Life Science Informatics Publications

Research Journal of Life Sciences, Bioinformatics, Pharmaceutical and Chemical Sciences

Journal Home page http://www.rjlbpcs.com/

RJLBPCS

ISSN 2454-6348

**Original Research Article** **DOI: 10.26479/2021.0704.05**

# PHYLOGENETIC, SEQUENCE ANALYSIS AND STRUCTURAL STUDIES OF MATURASE K PROTEINS FROM FERN SPECIES

**S. B. Thakar[1,4*], M. J. Dhanavade[2], K. D. Sonawane[2,3]**

1. Department of Biotechnology, Shivaji University, Kolhapur- 416 004, (M.S.), India.
2. Structural Bioinformatics Unit, Department of Biochemistry, Shivaji University, Kolhapur, India.
3. Department of Microbiology, Shivaji University, Kolhapur – 416 004 (M.S.), India.
4. School of Integrative Engineering, Chung-Ang University, 84 Heukseok-ro, Dongjak-gu, Seoul, Republic of Korea.

**ABSTRACT:** Ethnobotanical and remedial values of fern have been widespread by way of different researchers from time to time through a number of methods but adequate responsiveness has not been given in the direction of their medicinally valued applications, so it is essential to study isolated proteins from diverse ferns in feature at molecular level as well as end to end with their taxonomical connection. The objective of the current study is to applied bioinformatics systems to realize evolutionary importance as a result of building phylogenetic tree in direction to address association among ferns Maturase K (matK) protein sequences of a number of ferns were retrieved from NCBI and utilized for multiple sequence alignment and phylogenetic investigation. Bioinformatics techniques were performed to recognize evolutionary importance through building phylogenetic tree.3- dimensional structures matK proteins from ferns species were built by homology modeling. The current study specifies matK protein to be a suitable marker to identify fern species. This investigation also suggests that matK protein is a respectable candidate for plant systematics as well as DNA barcoding studies. The 3-D model of Maturase K protein also shows good quality and might be beneficial to realize the structural and functional connection among the different fern species to infer evolutionary significance. So, the predicted model of matK from fern may perhaps be helpful in the exploration of plant systematics as well as the structural analysis of Maturase K proteins of diverse fern species, which further can be used to detect structural alterations among Maturase K proteins during evolution.

**Keywords:** Ferns, Maturase K, Phylogenetic, Sequence analysis, Homology modeling

**Article History: Received: July 30, 2021; Revised: August 12, 2021; Accepted: August 27, 2021.**

**Corresponding Author: Dr. S. B. Thakar*** Ph.D.
School of Integrative Engineering, Chung-Ang University, 84 Heukseok-ro, Dongjak-gu, Seoul, Republic of Korea. Email Address: sbthakar@gmail.com

## 1.INTRODUCTION

The thousands of year's plants are being used in traditional remedies [1]. Currently, as per the World Health Organization (WHO) information, almost 80% of the world's people depend on traditional medications for their prime healthcare necessities [2]. Ferns are the largest second integral part of world flora as well as the oldest groups of primitive vascular plants on the earth [3]. Approximately 12,000 species of fern arise in the world flora out of which more than 1200 species are stated in India [4, 5]. Ethnobotany as well as medicinal values of fern has been defined by a number of research workers from time to time by different ways but adequate attention has not been professional towards their medicinally useful aspects [1]. So range of proteins/enzymes has been mined from fern plants. The matK gene of chloroplast codes for maturase similar to protein, which is contribution in Group II intron splicing in the plant systematics, MatK has newly occurred as a helpful gene because of its high phylogenetic signal associated with other genes [6]. MatK has an infrequent phylogenetic mode as well as relatively high substitution rates in the amino acids and nucleotide levels [7]. These phylogenetic applications have raised questions regarding functions of Maturase K. Strong evolutionary signal from matK as well as it is extremely advantageous gene in evolutionary studies and plant systematic [7]. In Plant systematics, Mat K utilized as molecular marker as identify the diversity that arises between the plants as well as animal species. The internal transcribed spacer (ITS) region of nuclear ribosomal cistron is usually exploited sequence locus for molecular systematic in plant studies [8]. Abundant nuclear genes, chloroplast and mitochondrial have been employed for studying sequence difference at genus level. [9] Maturase K is the 1500 bp long and inside located intron of the trnK as well codes for maturase similar protein, which is role in Group II intron splicing. The Maturase K surrounds high substitution rates amongst the species and is growing as prospective candidate to evaluation plant systematics as well as evolution [10]. A homology assessment for this gene confirmations shows that the 102 amino acids at the carboxyl terminus are basically connected to specific regions of maturase-like polypeptide and may be contributed in splicing of group II introns. It is one more emerging gene with potential role to plant evolution as well as molecular systematics [11]. Usually the matK-trnK gene complex is used for plant studies for evolution and the solution for several taxonomic levels [12, 13]. MatK genes of features are exploited to purpose family's relations level of among the fern species. The site of matK

