



OPEN Phylogenetic analysis of selected species of Asteraceae on the basis of *RPS 11* Gene

Syeda Anber Zahra¹, Javed Iqbal²✉, Banzeer Ahsan Abbasi³, Sobia Kanwal⁴, Mona S. Alwahibi⁵, Mohamed S. Elshikh⁵, Muhammad Rizwan⁶✉, Rashid Iqbal^{7,8}✉ & Tariq Mahmood⁹✉

The Asteraceae family is a prominent group of flowering plants found across the globe, with the exception of Antarctica. The Asteraceae family is a largest flowering family pivotal group in plant evolution and diversification. Despite its importance, the genetic diversity within this family remains understudied. We focused on the *rps-11* gene, a chloroplast marker previously utilized in phylogenetic studies, to investigate its potential in resolving Asteraceae relationships. The focus was on examining genetic diversity within sixteen specifically chosen species from the Asteraceae family. This assessment was based on an analysis of a chloroplast gene responsible for encoding the ribosomal protein of the smaller subunit 11 (*rps 11*). Nearly 417 bp of *rps 11* gene was amplified, sequenced, computationally translated into amino acid sequence and the data was used for phylogenetic analysis as well as for *rps 11* protein structure predictions. Based on nucleotide and amino acid sequences phylograms were drawn with the help of Molecular Evolutionary Genetic Analysis (MEGA 6), which exhibited clear genetic relationship among species under investigation. The observed genetic distance was 0.02 for Maximum likelihood tree based on nucleotide sequences whereas it was 0.05 for phylogram based on amino acid sequences. These values revealed that amino acid-based tree has demonstrated greater diversity among selected species in comparison to nucleotides-based tree. On the basis of pair wise distance calculations, genetic divergence values were found within the range of 0.015–0.309. Moreover, 3D protein modeling for *rps 11* protein of sixteen selected species was also carried out by iterative threading assembly refinement (I-Tasser) software. The models exhibiting the highest C-score were picked with satisfactory plot statistics (> 90%) and structurally validated by PROCHECK. Furthermore, Ramachandran plots displayed that the *rps 11* protein structures of *Tagetes minuta*, *Xanthium strumarium*, *Lactuca sativa* and *Chrysanthemum indicum* have best feature models with > 90% of residues in the allowed region and ≤ 2% in the disallowed region. The research is not enough to stand alone to validate the viability of the *rps11* gene as a prospective contender for phylogenetic analysis. In future we will focus on the maximum genetic diversity theory for phylogenetic analysis of this family.

Keywords Asteraceae, Genetic variability, *rps 11*, Phylogram, 3D protein modeling, Molecular evolutionary genetic analysis

The family Asteraceae is recognized as the sunflower, aster or daisy family. It is the largest flowering plant family and about 10% of Angiosperm plants belong to this family^{1,2}. According to the Royal Botanical Gardens of Kew, the family has around 1600 genera and 23,000 species³. Asteraceae is found worldwide except for Antarctica and

¹School of Food and Pharmacy, Shanghai Zhongqiao Vocational and Technical University, Shanghai 200000, China.

²Department of Botany, Bacha Khan University, Charsadda 24420, Khyber Pakhtunkhwa, Pakistan. ³Department of Botany, Rawalpindi Women University, 6th Road, Satellite Town, Rawalpindi 46300, Pakistan. ⁴Department of Biology and Environmental Sciences, Allama Iqbal Open University, Islamabad 45320, Pakistan. ⁵Department of Botany and Microbiology, College of Science, King Saud University, Riyadh 11451, Saudi Arabia. ⁶Institute of Crop Science and Resource Conservation (INRES), University of Bonn, 53115 Bonn, Germany. ⁷Department of Agronomy, Faculty of Agriculture and Environment, The Islamia University of Bahawalpur, Bahawalpur 63100, Pakistan. ⁸Department of Life Sciences, Western Caspian University, Baku, Azerbaijan. ⁹Department of Plant Sciences, Faculty of Biological Sciences, Quaid-i-Azam University, Islamabad 45320, Pakistan. ✉email: javed89qau@gmail.com; m.rizwan@uni-bonn.de; rashid.iqbal@iub.edu.pk; tmahmood@qau.edu.pk

has notable diversity, particularly in the subtropical and tropical regions of Southern Africa, the Mediterranean region, the Andes, North America, Eastern Brazil, Southwestern China and Central Asia^{4,5}. Most distinguishing feature of this family is the arrangement of florets on a receptacle in centripetally emerging heads enclosed by bracts and anthers bonded in a sphere. Plants belonging to this family are mostly herbaceous². Members of family *Asteraceae* are mostly perennial or biennial, annual herbs and rarely trees, sometimes bushes, with or without milky latex, sometimes with spines (Flora of Pakistan). Some of the genera of *Asteraceae* family *Aster*, *Helianthus*, *Chrysanthemum* and *Tagetes* etc. are ornamental plants and most of them have therapeutic values. The commercial importance of these plants can be assessed by the uses such as vegetable and food, as hair tonic etc.^{6,7}

Chloroplast DNA (cpDNA) is the most preferred DNA molecule for phylogenetic analysis. cpDNA offer an exceptional opportunity for studying phylogeny at intra and interspecies level due to its extraordinary genetic potential and conserve nature as compared to nuclear DNA^{8–11}. Several cpDNA genes (*matK*, *rbcl*, *atpB*, *trnL*, *trnL-F*, *rps11*, *rps14*, *rps16* and *ndhF*) have been used for understanding phylogeny in plants at different systematic levels in plant kingdom^{4,5,8–16}. For assessment of taxonomic status and investigation of the evolutionary relationship among selected genera of family *Asteraceae* molecular marker based on chloroplast DNA was used^{17,18}. In the present study, molecular and bioinformatics analyses were carried out to evaluate genetic variability among different species of family *Asteraceae* on the basis of a chloroplast gene *rps 11*.

