



Deciphering Millet Diversity: Proteomic Clusters and Phylogenetic Insights

Anitha Ravichandran ^a, Saranya N. ^{a*}, Jayakanthan Mannu ^a,
Bharathi N. ^a, Senthil Natesan ^a, Anandhi Venugopal ^b,
Ramanathan Sowdhamini ^c

^a Department of Plant Molecular Biology and Bioinformatics, Center for Plant Molecular Biology and Biotechnology, TNAU, Coimbatore-03, India.

^b Department of Physical Sciences and Information Technology, AEC&RI, TNAU, Coimbatore-03, India.

^c National Center for Biological Science, Bangalore, India.

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/IJPSS/2023/v35i203792

Open Peer Review History:

This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: <https://www.sdiarticle5.com/review-history/106594>

Original Research Article

Received: 11/07/2023

Accepted: 16/09/2023

Published: 18/09/2023

ABSTRACT

Millets are renowned for their climatic resilience and possess high nutritive value with wide genetic variations. In countries like India and Africa, millets are part of many people's regular diets with rich sources of protein, dietary fiber, polyphenols, minerals, vitamins, and other nutrients. The proteomic signatures of several millet species, including Fonio, Finger, Proso, Sorghum, and Foxtail millet, were examined in this study. We have performed orthologous analysis to discover both common and distinctive protein clusters among these species by using the OrthoFinder algorithm in conjunction with visualization tools. A total of 16,247 clusters were shared by all species, offering light on similar evolutionary or adaptation mechanisms. The strong representation of Gene Ontology (GO) categories related to osmotic stress, water deprivation, and temperature stresses in the research further highlighted the millets' powerful adaptative responses to various

*Corresponding author: E-mail: saranya.n@tnau.ac.in;

environmental difficulties. Intricate signaling mechanisms for wound, defense, and growth are also revealed by their efficient photosynthetic capacities. However, each species' distinctive clusters, particularly those in Finger millet, highlighted how it differed from other millets. The evolutionary links were further clarified by a phylogenetic tree built using the Maximum likelihood approach and the JTT+CAT evolutionary model, with Foxtail and Proso millets showing a closer kinship. The research sheds light on the complex genetic network of millets, evolutionary histories, and potential adaptive processes. The identification of 2,277 clusters, which are mainly shared by foxtail, proso, fonio, and sorghum millets and support the distinct evolutionary history of finger millet, was especially important. These millets' strong adaptive mechanisms, which are on display in clusters related to different response mechanisms, demonstrate their evolutionary skill and point to prospective directions for crop improvement and resilience techniques.

Keywords: Millets; proteomic signatures; OrthoFinder; gene ontology; evolutionary lineage; adaptive responses.

1. INTRODUCTION

Millets, a small-seeded plant species from the Poaceae (or Gramineae) family of grasses, are well known for their nutritional qualities. Their remarkable versatility allows them to flourish in a variety of climatic conditions, including low moisture, hot, salty, and cold conditions. This adaptability includes the poor, soils lacking in nutrients that can be found in many different parts of the planet [1,2]. Millets are an essential crop because of these qualities, as well as their rapid development and tolerance to abiotic challenges. Particularly in nations like Africa, Russia, India, and China, their grains and forage are irreplaceable sources of food and feed. Millets are a key component of many peoples' everyday diets in places like India and Africa. Millets contain favorable chemicals known as phenolic that are good for human health, including phenolic acids, flavonoids, and tannins. They are also abundant providers of dietary fiber, polyphenols, minerals, vitamins, fatty acids, protein, and other nutrients compared to other cereal crops [3-6].

The most well-known millet cultivars include sorghum (*Sorghum bicolor*), finger millet (*Eleusine coracana*), proso millet (*Panicum miliaceum*), foxtail millet (*Setaria italica*), and pearl millet (*Pennisetum glaucum*). Minor millets including tef (*Eragrostis tef*), kodo millet (*Paspalum scrobiculatum*), and fonio millet (*Digitaria exilis*) also add to the variety of this crop category. These millets have different genome sizes, ploidy, and chromosomal counts, but they have considerable genomic synteny with bioenergy grasses and non-millet cereals. The ancestor relationship showed divergence from a Poaceae family some 70 million years ago, which causes this overlap. Despite being important to agriculture, nothing is known about the genetic

diversity and evolutionary links among the various millet species [2].