in the trnK gene was resolute by concerning with a matK sequence of Trillium [14]. This data was operated to detection of molecular markers, which was utilized for detecting species of these taxa as well as delivers the valued information for conservative as well as studies of molecular plant breeding [15]. In the present study, we observed matK protein from fern species to run insight into functional characteristics of this commonly occurring gene. We have selected Mat K protein from 10 fern species as shown Table No.1 List of Maturase K fern species. Molecular phylogenetic studies are employed to order of evolutionary actions and signify them in evolutionary trees that graphically represent connections and homologous between fern species. Maturase K proteins point out that they have common ancestors in fern species and these proteins had been successful for resolve of phylogenetic relationships between taxa of fern plants.

## 2. MATERIALS AND METHODS

### 2.1. Software and Hardware

Homology modeling was completed with SWISS-MODEL to find out structural matches among maturase K proteins [17]. The expected models were assessed by online servers such as PROCHECK programs [16] for the superiority assessment of Ramachandran plot, PROSA [16]. The homology modeling study was performed on HP workstation in addition Rack/Blade server accessible as an in house facility. The studies of molecular structures as well as interactive visualization were carried out by chimera [24].

### 2.2. Sequence, phylogenetic analysis and homology modeling

The whole maturase K (matK) amino acid sequences from 10 fern species were mined from NCBI protein sequence database (Accession No.NP_848039, AAK69120, CAP04517, CAP04522, BAM65725, AAX12244, ABF51613, ACZ63164, BAM65728, BAM65735) (http://www.ncbi.nlm.nih.gov) [16]. Multiple sequence alignment study was completed with CLUSTAL 2.1 EBI Tool (www.ebi.ac.uk/Tools/msa/) [17]. Phylogenetic analysis was performed with Molecular Evolutionary Genetics Analysis (MEGA version 6.0) program [18]. The data sets for ten proteins contained within dissimilar range of taxa from fern species. The BLAST platform was making use of to search proper template obtainable in the PDB. The multiple sequence alignment of the template sequence (4B2Q) [25] as well as the maturase K sequence was through by the EMBOSS platform [23].

### 2.3. Maturase k model refinement and validation

The models developed with SWISS-MODEL [21] were authenticated through checkup of the Phi/ Psi Ramachandran plot [19] completed from PROCHECK analysis [23]. The model generated by SWISS-MODEL [14] was lastly in use for further studies on the basis of geometry, 3D alignment with the template also the results of PROCHECK and PROSA analyses[20].The PROSA program [20] has been performed to check the energy measures of predictable models in evaluation with known X-ray and NMR structures. The PROSA II energy plot was calculated to check the dealings

energies of whole residues of the matk expected model. The Ramachandran plot was achieved by using PROCHECK analysis [23]. The expected homology model was then matched with a template structure (PDBID: 4B2Q) [25] using chimera [24].