Materials and methods

Collection of plant material

Selected plants of the *Asteraceae* family were collected from different geographical areas of Pakistan including Islamabad, Gujar Khan and Peshawar following established protocol and permission was obtained. Our plant study complies with relevant institutional, national and international guidelines and legislation. Further, the plant material was identified by Javed Iqbal and taxonomically validated by well-known taxonomist Dr. Syed Afzal Shah, Assistant Professor at Department of Biological Sciences, National University of Medical Sciences Rawalpindi, Pakistan. The voucher specimens were submitted to gene bank of the National University of Medical Sciences Rawalpindi, Department of Biological Sciences, for allotment of voucher number SAS-675, multiplication and to make sure availability for future use. (Table 1).

DNA isolation and primer designing

Complete genomic DNA from fresh leaves was extracted using the CTAB (Cetyl Trimethyl Ammonium Bromide) method as demonstrated by Richards^{19,20}. Forward and reverse primer of *rps11* gene were designed by online web server primer 3 (version 4.0) (<http://primer3.sourceforge.net/>). The Tobacco cpDNA sequence (Accession # Z00044.2) available in Genbank was used to design *rps11* gene primers. The sequence of forward and reverse primers is;

rps 11 Forward: 5' TGGCAAAAGCTATACCGAAAA 3'.
rps 11 Reverse: 5' TTCGAGGTCTACAGCCATT 3'.

Amplification and sequencing of *rps 11* gene from different species of *Asteraceae*

To amplify *rps 11* gene, 25 µL of reaction mixture containing 1 µl template, 1 µl (25 pmol) of each primer, 12.5 µl PCR Master Mix (Fermentas) and 9.5 µl nano pure water was prepared. The amplification conditions were 35 cycles of denaturation at 94 °C for 1 min, annealing at 55 °C for 30 s, and extension at 72 °C for 1 min, followed by a single-step of final extension at 72 °C for 20 min. The PCR products were confirmed on 1.5% agarose gel.

Serial no.	Plant Species	Location	Accession no.
1	<i>Artemisia scoparia</i>	Islamabad	KU859955
2	<i>Erigeron conyzanthus</i>	Islamabad	KU859957
3	<i>Cosmos bipinnatus</i>	Islamabad	KU859958
4	<i>Chrysanthemum indicum</i>	Islamabad	KU859959
5	<i>Carthamus oxyacantha</i>	Islamabad	KU859960
6	<i>Bidens biternata</i>	Peshawar	KU859956
7	<i>Launaea sarmentosa</i>	Islamabad	KU862618
8	<i>Launaea procumbens</i>	Islamabad	KU862619
9	<i>Lactuca sativa</i>	Islamabad	KU862620
10	<i>Parthenium hysterophorus</i>	Islamabad	KU862621
11	<i>Sonchus asper</i>	Islamabad	KU862622
12	<i>Sonchus arvensis</i>	Islamabad	KU862623
13	<i>Silybum marianum</i>	Islamabad	KU862624
14	<i>Sonchus maritimus</i>	Islamabad	KU862625
15	<i>Tagetes minuta</i>	Gujar khan	KU862626
16	<i>Xanthium strumarium</i>	Islamabad	KU726095

Table 1. Area of collection and GenBank accession numbers of chloroplast *rps 11* gene of selected species of *Asteraceae*.

Confirmed PCR products were purified with the help of the JETquick (Genomed) PCR Product Purification Kit and samples were sent to Macrogen (South Korea) for sequencing.

Phylogenetic and protein structural analysis

Phylogenetic trees were constructed based on nucleotide as well as amino acid sequences of *rps 11* representing sixteen selected species of *Asteraceae* to study the phylogenetic relationship by using Maximum Likelihood (ML). For 3D structural analysis, protein models were built using i-TASSER and were evaluated by Ramachandran plot analysis and PROCHECK. Ramachandran plot analysis and PROCHECK was used for protein structure validation. 3D models pdb files were uploaded on SAVES (The Structure Analysis and Verification Server) for Ramachandran plot and PROCHECK analysis.

Results and discussion

Isolated genomic DNA of good quality from sixteen selected species of *Asteraceae* was run on 1% agarose gel for confirmation. After PCR, amplified products were loaded on 1.5% agarose gel and amplified product of ~ 417 bp was observed on the gel. All the sixteen amplified products representing sixteen species of *Asteraceae* were processed for sequencing and the final sequenced data was submitted to GenBank for the allotment of accession numbers as shown in the Table 1. Both nucleotide and amino acid sequences of *rps 11* were used for inferring genetic characterization. Estimates of genetic variability among *rps 11* sequences from sixteen plant species were calculated by pairwise distance using MEGA6 and moderate to low genetic diversity was observed which was in the range 0.015–0.309 (Fig. 1). The study does not provide sufficient evidence to support the *rps 11* gene's potential for use in phylogenetic analysis. Moreover, methodologies we used is based on neutral theory which is still debatable and uncertain. The validity of our findings is therefore contingent upon the limitations of this theoretical framework^{21,22}. We recognize that alternative theories, such as the maximum genetic diversity theory offer a more comprehensive understanding of molecular evolution by incorporating the role of physiological selection and organismal complexity^{23,1}. These alternative approaches may provide a more nuanced explanation of our results, and we encourage future studies to explore these perspectives^{24,25}. The greatest genetic diversity idea will be our primary emphasis in future phylogenetic analyses of this family.

