix were the dominant secondary structure elements, followed by extended strands and beta turns, according to SOPMA (Table 4). The percentage of alpha helix was more in all prolamin proteins when compared to other secondary structure elements.

**Table 4:** Predicted secondary structures present in prolamin proteins of foxtail millet

| S. No | Protein | Amino acids | Alpha helix | Extended strand | Beta turn | Random coil |
|---|---|---|---|---|---|---|
| 1 | KQL15786.1 (14.3 kDa) | 134 | 58.96% | 5.97% | 2.24% | 32.84% |
| 2 | KQK95359.1 (20.4 kDa) | 194 | 61.86% | 5.67% | 4.64% | 27.84% |
| 3 | KQK95356 (22 kDa) | 203 | 68.97% | 3.94% | 0.00% | 27.09% |
| 4 | KQL15787.1 (22.4 kDa) | 204 | 72.55% | 1.47% | 0.00% | 25.98% |
| 5 | KQK95355.1 (26 kDa) | 240 | 72.50% | 2.08% | 0.00% | 25.42% |

### 3.4 Multiple sequence alignment

Multiple sequence alignment was done to find the conserved regions among the prolamin proteins (Fig 1). The N-terminal signal peptide sequences of all the prolamin proteins showed highly conserved regions. These prolamin proteins were rich in certain amino acids like glutamine, proline, alanine and leucine. These foxtail millet prolamin proteins shows series of tandem repeats of Q residues.

