

# Original Research Article

## Deciphering Millet Diversity: Proteomic Clusters and Phylogenetic Insights

---

### 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 protein sources, 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 common and distinctive protein clusters among these species 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. However, each species' distinctive clusters, particularly those in Finger millet, highlighted how it differed from other millets. A phylogenetic tree built using the Maximum likelihood approach and the JTT+CAT evolutionary model further clarified the evolutionary links, 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. These results open the avenue for focused genomic studies and millet crop development techniques.

*Keywords: Millets, OrthoFinder, Crop improvement, Evolutionary relationship, Clusters*

### 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 various 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](McDonough, Rooney, Serna-Saldivar, & DEKKER-, 2000)[2]. Millets are an essential crop because of these qualities and their rapid development and tolerance to abiotic challenges [6]. Particularly in nations like Africa, Russia, India, and China, their grains and forage are irreplaceable food and feed sources. 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][5][4].

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, including 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, using tools like PhyloTreePruner [10] and TreeFam [11] to conduct analysis. The second is supported by tools like PanOCT [12], OrthoFinder [13], and OrthoMCL [14] which rely on graph-based approaches [15]. This study aims to understand the complex genetic background and phylogenetic linkages present in millets by utilizing these tools and approaches.

## 2. MATERIAL AND METHODS

### 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 details each millet species' informational sources (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]

### OrthoVenn3 Analysis:

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

### Orthologous Analysis:

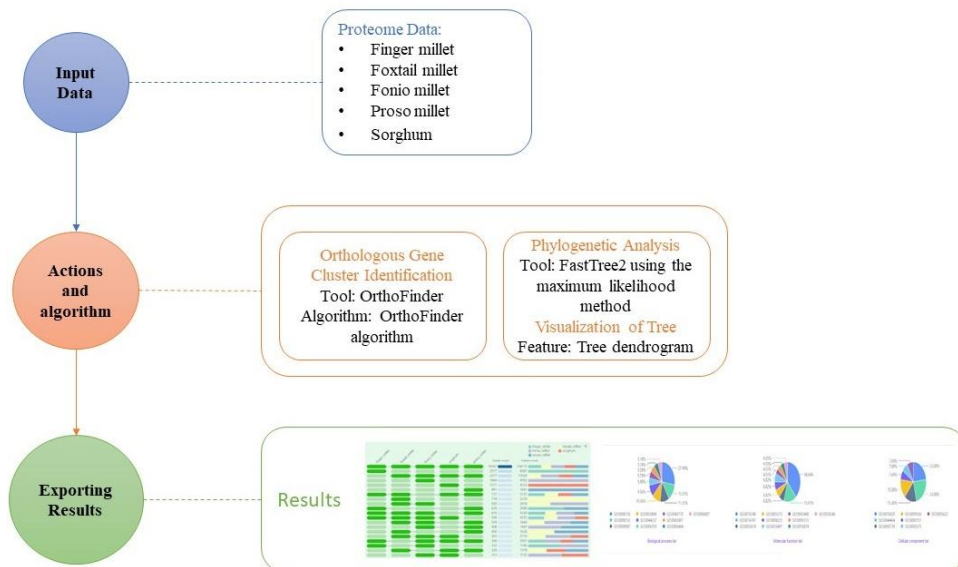
The OrthoFinder algorithm was utilized to infer orthogroups for our study which is recognized as a method for identifying balanced orthologous gene clusters using similarity search. OrthoFinder employs the Markov clustering algorithm to produce orthologous groups. Parameters such as e-value:  $1 \times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

### 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 considering rate heterogeneity across sites, ensures a thorough comprehension of protein evolution.