A phylogram was prepared for the *rps 11* gene sequences to establish the evolutionary relationship of sixteen species of *Asteraceae*. The phylogram displayed two major clusters, cluster I and cluster II (Fig. 2). Cluster I is comprised of eleven species namely *Artemisia scoparia*, *Bidens biternata*, *Erigeron conyzanthus*, *Cosmos bipinnatus*, *Chrysanthemum indicum*, *Carthamus oxyacantha*, *Launaea procumbens*, *Xanthium strumarium*, *Tagetes minuta*, *Parthenium hysterophorus* and *Lactuca sativa*. In this cluster bootstrap value was observed in the range of 4–91% which clearly indicated the phylogenetic relationship amongst the selected species. Highest bootstrap value 91% which shows highest similarity was found in *Erigeron conyzanthus*, *Bidens biternata* and *Cosmos bipinnatus* indicating common origin. On the basis of genetic divergence, this cluster was further divided into two subclusters, subcluster I and subcluster II. However, cluster II comprises of three species namely *Silybum marianum*, *Sonchus maritimus* and *Launaea sarmentosa*. *Sonchus asper* and *Sonchus arvensis* appeared as outgroups in cluster II.

The nucleotide data was translated into amino acid sequences using online tool TRANSLATE and translated amino acid sequence data was also used for inferring genetic characterization. A phylogenetic tree was

Plant species	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
<i>Artemisia scoparia</i>																
<i>Bidens biternata</i>	0.07															
<i>Erigeron conyzanthus</i>	0.11	0.10														
<i>Cosmos bipinnatus</i>	0.14	0.12	0.14													
<i>Chrysanthemum indicum</i>	0.015	0.08	0.10	0.13												
<i>Carthamus oxyacantha</i>	0.05	0.11	0.10	0.16	0.03											
<i>Launaea sarmentosa</i>	0.04	0.09	0.12	0.16	0.03	0.03										
<i>Launaea procumbens</i>	0.06	0.09	0.12	0.14	0.06	0.08	0.06									
<i>Lactuca sativa</i>	0.04	0.10	0.10	0.15	0.03	0.03	0.04	0.07								
<i>Parthenium hysterophorus</i>	0.08	0.11	0.11	0.15	0.06	0.07	0.07	0.09	0.08							
<i>Sonchus asper</i>	0.04	0.09	0.12	0.15	0.03	0.04	0.03	0.06	0.04	0.08						
<i>Sonchus arvensis</i>	0.26	0.32	0.38	0.39	0.25	0.28	0.26	0.30	0.26	0.31	0.22					
<i>Silybum marianum</i>	0.04	0.09	0.11	0.15	0.03	0.05	0.03	0.08	0.05	0.06	0.05	0.28				
<i>Sonchus maritimus</i>	0.04	0.09	0.11	0.15	0.03	0.04	0.03	0.07	0.04	0.06	0.04	0.27	0.02			
<i>Tagetes minuta</i>	0.05	0.09	0.09	0.15	0.03	0.04	0.05	0.09	0.03	0.08	0.06	0.28	0.06	0.06		
<i>Xanthium strumarium</i>	0.09	0.16	0.20	0.24	0.11	0.13	0.13	0.16	0.12	0.17	0.13	0.309	0.13	0.12	0.13	

Fig. 1. Estimates of evolutionary deviation for *rps 11* sequences from sixteen plant species, calculated by pairwise distance using MEGA6.