Because they frequently retain at least some of their ancestral biological function, orthologs hold great interest in evolutionary biology and genetics [7]. These genes, which develop from a common ancestor during speciation, provide crucial information for phylogenetic and taxonomy studies. Researchers can determine an organism's evolutionary history and degree of relatedness by analyzing the genetic divergence patterns among orthologs [8,9]. Proteomic and transcript data are exploding due to the rapid advancements in high-throughput technologies, and these data are frequently computationally translated into potential polypeptide sequences. As a result, the inclusion of more recent datasets does not restrict the accessibility of orthology data [7]. To find orthologous clusters, two main techniques have been created. The first is based on trees which uses tools like PhyloTreePruner [10] and TreeFam [11] to carry out analysis. The second is supported by tools like PanOCT [12], OrthoFinder [13], and OrthoMCL [14] which rely on graph-based approaches [15].

Our goal in this study is to understand the complex genetic background and phylogenetic linkages present in millets by utilizing these tools and approaches.

2. MATERIALS AND METHODS

2.1 Data Source

From the National Centre for Biotechnology Information (NCBI) database (<https://www.ncbi.nlm.nih.gov/>), whole proteome information for different millet species was obtained in fasta format. The table below specifics the informational sources for each millet species (Table 1).

Table 1. The data sources utilized in the study are summarized and organized in this table

Millet Species	Scientific Name	Common Name	Database	Bioproject	Biosample	Assembly Accession	Reference
Foxtail millet	<i>Setaria italica</i>	Foxtail millet	NCBI	PRJNA32913	SAMN02981383	GCF_000263155.2	[16]
Finger millet	<i>Eleusine coracana subsp. coracana</i>	Caracan millet	NCBI	PRJDB5606	SAMD00076255	GCA_021604985.1	[17]
Fonio millet	<i>Digitaria exilis</i>	Fonio millet	NCBI	PRJNA640067	SAMN15456850	GCA_015342445.1	[18]
Sorghum	<i>Sorghum bicolor</i>	Sorghum	NCBI	PRJNA13876	SAMN02953738	GCA_000003195.3	[19]
Proso millet	<i>Panicum miliaceum</i>	Broom millet	NCBI	PRJNA431363	SAMN08389585	GCA_003046395.2	[20]

2.2 OrthoVenn3 Analysis

In the present study, the OrthoVenn3 web tool was used to better understand the evolutionary connection and identify orthologous groups among the selected species (Shown in Fig. 2) [21]. The analysis was divided into two stages as described below:

2.3 Orthologous Analysis

The OrthoFinder algorithm was utilized to infer orthogroups for our study which is recognized as a method for identifying balanced orthologous gene clusters using similarity search. OrthoFinder employs the Markov clustering algorithm to produce orthologous groups. Parameters such as e-value: 1×10^{-2} were chosen as the criterion for sequence similarity. Further, the annotation and protein similarities option were enabled to functionally classify recognized orthologous groups and to display similarities in protein sequences between orthologous groups, respectively. Cluster Relationship Network was created to understand relationships between various orthologous groups graphically.

2.4 Phylogenetic Tree Construction

In order to find ancestral relationships among millets. FastTree2, which uses the maximum likelihood method to create the evolutionary relationship, was used to build the tree.

JTT+CAT, WAG+CAT, and LG+CAT were three evolutionary models that served as the analysis's foundational frameworks. Notably, the Jones, Taylor, and Thornton (or JTT) matrix is a model of amino acid replacement that was developed empirically. The CAT model, which complements it by taking rate heterogeneity across sites into account, ensures a thorough comprehension of protein evolution.

As part of a thorough look into the tree's dependability, the SH test was used to confirm each node's reliability. After creation, the tree is automatically saved in the Nwk format to support a variety of bioinformatics applications. The JTT + CAT model, which we chose for our study, appeared as the most consistent with our dataset after we considered the three evolutionary models. After setting our preferences and parameters, we clicked "Start Analysis" to begin the analysis.