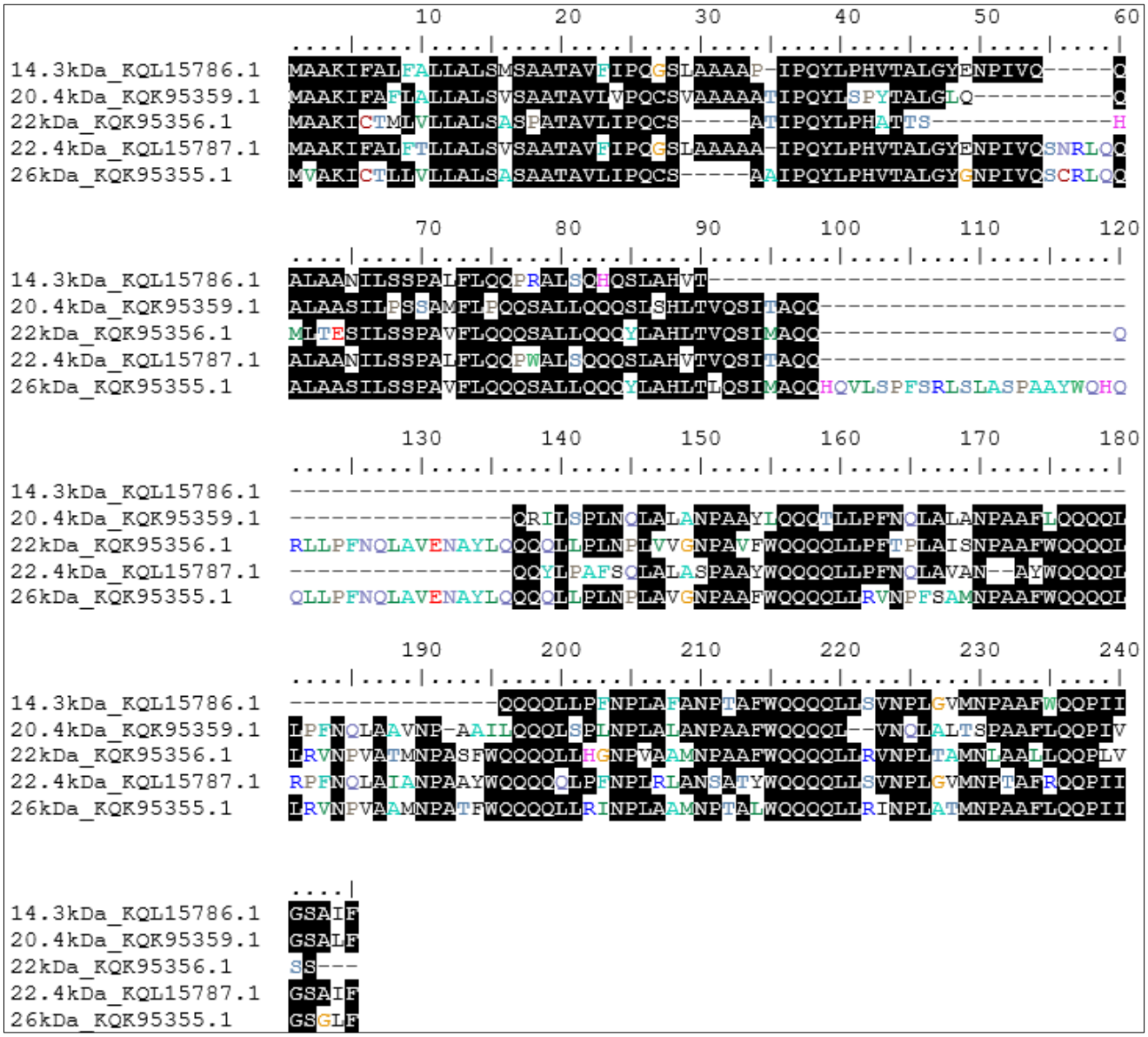


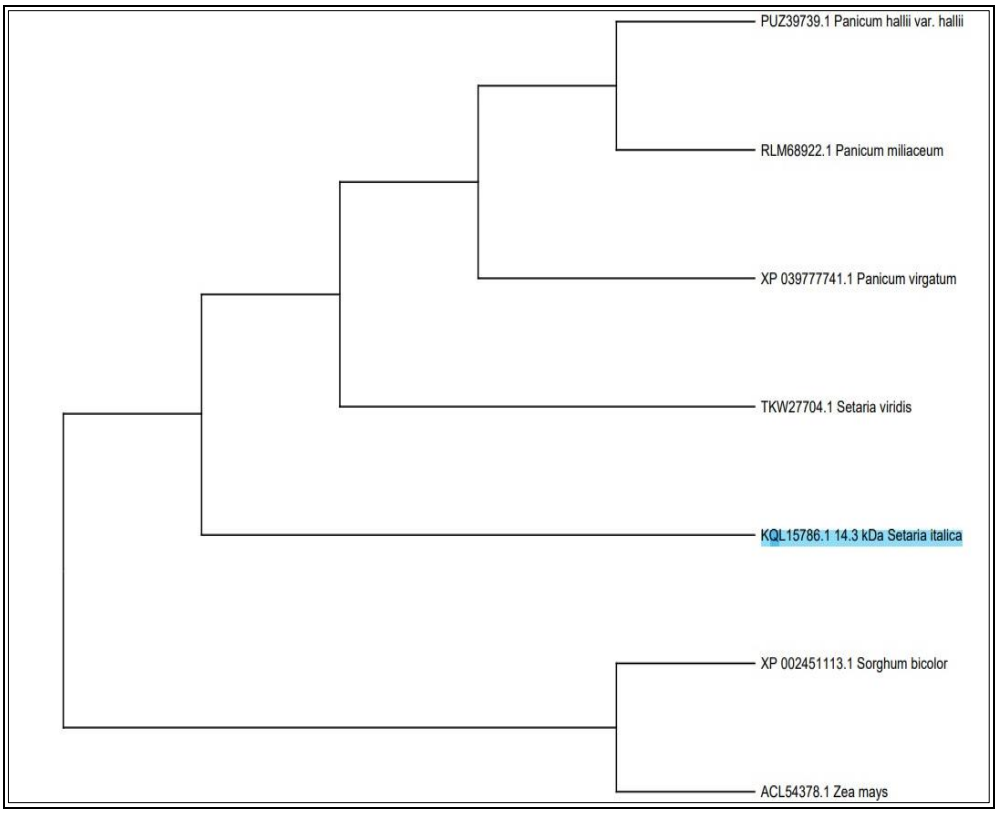

**Fig 1:** Multiple sequence alignment of prolamin proteins of foxtail millet

### 3.5 Phylogenetic analysis and Single peptide prediction

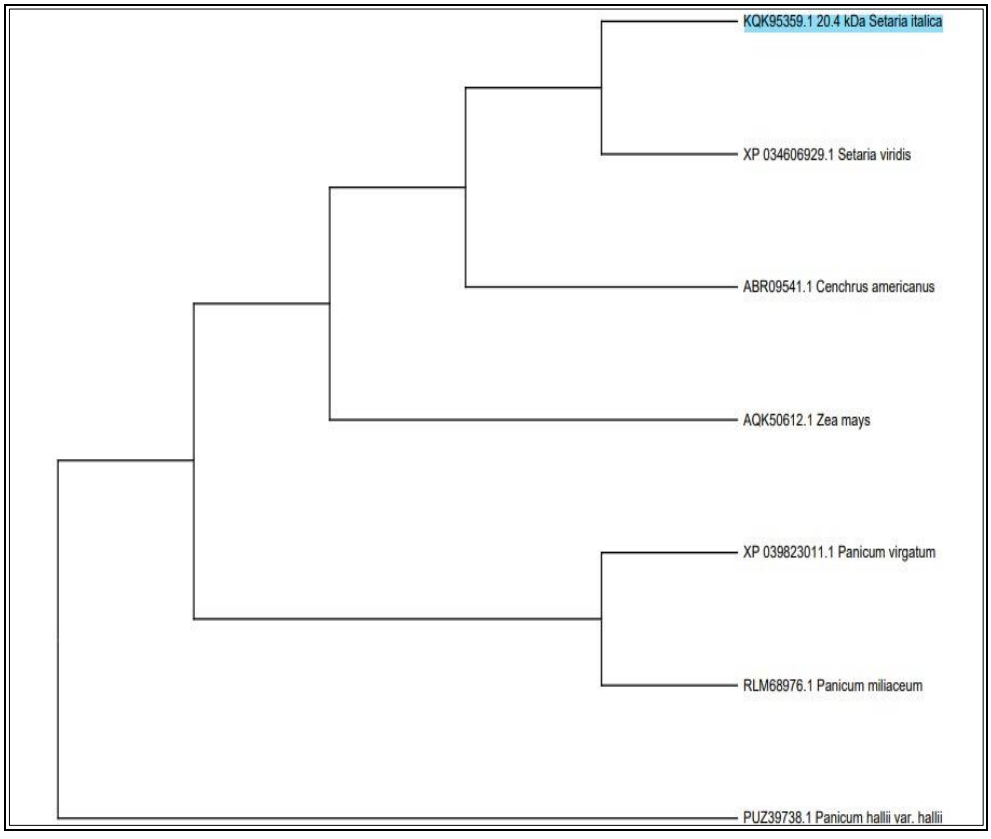
The phylogenetic analysis was done to the prolamin proteins of foxtail millet with other related species like *Panicum miliaceum*, *Panicum halliii var halliii*, *Panicum virgatum*, *Sorghum