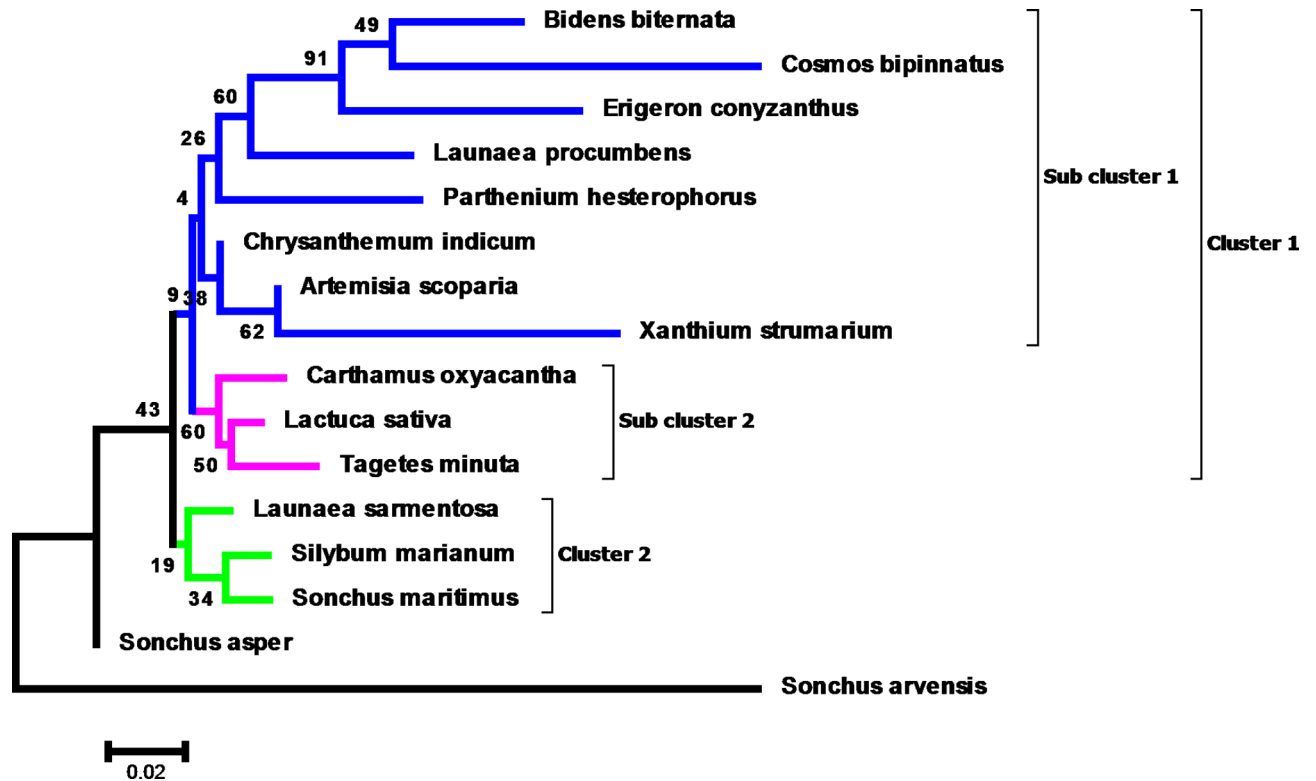


Fig. 2. Phylogenetic tree based on *rps 11* gene sequences demonstrating genetic variability among sixteen selected species of family *Asteraceae* using MEGA6.

constructed using MEGA 6, like the same was it was done for nucleotides data. Two major clusters represented by cluster I and cluster II (Fig. 3) were observed in this tree. In cluster I, there were seven species namely, *Carthamus oxyacantha*, *Launaea procumbens*, *Lactuca sativa*, *Silybum marianum*, *Sonchus maritimus*, *Sonchus asper* and *Launaea sarmentosa*. Highest bootstrap value 68 indicating maximum similarity was found in *Silybum marianum* and *Sonchus maritimus*, indicating their common origin.

Comparison of Maximum likelihood trees based on *rps 11* nucleotide and amino acid sequences showed that protein-based tree has demonstrated higher diversity among studied species in comparison to nucleotide-based tree. The genetic distance was 0.02 in Maximum likelihood tree based on nucleotide sequences whereas it was 0.05 for amino acid base tree.

The bootstrap value range of 4–91% was seen for nucleotides based phylogenetic tree whereas it was 1–100% for phylogenetic tree based on amino acid sequences. Moreover, it was observed that *Sonchus arvensis* is placed as basal group in both phylograms demonstrating that it is an ancestor of cluster II. Our results support the findings of Amar et al. (2016) reporting higher genetic diversity among different species of *Asteraceae*⁴.

Analysis of 3-dimensional protein structures

Three-dimensional Protein models were obtained using I-Tasser. I-Tasser produces five 3-dimensional models for each sequence alongside the confidence value. From 5 built structures for individually *rps 11* protein, topmost model was chosen on the basis of higher percentages of residues in maximum allowed areas and minimum percentage scores in disallowed areas. The outcome suggested that model 1 for all sixteen species predicted by I-TASSER was good model and should be analyzed further. Model 1 of *rps 11* from all sixteen species of *Asteraceae* was visualized using UCSF Chimera as shown in Fig. 4.

Protein structure analysis and validation

PROCHECK can be used to construct Ramachandran plots of 3D protein models and exhibit the placement of amino acids in the “disallowed region” or “allowed region”. Best quality model exhibit about $\leq 2\%$ residues in the disallowed region and $\geq 90\%$ residues in the favored region of the plot. Estimation and verification consequences from Ramachandran plot assessment displayed that the *rps 11* protein structures of *Tagetes minuta*, *Xanthium strumarium*, *Lactuca sativa* and *Chrysanthemum indicum* have best feature models with about $> 90\%$ of residues in the allowed region and $\leq 2\%$ in the disallowed region (Table 2). Keeping in mind the Ramachandran scores it was observed that Ramachandran plot for all the sixteen sequences were considered satisfactory (Fig. 5). In an earlier study, Ramachandran plot study was carried out for the structural investigation of wound-inducible proteinase inhibitor-II gene¹⁷. Similarly in silico analyses were conducted for structural modeling of antioxidant proteins of spinach by Ramachandran plot analysis¹⁸. To the best of our knowledge this is the first report describing the 3D protein structure and Ramachandran plot analysis for *rps 11* protein in *Asteraceae* family.