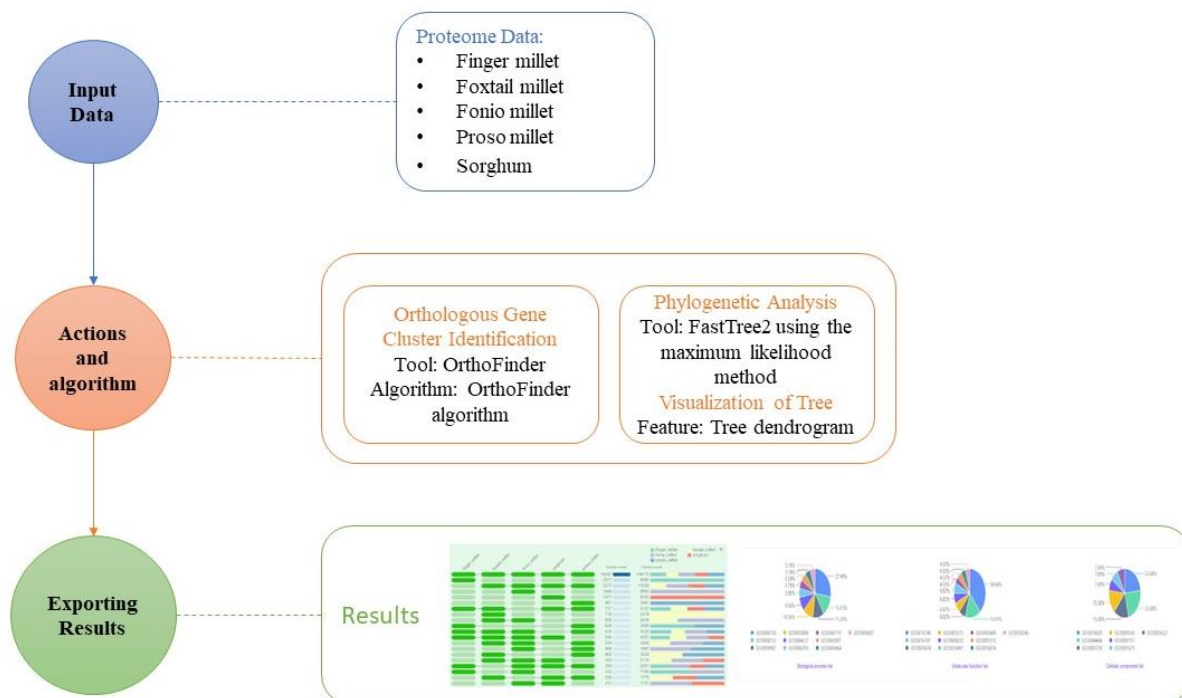


Fig. 1. Using OrthoVenn3, the flowchart shows how the orthologous gene cluster analysis and phylogenetic tree construction procedure works

3. RESULTS

3.1 Comparative Proteomic Analysis Across Millet Species

A total of 25,042 protein clusters containing 61,938 proteins have been found. Notably, 5,213 proteins were identified as singletons in fonio millet. A count of 57,180 proteins that were divided into 23,218 clusters comprising 7,155 proteins was reported as singletons in finger millet. In Proso millet, a total of 55,964 proteins grouped in 24,477 clusters, with 5,141 singletons were observed. In sorghum, 47,110 proteins were clustered into 23,677 number among which 5,514 proteins were singletons. In foxtail millet a total of 43,001 proteins were grouped under 24,311 clusters and 3,995 proteins were classified as singletons. Each millet species has distinct proteomic signatures, which are highlighted by variations in protein distribution and the occurrence of singletons. This highlights the distinctive evolutionary paths and functional potentials of each millet species (shown in Fig. 2).

3.2 Unique and Common Cluster Signatures Across Millets

Although it was shown that all millet species commonly shared 16,247 clusters, some clusters had a substantial number of proteins that had similar related activities. Among the 16,247 clusters, the highest number of sequences was

found with Cluster 3 (114 sequences), Cluster 4 (97 sequences), and Cluster 6 (96 sequences) respectively. The distinct clusters observed in each millet such as finger millet, fonio millet, sorghum, proso millet, and foxtail millet are 2577, 1644, 1277, 981, and 718. These distinct protein concentrations and clusters suggest possible adaptive and specialized roles in each millet species. (See Fig. 3). Important Go terms of the common shared clusters are listed in Table 2.