## 3. RESULTS AND DISCUSSION

### 3.1 Phylogenetic analysis of MatK

Invention in molecular biology and protein sequencing methods has supported to distinguish the proteome of several organisms rapidly. Also, a minor effort has been made in the current study to explore sequence as well as structural connections between fern species Phylogenetic investigates of the protein sequences of ten fern species are given that valued evidence regarding their protein makeup, taxonomy, plant systematics, DNA barcoding as well as common antecedent. The phylogenetic analysis of 10 ferns species (Table 1) confirmed their evolutionary connection and homology. From these 10 ferns species we have utilized maturase K protein sequences for the sequence exploration study. The outcomes point out that these fern species have common ancestors. exploration study. The outcomes point out that these fern species have common ancestors.

**Table 1 List of Maturase K Fern Species**

| SR .No | Fern Species | Family | Accession Number | Sequence |
|---|---|---|---|---|
| 1 | *Adiantum capillus-veneris* | Pteridaceae | NP_848039 | Maturase K |
| 2 | *Dioon spinulosum* | Zamiaceae | AAK69120 | Maturase K |
| 3 | *Equisetum arvense* | Equisetaceae | CAP04517 | Maturase K |
| 4 | *Equisetum ramosissimum* | Equisetaceae | CAP04522 | Maturase K |
| 5 | *Helminthostachys zeylanica* | Ophioglossaceae | BAM65725 | Maturase K |
| 6 | *Lycopodiella cernua* | Lycopodiaceae | AAX12244 | Maturase K |
| 7 | *Huperzia selago* | Huperziaceae | ABF51613 | Maturase K |
| 8 | *Lygodium microphyllum* | Lygodiaceae | ACZ63164 | Maturase K |
| 9 | *Ophioglossum pendulum* | Ophioglossaceae | BAM65728 | Maturase K |
| 10 | *Ophioglossum reticulatum* | Ophioglossaceae | BAM65735 | Maturase K |

In this analysis we have done multiple sequence alignment and phylogenetic exploration of Maturase K protein sequences of fern species. The multiple sequence alignment displays identical conserved domain as depicted in Fig.3.1.


**Fig.1: Multiple Sequence Alignment of maturase K protein sequences of fern**

Used for Mat K proteins, the trees were achieved using the bootstrap methods. We show the NJ consensus trees in Fig. 2. We carry out combined analysis of ten matK fern protein sequences using MEGA 6 software such as inferred inherited sequences of taxa with Maximum Likelihood method.[14]