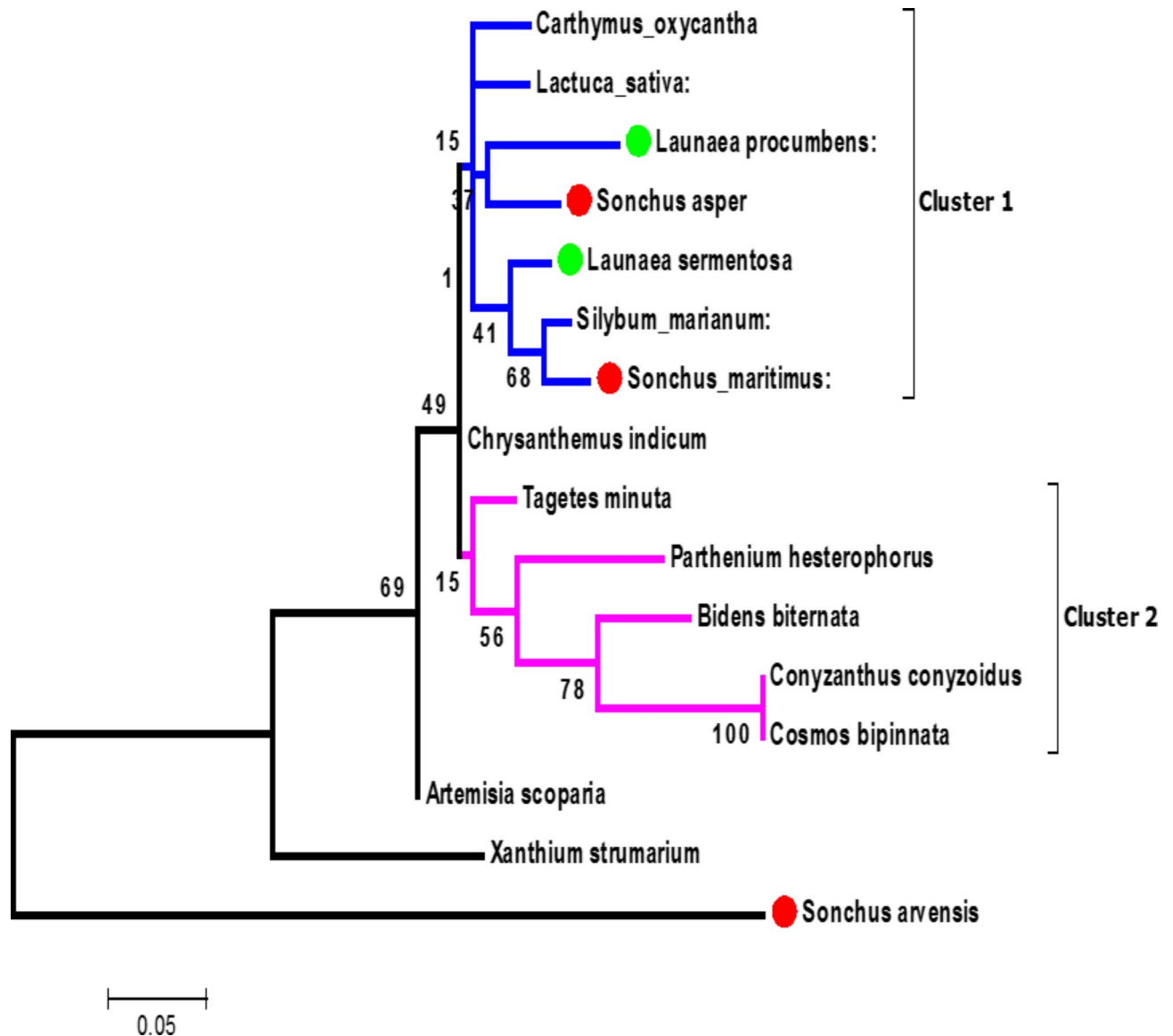


Fig. 3. Phylogram representing genetic variability among sixteen selected species of family *Asteraceae* using MEGA6 based on *rps 11* protein sequences.

Conclusion

Our study demonstrates the utility of the *rps11* primer for phylogenetic analysis in the Asteraceae family. The genetic diversity revealed among the sixteen selected species highlights the complexity and richness of this family. The generated 3D protein models of the *rps11* protein showed high structural quality, with satisfactory plot statistics and Ramachandran plots indicating accurate predictions. Notably, the models for *Tagetes minuta*, *Xanthium strumarium*, *Lactuca sativa*, and *Chrysanthemum indicum* exhibited exceptional quality, with over 90% of residues in the allowed region and less than 2% in the disallowed region. Furthermore, the 3D protein modeling of the *rps11* protein offers a new perspective on the structural conservation and variation within this family. While our current research provides valuable insights, it is insufficient to independently confirm the *rps11* gene's potential as a reliable marker for phylogenetic analysis. In future studies, we plan to explore the application of the maximum genetic diversity theory to phylogenetic analysis of this family, which may offer a more comprehensive understanding of its evolutionary relationships.