3.3 Phylogenetic Tree Analysis Findings

The JTT+CAT evolutionary model is used to generate the phylogenetic tree, which illustrates species relationships using the maximum likelihood method.

The phylogenetic tree's branching structure and branch lengths shed light on the evolutionary links between the many millet species. For instance, it shows the more recent common ancestor, indicating a closer evolutionary relationship, For example, Foxtail and Proso millet are shown on the tree closer together than the Finger millet.

These conclusions are supported by the clustering pattern, which also suggests possible evolutionary links and distances. Particularly, Fonio millet stands out with the most ranged orthogroups based on protein cluster counts. Sorghum and finger millet, on the other hand, exhibit considerably fewer orthogroups.

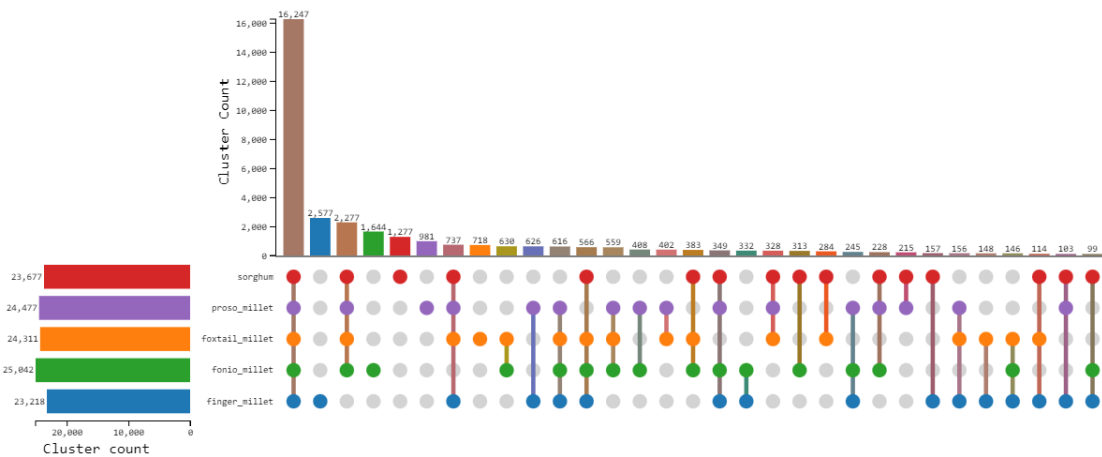


Fig. 2. The graphical representations give a precise picture of the protein among several millet species. They outline how these proteins are organized into clusters and show the overall number of proteins that have been found for each species. By highlighting both the common and distinctive protein clusters they contain, this graphic representation highlights the distinctive proteomic profiles of each millet species. They also show the number of genes under each orthologous groupings, and singletons in each species

Table 2. Consists of GO terms along with its ID

GO terms	GO id
Response to jasmonate	GO:0009753
Response to gibberellin	GO:0009739
Response osmotic stress	GO:0006970
Response water deprivation	GO:0009414
Response to heat stress	GO:0009408
Response to cold stress	GO:0009409
Photosystem II stabilization	GO:0042549
Photosynthesis light reaction	GO:0019684
Response to red or far-red light	GO:0009639
Defense response	GO:0006952
Plant-type hypersensitive response	GO:0009626
Response to nematode	GO:0009624
Negative regulation of gene silencing	GO:0060969
Positive regulation of cytokinin-activated signaling	GO:0080038
Response to salicylic acid	GO:0009751