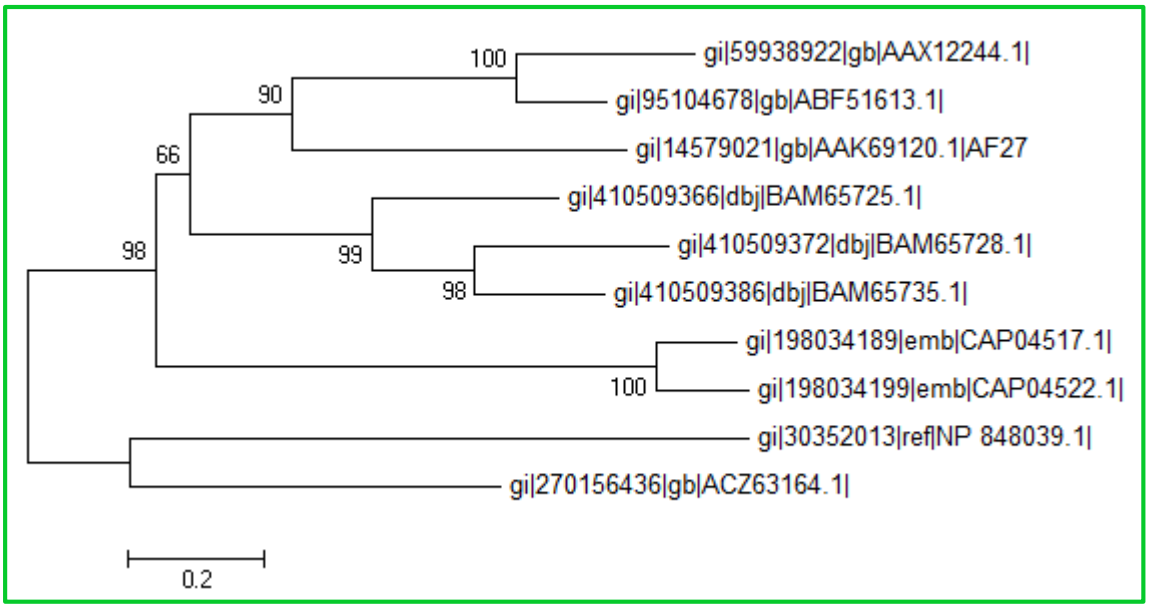


**Fig. 2: Molecular Phylogenetic analysis by Maximum Likelihood method**

The evolutionary history was inferred using the Neighbor-Joining method [22]. The optimal tree with the sum of branch length = 4.01089754 is shown. The confidence probability (multiplied by 100) that the interior branch length is greater than 0, as estimated using the bootstrap test (500 replicates is shown next to the branches [28, 29]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method [30] and are in the units of the number of amino acid substitutions per site. The analysis involved 10 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 219 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 [26] Sequence analysis of MatK. The data group used for ten proteins comprised from ferns for which the sequences are accessible in NCBI protein database to discover the interspecies difference. The Mat K protein sequences of different 10 fern species as displayed in table 1 as well as belonging to 5 families were mined from NCBI protein sequence database. [12]. Multiple sequence alignment was done by using CLUSTAL 2.1 EBI Tools (www.ebi.ac.uk/Tools/msa/) [13]. Structural prediction was completed with Swiss Model Program [17]

**4.4.3.2 Structural analysis of Predicted models**

We have implemented multiple sequence alignment as well as studies of homology modeling in Maturase K protein sequences of ferns species. The multiple sequence alignment displays conserve domain as represented in Fig 1. Exploration of amino acid sequences for matK proteins specify that fern species were comparatively distantly related between them 10 ferns were related. MatK

protein of Homology model predominately encompasses Helix, Sheet as well as loops (Fig. 3). Then more the expected model of MatK constructed by SWISS-MODEL (Figure 3) [17] was used to check the model quality.

The expected model quality was valued using programs like PROSA, RamPage and the RamPage analysis model of MatK constructed using SWISS-MODEL displays that Swiss Model 1 is having a respectable quality in which 89.1% residues are found in best number one regions, 6.2% residues in allowed regions and 4.7% residues in outlier regions (Fig. 4, Table No 2).