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

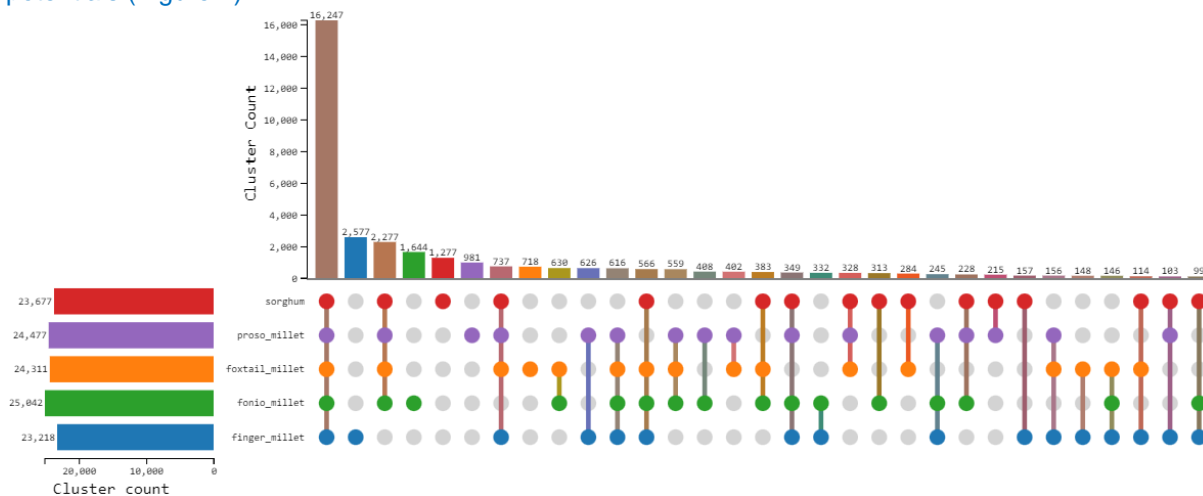


**Figure 1:** Using OrthoVenn3, the flowchart shows how the orthologous gene cluster analysis and phylogenetic tree construction procedure works.

### 3. RESULTS

#### 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 numbers 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 protein distribution variations and singletons' occurrence. This highlights each millet species' distinctive evolutionary paths and functional potentials (Figure 2).



**Figure 2:** The graphical representations precisely picture 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 group, and singletons in each species.

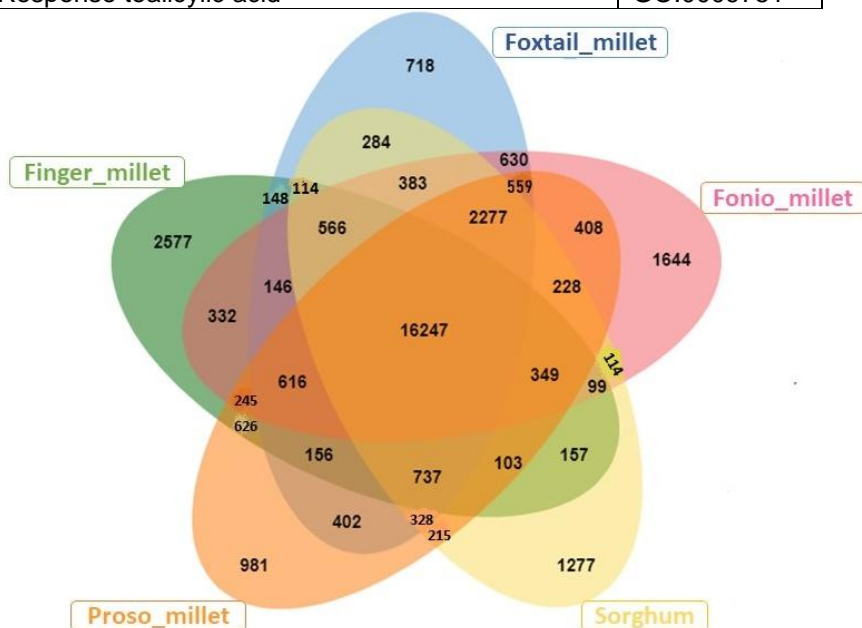
#### Unique and Common Cluster Signatures Across Millets:

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 each millet species' possible adaptive and specialized roles. (See Figure 3). Important Go terms of the common shared clusters are listed in Table 2.