bicolor*, *Zea mays* and *Cenchrus americanus* (Fig 2A-2E). The 14.3 kDa and 22.4 kDa prolamin proteins were closely related to *Setaria viridis* and shows distant lineage with *Zea mays*. The 20.4 kDa and 22 kDa prolamin proteins were closely

related to *S.viridis* and shows distant lineage with *Panicum halliii*. The 26 kDa prolamin protein was closely related to *S. viridis* and shows distant lineage with *Panicum virgatum.*

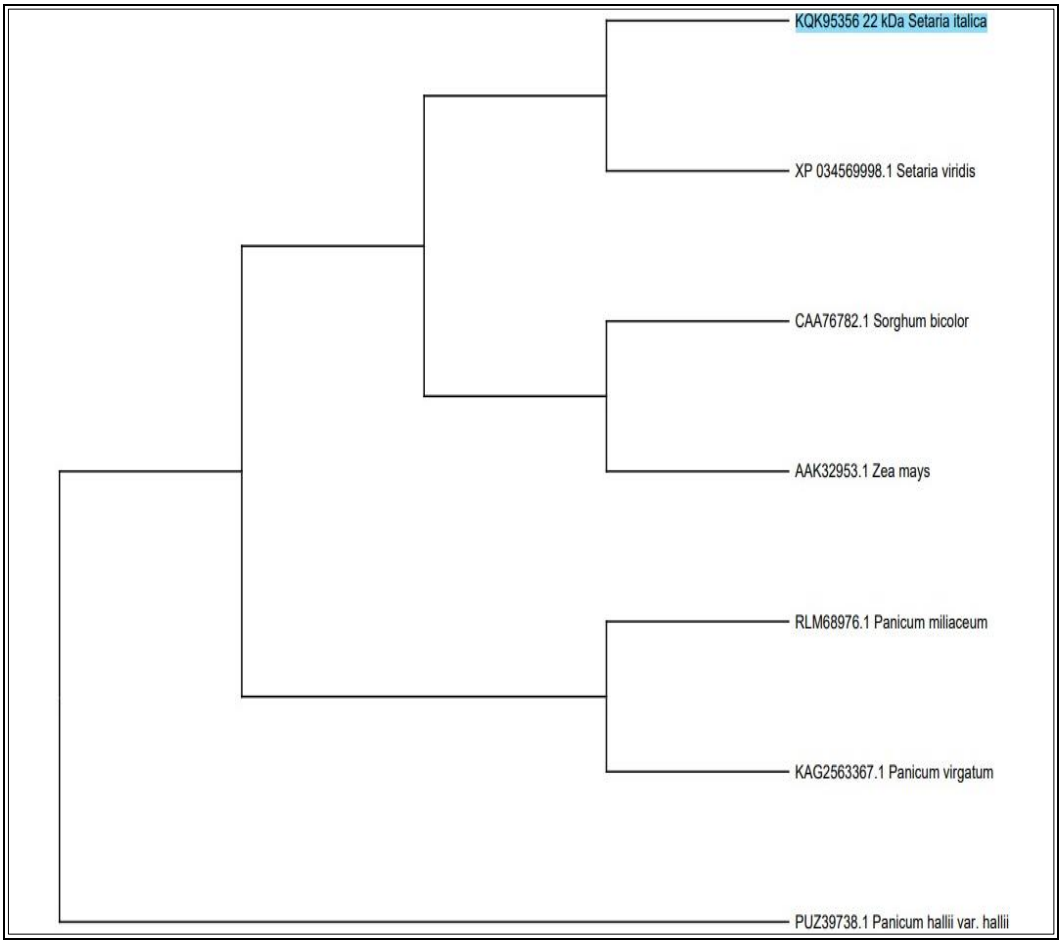
Signal peptides of the prolamin proteins were predicted and the cleavage site was found between the 20 to 21 amino acids (Table 5). The amino acid count and molecular weight of all the prolamin proteins were calculated after the removal of the signal peptide sequences that ranges from 12.2 to 24.3 kDa. Earlier study indicates the presence of 12–33 kDa range in the foxtail millet's SDS-PAGE pattern (Kumar and Parameswaran, 1998) [13].