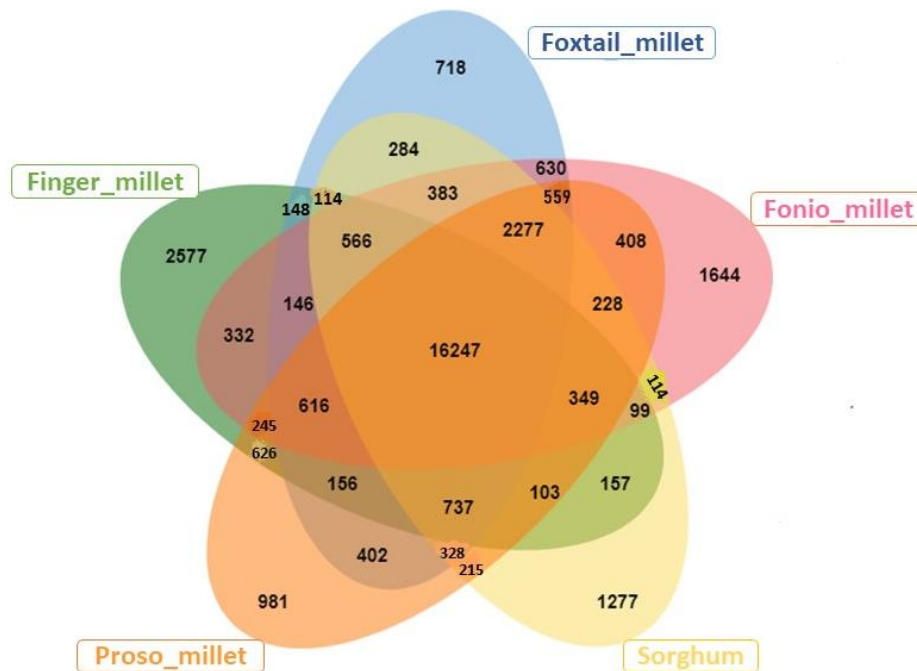


Fig. 3. Visual representation of the cluster distribution among all the millets

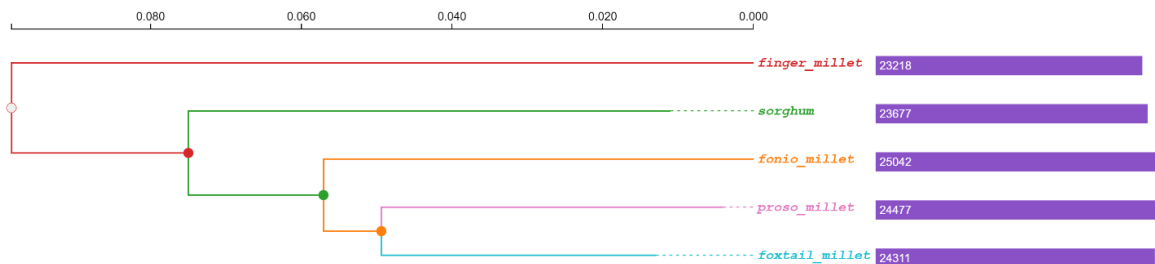


Fig. 4. The above graph gives a thorough review of the significance of orthogroups and evolutionary relationships in the five millet species

4. DISCUSSION

There are 16,247 clusters shared among all millet species, and these clusters are particularly rich in certain Gene Ontology (GO) terminology indicating shared biological processes and molecular function (refer to the supplementary files). These consist of responses to hormonal signals in which the response to jasmonate highlights the species' potential role in signaling for wound and defense. Response to gibberellin, which plays a role in growth and development [22,23]. The protein sequences are linked to adaptive responses to stress, including those to osmotic stress, water deprivation, heat, and cold stress. These point to the millet's ability to adapt to various environmental difficulties [24,25,26]. GO phrases like photosystem II stabilization and photosynthesis light reaction highlight their photosynthetic capacities, while the response to red or far-red light may provide a hint as to their photomorphogenic responses [27,28]. Clusters of proteins involved in defense response, plant-type hypersensitive response, and response to nematode are revealed, demonstrating the plant's resistance to pathogens and pests [29,30].

Some of the protein sequences involved in regulation and feedback mechanisms that emphasize their regulatory roles in developmental and defense pathways, such as negative regulation of gene silencing, positive regulation of cytokinin-activated signaling, and response to salicylic acid [30,31,32]. The presence of 2,277 clusters that were shared only by foxtail, proso, fonio, and sorghum millets is a curious finding in addition to the clusters that were universally shared among all species. This unusual grouping suggests that finger millet differs from the other millets in our study more than the others. Additionally, foxtail and proso millets share 402 clusters when the overlap of clusters across individual species is taken into account which shows a closer evolutionary relationship or shared environmental pressures compared to other millet pairs. When compared to other pairwise combinations, they have a much higher number of common clusters, which shows a stronger evolutionary link. Since many of these shared clusters are involved in drought tolerance, this information may shed light on the typical environmental problems that millets encountered throughout their evolutionary history. With 25,042 clusters, Fonio millet has the most varied protein assembly, whereas Finger millet and sorghum had 23,218 and 23,677

clusters, respectively. This could be a connection to the complex genetic and environmental interactions that shape each species.