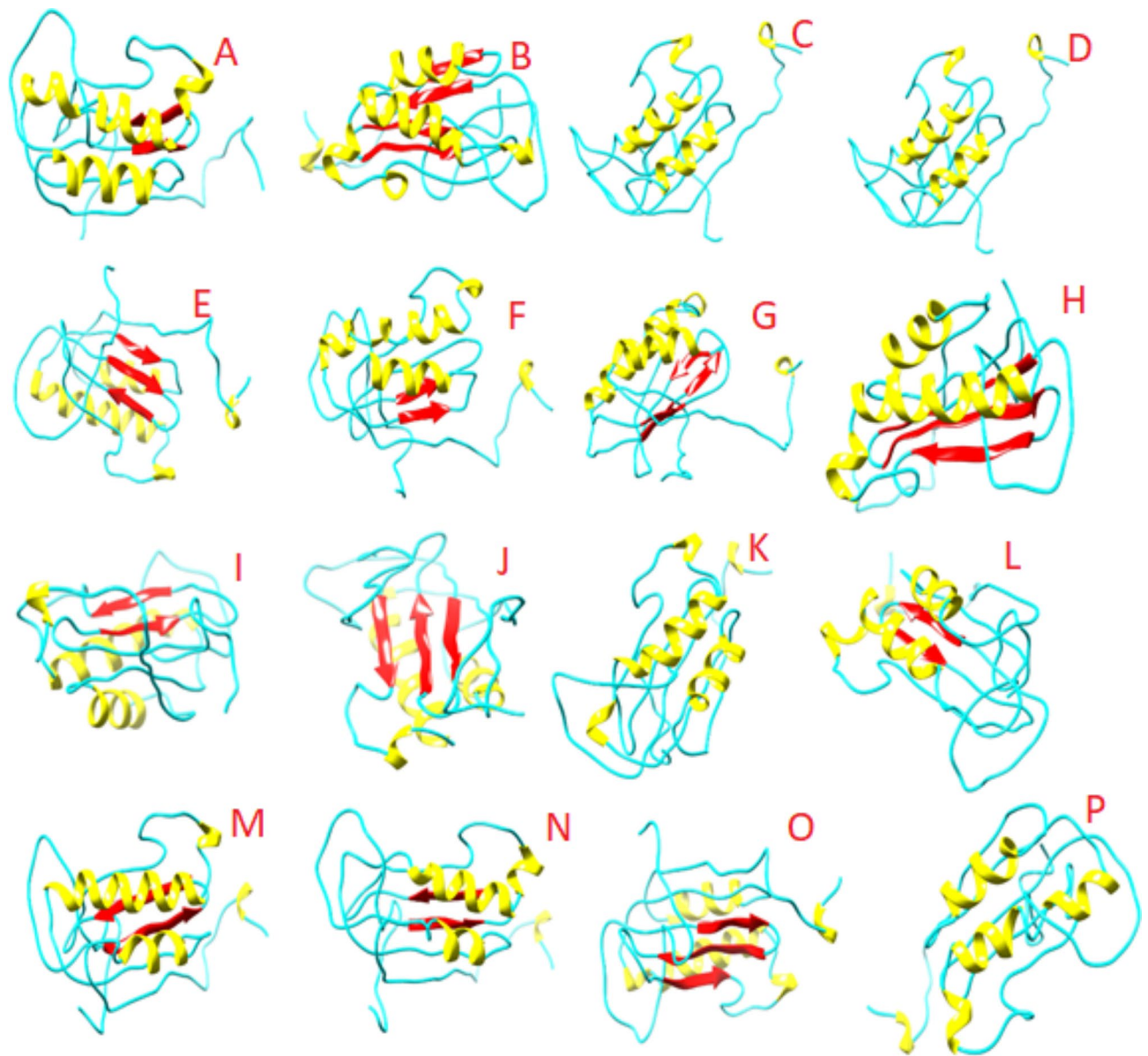


Fig. 4. The 3D protein model with the highest C-score built by I-TASSER for different species of family Asteraceae A: *Artemisia scoparia*, B: *Bidens biternata*, C: *Erigeron conyzanthus*, D: *Cosmos bipinnatus*, E: *Chrysanthemum indicum*, F: *Carthamus oxyacantha*, G: *Launaea sermentosa*, H: *Launaea procumbens*, I: *Lactuca sativa*, J: *Parthenium histerophorus*, K: *Sonchus asper*, L: *Sonchus arvensis*, M: *Silybum marianum*, N: *Sonchus maritimus*, O: *Tagetes minuta*, P: *Xanthium strumarium*

S.N	Plant species	Allowed region	Disallowed region
1	<i>Artemisia scoparia</i>	93%	7.0%
2	<i>Bidens biternata</i>	94.8%	5.2%
3	<i>Carthamus oxyacantha</i>	97.4%	2.6%
4	<i>Chrysanthemum indicum</i>	98.2%	1.8%
5	<i>Erigeron conyzanthus</i>	95.7%	4.3%
6	<i>Cosmos bipinnatus</i>	95.7%	4.3%
7	<i>Lactuca sativa</i>	98.2%	1.8%
8	<i>Launaea procumbens</i>	96.5%	3.5%
9	<i>Launaea sermentosa</i>	94.7%	5.3%
10	<i>Parthenium histerophorus</i>	97.4%	2.6%
11	<i>Silybum marianum</i>	95.6%	4.4%
12	<i>Sonchus arvensis</i>	93.6%	6.4%
13	<i>Sonchus asper</i>	95.7%	4.3%
14	<i>Sonchus maritimus</i>	97.4%	2.6%
15	<i>Tagetes minuta</i>	98.3%	1.7%
16	<i>Xanthium strumarium</i>	98.2%	1.8%

Table 2. Ramachandran scores of models for all the sixteen sequences predicted by PROCHECK.