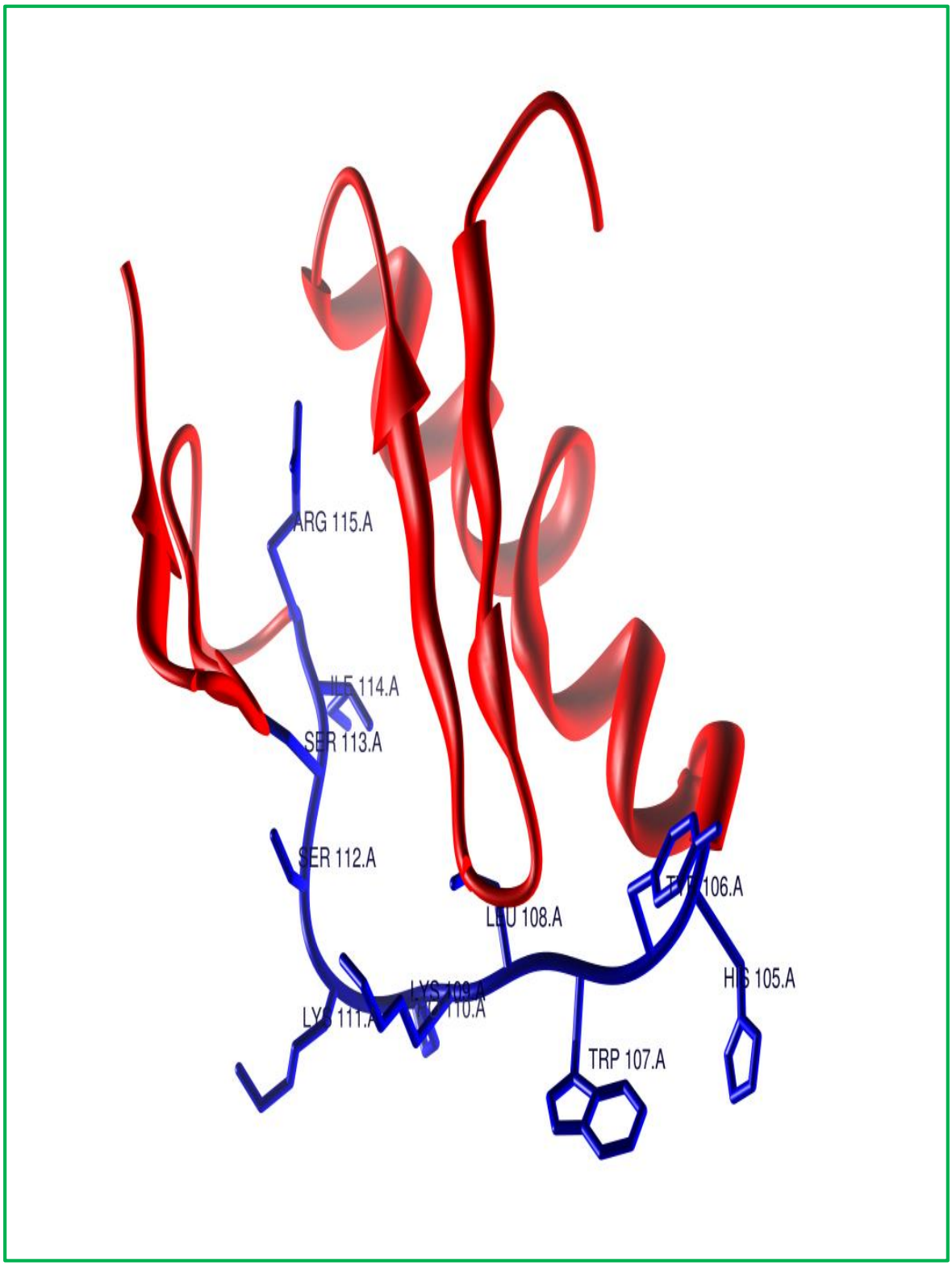


**Fig. 3: Predicted model of matK protein from *Adiantum capillus-veneris***

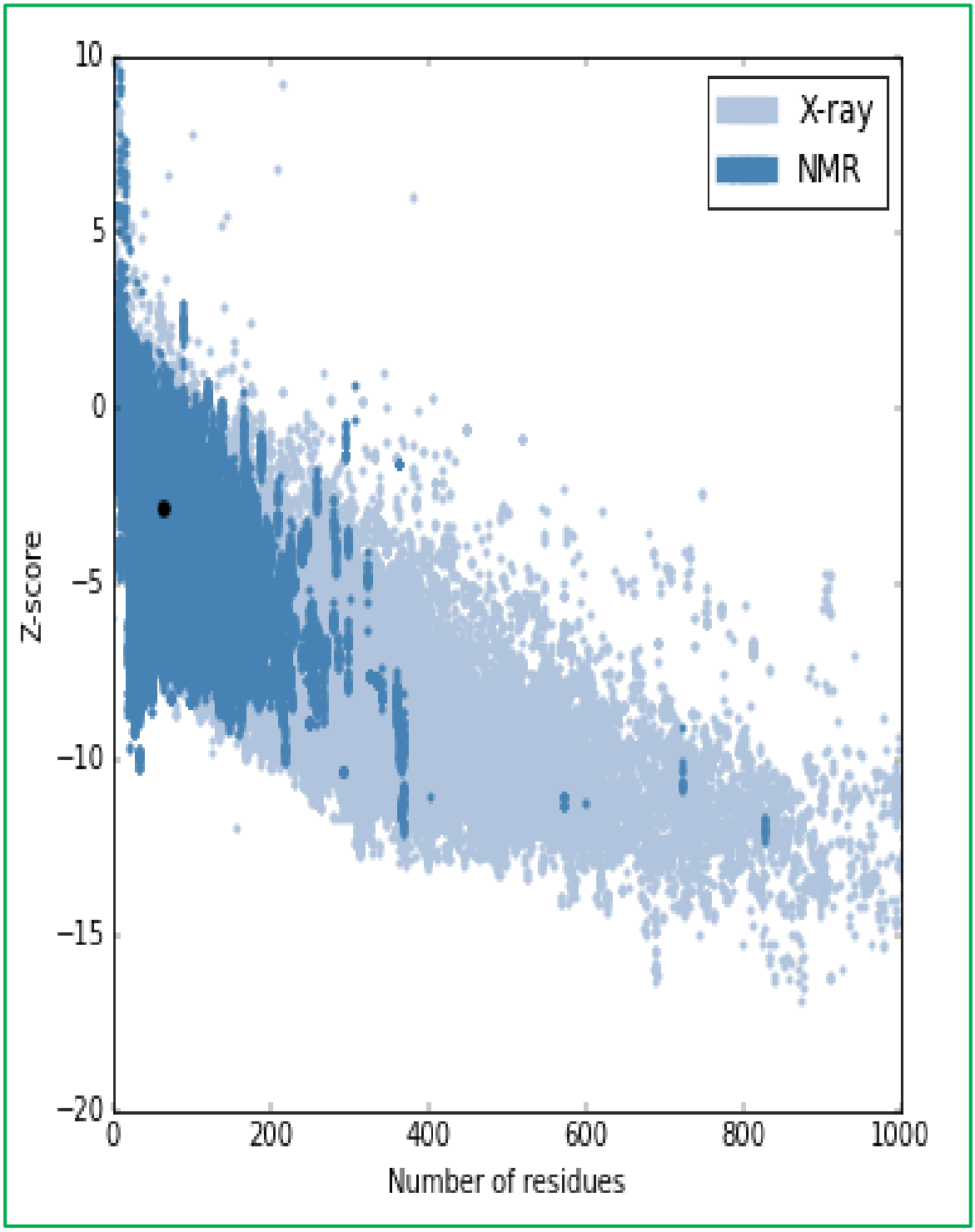


**Fig. 3: Validation of MatK model using ProSA The black dot shows the Similarity of model with X-ray and NMR structures**

**Table 2: Stereo-chemical quality by PROCHECK and model evaluation through PROSA for Maturase K 3D Structures**

| Sr.No | Generated Model | Rampage analysis showing residues at various regions | | | PROSA Z-score |
|---|---|---|---|---|---|
| | | Favoured region (%) | Allowed region (%) | Outlier region (%) | |
| 1 | Swiss Model 1 | 89.1% | 6.2% | 4.7% | -2.87 |
| 2 | Swiss Model 2 | 82.8% | 9.5% | 7.6% | -2.17 |
| 3 | Swiss Model 3 | 76.1% | 16.4% | 7.5% | 0.73 |
| 4 | Swiss Model 4 | 93.2% | 4.5% | 2.3% | -2.32 |
| 5 | Swiss Model 5 | 87.5% | 7.8% | 4.7% | -2.7 |
| 6 | Swiss Model 6 | 84.0% | 10.4% | 5.7% | -2.26 |
| 7 | Swiss Model 7 | 82.1% | 7.1% | 10.7% | -2.27 |
| 8 | Swiss Model 8 | 81.1% | 10.8% | 8.1% | -0.15 |
| 9 | Swiss Model 9 | 85.9% | 7.8% | 6.2% | -2.95 |
| 10 | Swiss Model 10 | 87.3% | 9.5% | 3.2% | -1.74 |