These millets' adaptive strategies may be shown by the presence of specific clusters relating to response mechanisms, such as "response to jasmonic acid", "response to osmotic stress," and "response to gibberellin." It would be fascinating to investigate whether these particular response mechanisms are related in any way to the particular environmental circumstances where these millet species are mostly grown.

5. CONCLUSION

We have outlined the complicated proteome landscape that distinguishes different millet species through our thorough study of those species. The extraordinary genetic variety of these crops is shown by the diverse protein clusters seen in different species, particularly the notable clusters found in Fonio millet. The distinctive clusters found in each species highlight their unique evolutionary pathways and potential adaptation processes, even if a sizable number of clusters are similar across all species, suggesting a common evolutionary track or adaptive response. Notably, shared clusters show a closer evolutionary link between Foxtail and Proso millets, suggesting that they may have shared ancestry or environmental adaptations. Exciting possibilities about the adaptive strategies used by these millets over evolutionary periods are presented by the identified response mechanisms, especially those linked to environmental challenges. Our research, which is based on careful examination on evolutionary models, opens the way for further investigation of millets' functional biology and genomes, which may eventually inform conservation and crop-improvement plans.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. McDonough CM, Rooney LW, Serna-Saldivar SO. The millets. Food Science And Technology-New York-Marcel Dekker. 2000;177-202.
2. Muthamilarasan M, Dhaka A, Yadav R, Prasad M. Exploration of millet models for

- developing nutrient rich graminaceous crops. *Plant Science* 2016;242:89-97.
3. Amadou I, Gounga ME, Le GW. Millets: Nutritional composition, some health benefits and processing-A review. *Emirates Journal of Food and Agriculture*. 2013;501-508.
 4. Hassan ZM, Sebola NA, Mabelebele M. The nutritional use of millet grain for food and feed: A review. *Agriculture & food security*. 2021;10:1-14.
 5. Saleh AS, Zhang Q, Chen J, Shen Q. Millet grains: Nutritional quality, processing, and potential health benefits. *Comprehensive reviews in food science and food safety*. 2013;12(3):281-295.
 6. Lata C. Advances in omics for enhancing abiotic stress tolerance in millets. In *Proc. Indian Natl. Sci. Acad.* 2015;81:397-417.
 7. Lechner M, Findeiß S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: Detection of (co-) orthologs in large-scale analysis. *BMC bioinformatics*. 2011;12(1):1-9.
 8. Van Vlierberghe M, Philippe H, Baurain D. Broadly sampled orthologous groups of eukaryotic proteins for the phylogenetic study of plastid-bearing lineages. *BMC Research Notes*. 2021;14(1):1-4.
 9. Gabaldón T, Koonin EV. Functional and evolutionary implications of gene orthology. *Nature Reviews Genetics*. 2013;14(5):360-366.
 10. Kocot KM, Citarella MR, Moroz LL, Halanych KM. PhyloTreePruner: A phylogenetic tree-based approach for selection of orthologous sequences for phylogenomics. *Evolutionary Bioinformatics*. 2013;9. EBO-S12813.
 11. Schreiber F, Patricio M, Muffato M, Pignatelli M, Bateman A. TreeFam v9: A new website, more species and orthology-on-the-fly. *Nucleic acids research*. 2014;42(D1):D922-D925.
 12. Fouts DE, Brinkac L, Beck E, Inman J, Sutton G. PanOCT: Automated clustering of orthologs using conserved gene neighborhood for pan-genomic analysis of bacterial strains and closely related species. *Nucleic acids research*. 2012;40(22): e172-e172.
 13. Emms DM, Kelly S. OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome biology*. 2015;16(1):1-14.
 14. Li L, Stoeckert CJ, Roos DS. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome research*. 2003;13(9):2178-2189.
 15. Xu L, Dong Z, Fang L, Luo Y, Wei Z, Guo H, Wang Y. OrthoVenn2: A web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic acids research*. 2019;47(W1):W52-W58.
 16. Bennetzen JL, Schmutz J, Wang H, Percifield R, Hawkins J, Pontaroli AC, Devos KM. Reference genome sequence of the model plant *Setaria*. *Nature biotechnology*. 2012;30(6):555-561.
 17. Hatakeyama M, Aluri S, Balachadran MT, Sivarajan SR, Patrignani A, Grüter S, Shimizu KK. Multiple hybrid de novo genome assembly of finger millet, an orphan allotetraploid crop. *DNA Research*. 2018;25(1):39-47.
 18. Wang X, Chen S, Ma X, Yssel AE, Chaluvadi SR, Johnson MS, Van Deynze A. Genome sequence and genetic diversity analysis of an under-domesticated orphan crop, white fonio (*Digitaria exilis*). *GigaScience*. 2021;10(3):giab013.
 19. Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Rokhsar DS. The Sorghum bicolor genome and the diversification of grasses. *Nature*. 2009;457(7229):551-556.
 20. Zou C, Li L, Miki D, Li D, Tang Q, Xiao L, Zhang H. The genome of broomcorn millet. *Nature communications*. 2019;10(1):436.
 21. Sun J, Lu F, Luo Y, Bie L, Xu L, Wang Y. OrthoVenn3: An integrated platform for exploring and visualizing orthologous data across genomes. *Nucleic Acids Research*. 2023;gkad313.
 22. Li L, Li C, Lee GI, Howe GA. Distinct roles for jasmonate synthesis and action in the systemic wound response of tomato. *Proceedings of the National Academy of Sciences*. 2002;99(9):6416-6421.
 23. Attaran E, Major IT, Cruz JA, Rosa BA, Koo AJ, Chen J, Howe GA. Temporal dynamics of growth and photosynthesis suppression in response to jasmonate signaling. *Plant Physiology*. 2014;165(3):1302-1314.
 24. Kido EA, Ferreira Neto JRC, Silva RLDO, Pandolfi V, Guimaraes ACR, Veiga DT, Benko-Iseppon AM. New insights in the sugarcane transcriptome responding to drought stress as revealed by supersage. *The Scientific World Journal*; 2012.