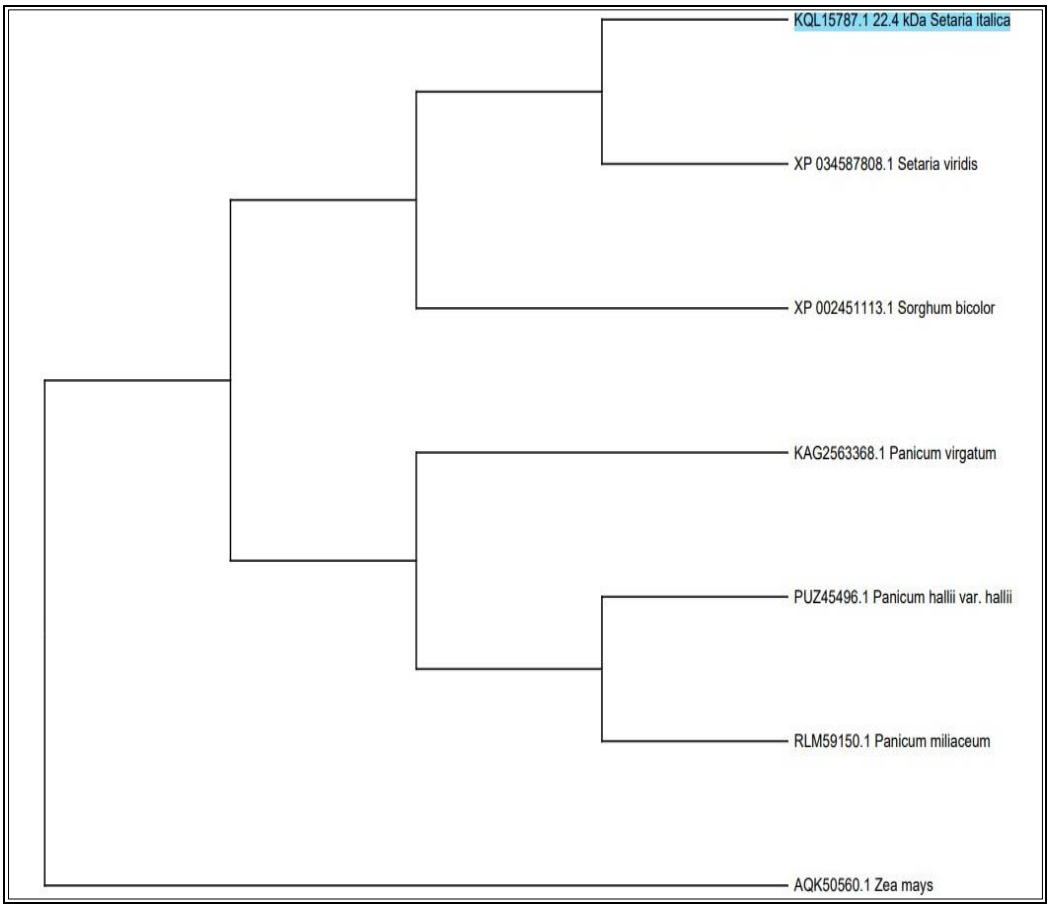

**A) KQL15786.1 (14.3 kDa)**



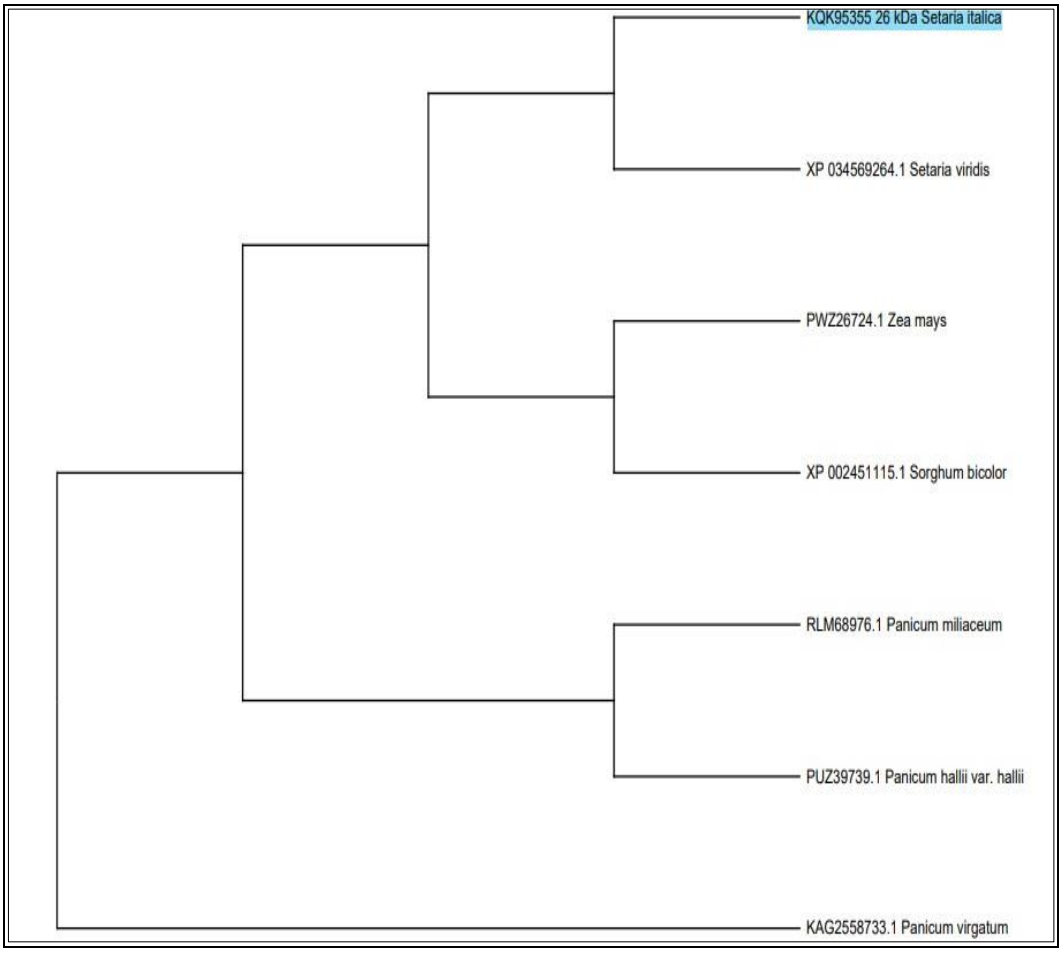

**B) KQK95359.1 (20.4 kDa)**




**C) KQK95356 (22kDa)**




**D) KQL15787.1 (22.4 kDa)**




**E) KQK95355.1 (26kDa)**

**Fig 2:** Phylogenetic analysis of prolamin proteins of foxtail millet

**Table 5:** Signal peptide prediction of prolamin proteins of foxtail millet

| S No | Protein Name | Full Length Protein | | Signal peptide score | Cleavage site | Protein without signal peptide | |
|---|---|---|---|---|---|---|---|
| | | AA | MW (kDa) | | | AA | MW (kDa) |
| 1 | KQL15786.1 (14.3 kDa) | 134 | 14.3 | 0.9713 | 21 to 22 | 113 | 12.2 |
| 2 | KQK95359.1 (20.4 kDa) | 194 | 20.4 | 0.9567 | 21 to 22 | 173 | 18.4 |
| 3 | KQK95356 (22 kDa) | 203 | 22.4 | 0.9895 | 21 to 22 | 182 | 20.3 |
| 4 | KQL15787.1 (22.4 kDa) | 204 | 22.4 | 0.9544 | 21 to 22 | 183 | 20.3 |
| 5 | KQK95355.1 (26 kDa) | 240 | 26.4 | 0.9828 | 21 to 22 | 219 | 24.3 |

## 4. Conclusion

Foxtail millet has been recognised as a wholesome nutritious cereal crop. The study revealed the presence of favourable essential amino acid composition in all the foxtail millet prolamins. The phylogenetic analysis of prolamin proteins of foxtail millet showed close homology with *Setaria viridis*. Secondary structural analysis demonstrates that alpha helix as the major secondary structure. High level of sequence homology was observed among the foxtail millet prolamins. The characterization of foxtail millet prolamin helps in determining the nutritional value of the prolamin and to predict the structure of prolamin proteins.

## 5. Acknowledgments

The authors express sincere thanks to the Department of Plant Biotechnology, Tamil Nadu Agricultural University, Coimbatore for providing facilities to conduct the research

## 6. References

1. Arvinder S, Yogesh K. Manipulating seed storage proteins for enhanced grain quality in cereals. African Journal of Food Science. 2009;3(13):439-446.
2. Brutnell TP, Wang L, Swartwood K, Goldschmidt A, Jackson D, Zhu XG, *et al*. *Setaria viridis*: a model for C4 photosynthesis. The Plant Cell. 2010;22(8):2537-2544.