**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



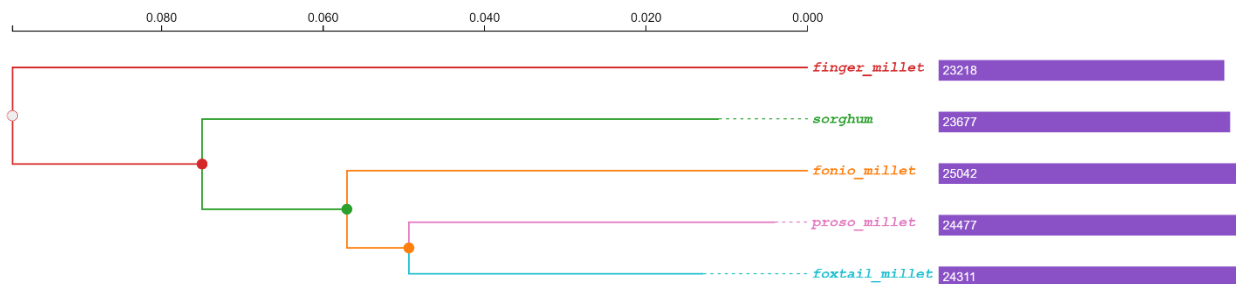
**Figure 3:** Visual representation of the cluster distribution among all the millets.

**Phylogenetic tree analysis findings:**

The JTT+CAT evolutionary model generates 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.



**Figure 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 response 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 [21][22]. 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 [23][24]. GO phrases like photosystem II stabilization and photosynthesis light reaction highlight their photosynthetic capacities, while the response to red or far-red light may hint at their photomorphogenic responses [26][27]. 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 [28][29].

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 [29][30][31]. 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.