25. Moothoo-Padayachie A, Macdonald A, Varghese B, Pammenter NW, Govender P. Uncovering the basis of viability loss in desiccation sensitive *Trichilia dregeana* seeds using differential quantitative protein expression profiling by iTRAQ. *Journal of plant physiology*. 2018;221:119-131.
26. Estravis-Barcala M, Mattera MG, Soliani C, Bellora N, Opgenoorth L, Heer K, Arana MV. Molecular bases of responses to abiotic stress in trees. *Journal of Experimental Botany*. 2020;71(13):3765-3779.
27. Li YY, Han M, Wang RH, Gao MG. Comparative transcriptome analysis identifies genes associated with chlorophyll levels and reveals ph Li P, Yu J, Feng N, Weng J, Rehman A, Huang J, Niu Q. Physiological and transcriptomic analyses uncover the reason for the inhibition of photosynthesis by phosphate deficiency in *Cucumis melo* L. *International Journal of Molecular Sciences*. 2022;23(20):12073.
28. Li P, Yu J, Feng N, Weng J, Rehman A, Huang J, Niu Q. Physiological and transcriptomic analyses uncover the reason for the inhibition of photosynthesis by phosphate deficiency in *Cucumis melo* L. *International Journal of Molecular Sciences*. 2022;23(20):12073.
29. Altegoer F, Weiland P, Giammarinaro PI, Freibert SA, Binnebesel L, Han X, Bange G. The two paralogous kiwellin proteins KWL1 and KWL1-b from maize are structurally related and have overlapping functions in plant defense. *Journal of Biological Chemistry*. 2020;295(23):7816-7825.
30. McDowell JM, Dangl JL. Signal transduction in the plant immune response. *Trends in biochemical sciences*. 2000;25(2):79-82.
31. Zubo YO, Schaller GE. Role of the cytokinin-activated type-B response regulators in hormone crosstalk. *Plants*. 2020;9(2):166.
32. Argueso CT, Ferreira FJ, Epple P, To JP, Hutchison CE, Schaller GE, Kieber JJ. Two-component elements mediate interactions between cytokinin and salicylic acid in plant immunity. *PLoS Genetics*. 2012;8(1):e1002448.

© 2023 Ravichandran et al.; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:

The peer review history for this paper can be accessed here:
<https://www.sdiarticle5.com/review-history/106594>