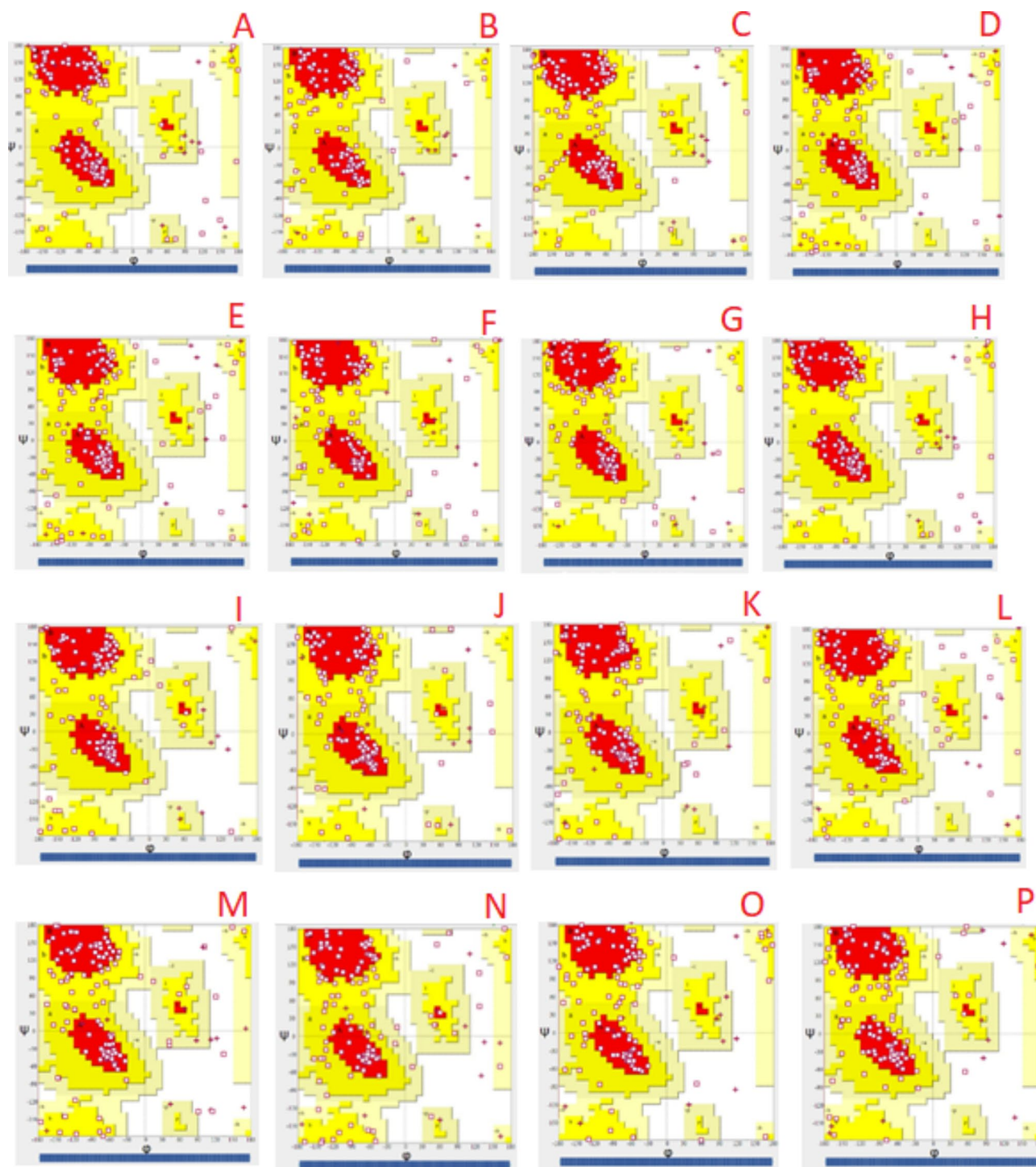


Fig. 5. The Ramachandran plot for *rps 11* protein of different species of family *Asteraceae* built by SAVES A: *Artemisia scoparia*, B: *Bidens biternata*, C: *Erigeron conyzanthus*, D: *Cosmos bipinnatus*, E: *Chrysanthemum indicum*, F: *Carthamus oxyacantha*, G: *Launaea sermentosa*, H: *Launaea procumbens*, I: *Lactuca sativa*, J: *Parthenium histerophorus*, K: *Sonchus asper*, L: *Sonchus arvensis*, M: *Silybum marianum*, N: *Sonchus maritimus*, O: *Tagetes minuta*, P: *Xanthium strumarium*.

Data availability

All the raw data in this research can be obtained from the corresponding authors upon reasonable request.