Hence, the total 89.1% of residues have been discover in the preferred regions and only 4.7% of residues are detected in the outer region proposing the respectable model quality (Table 2). In addition the PROSA program [16] was perform to check the model superiority using the Z score of the structure. The Z score is revealing of complete model superiority and is utilized to check whether the target structure is inside the range of scores usually discover in innate proteins of similarsize. The Z score for maturase k model was -2.87.The outcomes obtained using PROSA program [16] revealed that the maturase k structure was inside the satisfactory range of X-ray as well as NMR studies (Fig. 4). Generally the RamPage program [15] and PROSA program [16] studies revealed respectable result of the model constructed using SWISS-MODEL [17] the expected model quality was evaluated using platforms like RamPage and PROSA program. The RamPage investigation of a MatK model constructed through SWISS-MODEL displays that Swiss Model 1 is having a respectable quality. In addition PROSA program [16] was utilized to confirm the respectable quality of model. The Z score for maturase k model was -2.87.The outcomes obtained by PROSA programs recommend that maturase k structure was within the suitable range of X-ray as well as NMR studies.

Therefore, finally the structural analysis of Mat K protein of different fern species might be used to recognize the structural variation among Mat K proteins throughout evolution as well as interspecies correlation among the fern species.

## 4. CONCLUSION

The current study specifies *mat*K protein to be a suitable marker in discovering fern species. A portable software Molecular Evolutionary Genetics Analysis (MEGA) outline for competent discovering of amino acid sequences of fern species which is delivered with inter as well as intra species correlation with common ancestor and pattern between ancestries are identical (Homologous) as well as rates between places are identical inside the *matk* amino acid sequences. The combined tree analysis displays that the group has higher boot strap values building the evolutionary intelligence among the 10 species of ferns. As a result, from this investigation it might be recommended that *matK* amino acid is a respectable candidate for plant systematics as well as DNA barcoding infers species as well as the current study revealed that expected Maturase K model from ferns through Swiss model has a respectable quality. This expected model might be beneficial to realize the structural and functional connection among the different fern species and Mat K from fern might be also significant as it has strong evolutionary indicator. Consequently, the predicted model of MatK from fern plant source may perhaps be helpful in the exploration of Plant systematics as well as the structural analysis of Mat K protein of diverse fern species can be used to detect structural alteration among Mat K proteins during evolution.

## ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

## HUMAN AND ANIMAL RIGHTS

No Animals/Humans were used for studies that are base of this research.

## CONSENT FOR PUBLICATION

Not applicable.

## AVAILABILITY OF DATA AND MATERIALS

The author confirms that the data supporting the findings of this research are available within the article.

## FUNDING

None

## CONFLICT OF INTEREST

The authors declare that there is no conflict of interests.

**REFERENCES**


1. Thakar, S.B, Ghorpade, P.N, Kale M.V, Sonawane, K.D, FERN Ethnomedicinal Plant Database: Exploring fern ethnomedicinal plants knowledge for computational drug discovery Current computer-aided drug design, 2015; 11 (3), 266-271
2. May, L.W.The economic uses and associated folklore of ferns and fern allies, Bot. Rev, 1978; 44: 491-528
3. Singh sweta, Rita sing Utilization of ferns of Achankmar Amarkanatak Biosphere Reserve, Central India In women's health and Beauty care Practices. IRJP 2013; 4:1
4. Abu rabia et al.Urinary diseases and ethnobotany among pastoral nomads in the Middle East Journal of Ethnobiology and Ethnomedicine 2005;1:4.
5. Chellaiah Muthu, et al.Medicinal plants used by traditional healers Kancheepuram District of Tamil Nadu, India. Journal of Ethnobiology and Ethnomedicine, 2006; 2:43
6. Muller, K. F., Borsch, T and khidir H. Phylogenetic utility of rapidly evolving DNA at high taxonomical levels: contrasting matK, trnT-F and rbcL in basal angiosperms. Molecular Phylogenetic and Evolution 2006; 41: 99–117.
7. Michelle, B. and khidir, H. Expression of matk: functional and Evolutionary implications. American Journal of Botany, 2007; 94(8): 1402–1412
8. Kress ,W.J, Wurdack, K.J, Zimmer, E.A, Weigt, L.A, Janzen, D.H (2005) Use of DNA barcodes to identify flowering plants. Proc Natl Acad Sci USA 102(23): 8369–8374. PMID: 15928076
9. M. W. Chase et al., Phylogenetics of seed plants: An analysis of nucleotides sequences from the plastid gene rbcL. Annals of the Missouri Botanic Garden, 1993; 80: 528.
10. Notredame C, Higgins, D.G and Heringa J. Coffee: A novel method for fast and accurate multiple sequence alignment. Journal of Molecular Biology,; 2000 205: 217
11. Khidir, W.H and Hongping, L. The matK gene: sequence variation and application in plant systematics. American Journal of Botany 1997; 19: 830–839.
12. Ito, M and Kawamoto, A, Phylogenetic Relationships of Amaryllidaceae Based on matK Sequence Data, Journal of Plant Research, 1999; 207-216.
13. Wolfe, K. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proceedings of the National Academy of Science, 1987; 9054: 9058
14. Kazempour Osaloo S, Utech FH, Ohara M and Kawano S. Molecular systematics of the Trilliaceae. I. Phylogenetic analysis of Trillium using matK gene sequences. Journal of Plant Research 1999; 112: 35–49.
15. Pedersen, L.B. Phylogenetic analysis of the subfamily Alpinioideae (Zingiberaceae), particularly Etlingera Giseke, based on nuclear and plastid DNA. Plant Systematics and Evolution 2004; 245:239–258.
16. http://www.ncbi.nlm.nih.gov/