3. Gaur VS, Sood S, Tiwari S, Kumar A. Genome-wide identification and characterization of seed storage proteins (SSPs) of foxtail millet (*Setaria italica* (L.) *P. Beauv*.). 3 Biotech. 2018;8(10):1-16.
4. Holding DR. Recent advances in the study of prolamin storage protein organization and function. Frontiers in plant science. 2014;5:276.
5. http://www.ncbi.nlm.nih.gov/
6. https://web.expasy.org/protparam
7. https://PRPSITE.expasy.org
8. http://pfam.xfam.org/search/sequence
9. https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa%20_sopma.html
10. https://www.megasoftware.net/dload_mac_beta
11. https://services.healthtech.dtu.dk/service.php?SignalP-5.0

12. Kohama, Keiko, Takashi Nagasawa, Naoyuki Nishizawa. Polypeptide compositions and $NH_2$-terminal amino acid sequences of proteins in foxtail and Proso millets. Bioscience, biotechnology, and biochemistry. 1999;63(11):1921-1926.
13. Kumar KK, Parameswaran KP. Characterisation of storage protein from selected varieties of foxtail millet (*Setaria italica* (L) Beauv). Journal of the Science of Food and Agriculture. 1998;77(4):535-542.
14. Magdeldin S, Yoshida Y, Li H, Maeda Y, Yokoyama M, Enany S, *et al*. Murine colon proteome and characterization of the protein pathways. Bio Data mining. 2012;5(1):1-14.
15. Mandal S, Mandal RK. Seed storage proteins and approaches for improvement of their nutritional quality by genetic engineering. Current Science; c2000. p. 576-589.
16. Nakayama H, Komatsu S, Namai H, Okuno K. N-terminal amino acid sequences of prolamins encoded by the alleles at the Pro1 and Pro2 loci in foxtail millet, *Setaria italica* (L.) P. Beauv. Genes & genetic systems. 1999;74(6):309-314.
17. Osborne TB. The vegetable proteins. Longmans, Green and Company; 1924.
18. Shewry, Peter R, Rod Casey. Seed proteins. In Seed proteins. Springer, Dordrecht; c1999. pp. 1-10.
19. Shewry PR, Miflin BJ, Forde BG, Bright SWJ. Conventional and novel approaches to the improvement of the nutritional quality of cereal and legume seeds. Science Progress; c1933-1981. p. 575-600.
20. Sigrist CJ, Cerutti L, Hulo N, Gattiker A, Falquet L, Pagni M, *et al*. PRPSITE: A documented database using patterns and profiles as motif descriptors. Briefings in bioinformatics. 2002;3(3):265-274.