Received: 16 June 2024; Accepted: 9 October 2024

Published online: 22 October 2024

References

1. Anderberg, A. A., Baldwin, B. G., Bayer, R. G., Breitwieser, J., Jeffrey, C., Dillon, M. O., Eldenäs, P., Funk, V., Garcia-Jacas, N., Hind, D. J. N. and Karis, P. O., 2007. *Compositae: Compositae Adans., Fam. Pl. 2: 103 (1763), nom. alt. et cons. Asteraceae Martynov, Tekhno-Bot. Slovar: 55 (1820), nom. cons pp. (61–588)*. Springer Berlin Heidelberg.
2. Funk, V. A., Susanna, A., Stuessy, T. F. & Robinson, H. Classification of compositae. *Syst. Evolut. Biogeogr. Compos.* 171–192 (2009).
3. Wang, Y. et al. Complete chloroplast genome sequence of *Aquilaria sinensis* (Lour.) Gilg and evolution analysis within the Malvales order. *Front. Plant. Sci.* 7, 280 (2016).
4. Rehman, S., Chaudhary, H. J., Rasheed, A. & Mahmood, T. Phylogenetic relationship of selected Pakistani wheat varieties based on a chloroplast rps11 gene. *J. Anim. Plant. Sci.* 25, 442–447 (2015).
5. Amar, M., Hassan, A. & El Sherbeny, E. A. Assessment of genetic diversity in some wild plants of Asteraceae family by ribosomal DNA sequence. *Egypt. J. Genet. Cytol.* 41, 195–208 (2016).
6. Rahman, A. H. et al. Study of species diversity on the family Asteraceae (Compositae) of the Rajshahi division. *Res. J. Agric. Biol. Sci.* 4, 794–797 (2008).
7. Sambamurty, A. V. *Taxonomy of Angiosperms* 408–417 (IK International Pvt Ltd, 2005).
8. Small, R. L., Lickey, E. B., Shaw, J. & Hauk, W. D. Amplification of noncoding chloroplast DNA for phylogenetic studies in lycophytes and monilophytes with a comparative example of relative phylogenetic utility from Ophioglossaceae. *Mol. Phylogenet. Evol.* 36, 509–522 (2005).
9. Melotto-Passarin, D. M. et al. Phylogenetic relationships in Solanaceae and related species based on cpDNA sequence from plastid trne-trnt region. *Crop Breed. Appl. Biotechnol.* 8, 85–95 (2008).
10. Li, X. et al. Plant DNA barcoding: from gene to genome. *Biol. Rev.* 90(1), 157–166 (2015).
11. Richardson, J. E., Fay, M. F., Cronk, Q. C., Bowman, D. & Chase, M. W. A phylogenetic analysis of Rhamnaceae using rbcL and trnL-F plastid DNA sequences. *Am. J. Bot.* 87, 1309–1324 (2000).
12. Mahmood, T., Hassan, N., Nazar, N. & Naveed, I. Phylogenetic analysis of different *Artemisia* species based on chloroplast gene rps11. *Arch. Biol. Sci.* 63, 661–665 (2011).
13. Ibrahim, M., Nazar, N., Ilyas, M. & Mahmood, T. Assessment of genetic variability among selected species of Apocynaceae on the basis of rps 11 gene. *J. Anim. Plant. Sci.* 24, 1266–1269 (2014).
14. Jamil, I. et al. Subfamilial relationships within Solanaceae as inferred from atpβ-rbcL intergenic spacer. *Pak J. Bot.* 46(6), 2095–2100 (2014).
15. Akhtar, W., Rasheed, A. W., Shinwari, Z. K., Naqvi, S. M. & Mahmood, T. Genetic characterization of different Pakistani date palm varieties. *Pak J. Bot.* 46(6), 2095–2100 (2014).
16. Munir, F., Naqvi, S. M. S. & Mahmood, T. In vitro and in silico characterization of *Solanum lycopersicum* wound-inducible proteinase inhibitor-II gene. *Turk. J. Biol.* 37, 1–10 (2013).
17. Richards, E. J. Preparation of plant DNA using CTAB. *Short. Protoc. Mol. Biol.* 3, 2–10 (1997).
18. Dinçtürk, B. Recent advances in protein folding. *Turk. J. Biol.* 20, 147–152 (1996).
19. Sambamurty, A. V. *Taxonomy of Angiosperms* (IK International Pvt Ltd, 2005).
20. Samuel, R., Stuessy, T. F., Tremetsberger, K., Baeza, C. M. & Siljak-Yakovlev, S. Phylogenetic relationships among species of *Hypochoeris* (Asteraceae, Cichorieae) based on ITS, plastid trnL intron, trnL-F spacer, and matK sequences. *Am. J. Bot.* 90, 496–507 (2003).
21. Kreitman, M. The neutral theory is dead. Long live the neutral theory. *Bioessays*. 18(8), 678–683 (1996).
22. Kern, A. D. & Hahn, M. W. The Neutral Theory in Light of Natural selection. *Mol. Biol. Evol.* 35(6), 1366–1371 (2018).
23. Huang, S. The maximum genetic diversity theory of molecular evolution. *Commun. Inf. Syst.* 23(4), 359–392 (2023).
24. Gillespie, J. H. *The Causes of Molecular Evolution* Vol. 2 (Oxford University Press, 1991).
25. Eyre-Walker, A. & Keightley, P. D. The distribution of fitness effects of new mutations. *Nat. Rev. Genet.* 8(8), 610–618 (2007).

Acknowledgements

The authors extend their appreciation to the Researchers supporting project number (RSP2024R173), King Saud University, Riyadh, Saudi Arabia.

Author contributions

S.A.Z, J.I, B.A.A, and S.K, designed and conceived the study idea. S.A.Z, completed the experiments. R.I and T.M, analyzed the data and performed visualizations and statistical data analysis. S.A.Z, J.I, B.A.A, and S.K wrote the original draft. M.S.A, M.S.E, M.R, R.I and T.M, reviewed and edited the manuscript. S.A.Z, J.I, B.A.A, and S.K, provided the resources and supervision. All authors made valuable revisions, edited the manuscript, and approved the last version.

Funding

Open Access funding enabled and organized by Projekt DEAL. Researchers Supporting Project number (RSP2024R173) King Saud University, Riyadh, Saudi Arabia.

Declarations

Ethics approval and consent to participate

This study does not include human or animal subjects.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Statement on guidelines

All experimental studies and experimental materials involved in this research are in full compliance with relevant institutional, national and international guidelines and legislation posing a conflict or bias.

Additional information

Correspondence and requests for materials should be addressed to J.I., M.R., R.I. or T.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024