## REFERENCES

1. MCDONOUGH, C. M., ROONEY, L. W., & SERNA-SALDIVAR, S. O. (2000). THE MILLETS. *FOOD SCIENCE AND TECHNOLOGY-NEW YORK-MARCEL DEKKER-*, 177-202.
2. MUTHAMILARASAN, M., DHAKA, A., YADAV, R., & PRASAD, M. (2016). EXPLORATION OF MILLET MODELS FOR DEVELOPING NUTRIENT RICH GRAMINACEOUS CROPS. *PLANT SCIENCE*, 242, 89-97.
3. AMADOU, I., GOUNGA, M. E., & LE, G. W. (2013). MILLETS: NUTRITIONAL COMPOSITION, SOME HEALTH BENEFITS AND PROCESSING-A REVIEW. *EMIRATES JOURNAL OF FOOD AND AGRICULTURE*, 501-508.
4. HASSAN, Z. M., SEBOLA, N. A., & MABELEBELE, M. (2021). THE NUTRITIONAL USE OF MILLET GRAIN FOR FOOD AND FEED: A REVIEW. *AGRICULTURE & FOOD SECURITY*, 10, 1-14.
5. SALEH, A. S., ZHANG, Q., CHEN, J., & SHEN, Q. (2013). MILLET GRAINS: NUTRITIONAL QUALITY, PROCESSING, AND POTENTIAL HEALTH BENEFITS. *COMPREHENSIVE REVIEWS IN FOOD SCIENCE AND FOOD SAFETY*, 12(3), 281-295.
6. LATA, C. (2015, MARCH). ADVANCES IN OMICS FOR ENHANCING ABIOTIC STRESS TOLERANCE IN MILLETS. IN *PROC. INDIAN NATL. SCI. ACAD* (VOL. 81, PP. 397-417).
7. LECHNER, M., FINDEIß, S., STEINER, L., MARZ, M., STADLER, P. F., & PROHASKA, S. J. (2011). PROTEINORTHO: DETECTION OF (CO-) ORTHOLOGS IN LARGE-SCALE ANALYSIS. *BMC BIOINFORMATICS*, 12(1), 1-9.
8. VAN VLIERBERGHE, M., PHILIPPE, H., & BAURAIN, D. (2021). BROADLY SAMPLED ORTHOLOGOUS GROUPS OF EUKARYOTIC PROTEINS FOR THE PHYLOGENETIC STUDY OF PLASTID-BEARING LINEAGES. *BMC RESEARCH NOTES*, 14(1), 1-4.
9. GABALDÓN, T., & KOONIN, E. V. (2013). FUNCTIONAL AND EVOLUTIONARY IMPLICATIONS OF GENE ORTHOLOGY. *NATURE REVIEWS GENETICS*, 14(5), 360-366.
10. KOCOT, K. M., CITARELLA, M. R., MOROZ, L. L., & HALANYCH, K. M. (2013). PHYLOTREEPRUNER: A PHYLOGENETIC TREE-BASED APPROACH FOR SELECTION OF ORTHOLOGOUS SEQUENCES FOR PHYLOGENOMICS. *EVOLUTIONARY BIOINFORMATICS*, 9, EBO-S12813.
11. SCHREIBER, F., PATRICIO, M., MUFFATO, M., PIGNATELLI, M., & BATEMAN, A. (2014). TREEFAM V9: A NEW WEBSITE, MORE SPECIES AND ORTHOLOGY-ON-THE-FLY. *NUCLEIC ACIDS RESEARCH*, 42(D1), D922-D925.
12. FOUTS, D. E., BRINKAC, L., BECK, E., INMAN, J., & SUTTON, G. (2012). PANOCT: AUTOMATED CLUSTERING OF ORTHOLOGS USING CONSERVED GENE NEIGHBORHOOD FOR PAN-GENOMIC ANALYSIS OF BACTERIAL STRAINS AND CLOSELY RELATED SPECIES. *NUCLEIC ACIDS RESEARCH*, 40(22), E172-E172.
13. EMMS, D. M., & KELLY, S. (2015). ORTHOFINDER: SOLVING FUNDAMENTAL BIASES IN WHOLE GENOME COMPARISONS DRAMATICALLY IMPROVES ORTHOGROUP INFERENCE ACCURACY. *GENOME BIOLOGY*, 16(1), 1-14.
14. LI, L., STOECKERT, C. J., & ROOS, D. S. (2003). ORTHOMCL: IDENTIFICATION OF ORTHOLOG GROUPS FOR EUKARYOTIC GENOMES. *GENOME RESEARCH*, 13(9), 2178-2189.
15. XU, L., DONG, Z., FANG, L., LUO, Y., WEI, Z., GUO, H., ... & WANG, Y. (2019). ORTHOVENN2: A WEB SERVER FOR WHOLE-GENOME COMPARISON AND ANNOTATION OF ORTHOLOGOUS CLUSTERS ACROSS MULTIPLE SPECIES. *NUCLEIC ACIDS RESEARCH*, 47(W1), W52-W58.
16. BENNETZEN, J. L., SCHMUTZ, J., WANG, H., PERCIFIELD, R., HAWKINS, J., PONTAROLI, A. C., ... & DEVOS, K. M. (2012). REFERENCE GENOME SEQUENCE OF THE MODEL PLANT SETARIA. *NATURE BIOTECHNOLOGY*, 30(6), 555-561.
17. HATAKEYAMA, M., ALURI, S., BALACHADRAN, M. T., SIVARAJAN, S. R., PATRIGNANI, A., GRÜTER, S., ... & SHIMIZU, K. K. (2018). MULTIPLE HYBRID DE NOVO GENOME ASSEMBLY OF FINGER MILLET, AN ORPHAN ALLOTETRAPLOID CROP. *DNA RESEARCH*, 25(1), 39-47.
18. WANG, X., CHEN, S., MA, X., YSSEL, A. E., CHALUVADI, S. R., JOHNSON, M. S., ... & VAN DEYNZE, A. (2021). GENOME SEQUENCE AND GENETIC DIVERSITY ANALYSIS OF AN UNDER-DOMESTICATED ORPHAN CROP, WHITE FONIO (*DIGITARIA EXILIS*). *GIGASCIENCE*, 10(3), GIAB013.