17. http://www.ebi.ac.uk/Tools/msa/
18. Kumar S., Tamura K, Nei M. MEGA: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment //Brief. in Bioinform. 2004; V. 5. P. 150–163.
19. S.C. Lovell, I.W. Davis, W.B. Arendall III, P.I.W. de Bakker, J.M. Word, M.G. Prisant, J.S. Richardson and D.C. Richardson. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins: Structure, Function & Genetics. 2002; 50: 437-450.
20. M. Wiederstein, M.J. Sippl. ProSA web: interactive web service for the recognition of errors in three dimensional structures of proteins, Nucleic Acids Research 2007; 35.W407 W410.
21. T. Schwede, J. Kopp, N. Guex, M. C. Peitsch, SWISS MODEL: anautomated proteinhomolog y-modeling server, Nucleic Acids Research 31(2003) 3381–3385.
22. Saitou N. and Nei M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 1987; 4:406-425.
23. Dopazo J. Estimating errors and confidence intervals for branch lengths in phylogenetic trees by a bootstrap approach. Journal of Molecular Evolution 1994; 38:300-304.
24. Rzhetsky A. and Nei M. A simple method for estimating and testing minimum evolution trees. Molecular Biology and Evolution 1992; 9:945-967.
25. Zuckerkandl E. and Pauling L. (1965). Evolutionary divergence and convergence in proteins. Edited in Evolving Genes and Proteins by V. Bryson and H.J. Vogel, pp. 97-166. Academic Press, New York.
26. Tamura K., Stecher G., Peterson D., Filipski A., and Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Molecular Biology and Evolution 2013; 30: 2725-2729.
27. P.Rice,I.Longden,A.Bleasby. EMBOSS: the European molecular biology open Software suite, Trends in Genetics 16 (6) (2000) 276–277
28. R.A. Laskowaski, M.W. McArther, D.S.Moss, J.M.Thornton, PROCHECK a program to check sterio-chemical quality of a protein structures, Journal of Applied Crystallography, 1993; 26, 283–291.
29. E. F. Pettersen, T.D. Goddard, C. C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E.Fer rin.UCSF Chimera a visualization system for exploratory research and analysis, Journal of Computational Chemistry, 2004; 25, 1605–1612.
30. K.M., Davies, C., Anselmi, I., Wittig, J.D., Faraldo-Gomez, W. Kuhlbrandt (2012) Structure of the Yeast F1Fo-ATP Synthase Dimer and its Role in Shaping the Mitochondrial Cristae. Proc.Natl.Acad.Sci.USA