19. PATERSON, A. H., BOWERS, J. E., BRUGGMANN, R., DUBCHAK, I., GRIMWOOD, J., GUNDLACH, H., ... & ROKHSAR, D. S. (2009). THE SORGHUM BICOLOR GENOME AND THE DIVERSIFICATION OF GRASSES. *NATURE*, 457(7229), 551-556.
20. ZOU, C., LI, L., MIKI, D., LI, D., TANG, Q., XIAO, L., ... & ZHANG, H. (2019). THE GENOME OF BROOMCORN MILLET. *NATURE COMMUNICATIONS*, 10(1), 436.
21. LI, L., LI, C., LEE, G. I., & HOWE, G. A. (2002). DISTINCT ROLES FOR JASMONATE SYNTHESIS AND ACTION IN THE SYSTEMIC WOUND RESPONSE OF TOMATO. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES*, 99(9), 6416-6421.
22. ATTARAN, E., MAJOR, I. T., CRUZ, J. A., ROSA, B. A., KOO, A. J., CHEN, J., ... & HOWE, G. A. (2014). TEMPORAL DYNAMICS OF GROWTH AND PHOTOSYNTHESIS SUPPRESSION IN RESPONSE TO JASMONATE SIGNALING. *PLANT PHYSIOLOGY*, 165(3), 1302-1314.
23. KIDO, E. A., FERREIRA NETO, J. R. C., SILVA, R. L. D. O., PANDOLFI, V., GUIMARAES, A. C. R., VEIGA, D. T., ... & BENKO-ISEPPON, A. M. (2012). NEW INSIGHTS IN THE SUGARCANE TRANSCRIPTOME RESPONDING TO DROUGHT STRESS AS REVEALED BY SUPERSAGE. *THE SCIENTIFIC WORLD JOURNAL*, 2012.
24. MOOTHOO-PADAYACHIE, A., MACDONALD, A., VARGHESE, B., PAMMENTER, N. W., & GOVENDER, P. (2018). UNCOVERING THE BASIS OF VIABILITY LOSS IN DESICCATION SENSITIVE TRICHILIA DREGEANA SEEDS USING DIFFERENTIAL QUANTITATIVE PROTEIN EXPRESSION PROFILING BY ITRAQ. *JOURNAL OF PLANT PHYSIOLOGY*, 221, 119-131.
25. ESTRAVIS-BARCALA, M., MATTERA, M. G., SOLIANI, C., BELLORA, N., OPGENOORTH, L., HEER, K., & ARANA, M. V. (2020). MOLECULAR BASES OF RESPONSES TO ABIOTIC STRESS IN TREES. *JOURNAL OF EXPERIMENTAL BOTANY*, 71(13), 3765-3779.
26. LI, Y. Y., HAN, M., WANG, R. H., & GAO, M. G. (2021). COMPARATIVE TRANSCRIPTOME ANALYSIS IDENTIFIES GENES ASSOCIATED WITH CHLOROPHYLL LEVELS AND REVEALS PHLI, P., YU, J., FENG, N., WENG, J., REHMAN, A., HUANG, J., ... & NIU, Q. (2022). PHYSIOLOGICAL AND TRANSCRIPTOMIC ANALYSES UNCOVER THE REASON FOR THE INHIBITION OF PHOTOSYNTHESIS BY PHOSPHATE DEFICIENCY IN CUCUMIS MELO L. *INTERNATIONAL JOURNAL OF MOLECULAR SCIENCES*, 23(20), 12073.
27. LI, P., YU, J., FENG, N., WENG, J., REHMAN, A., HUANG, J., ... & NIU, Q. (2022). PHYSIOLOGICAL AND TRANSCRIPTOMIC ANALYSES UNCOVER THE REASON FOR THE INHIBITION OF PHOTOSYNTHESIS BY PHOSPHATE DEFICIENCY IN CUCUMIS MELO L. *INTERNATIONAL JOURNAL OF MOLECULAR SCIENCES*, 23(20), 12073.
28. ALTEGOER, F., WEILAND, P., GIAMMARINARO, P. I., FREIBERT, S. A., BINNEBESEL, L., HAN, X., ... & BANGE, G. (2020). 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*, 295(23), 7816-7825.
29. MCDOWELL, J. M., & DANGL, J. L. (2000). SIGNAL TRANSDUCTION IN THE PLANT IMMUNE RESPONSE. *TRENDS IN BIOCHEMICAL SCIENCES*, 25(2), 79-82.
30. ZUBO, Y. O., & SCHALLER, G. E. (2020). ROLE OF THE CYTOKININ-ACTIVATED TYPE-B RESPONSE REGULATORS IN HORMONE CROSSTALK. *PLANTS*, 9(2), 166.
31. ARGUESO, C. T., FERREIRA, F. J., EPPLE, P., TO, J. P., HUTCHISON, C. E., SCHALLER, G. E., ... & KIEBER, J. J. (2012). TWO-COMPONENT ELEMENTS MEDIATE INTERACTIONS BETWEEN CYTOKININ AND SALICYLIC ACID IN PLANT IMMUNITY. *PLOS GENETICS*, 8(1), E1002448.