Check for updates

Plant Biotechnology Journal
Open Access

Plant Biotechnology Journal (2023) **21**, pp. 2348–2357 doi: 10.1111/pbi.14136

Milletdb: a multi-omics database to accelerate the research of functional genomics and molecular breeding of millets

Min Sun[1,†], Haidong Yan[1,2,3,†], Aling Zhang[1,†], Yarong Jin[1], Chuang Lin[1], Lin Luo[4], Bingchao Wu[1], Yuhang Fan[1], Shilin Tian[5,6], Xiaofang Cao[5], Zan Wang[7], Jinchan Luo[1], Yuchen Yang[1], Jiyuan Jia[1], Puding Zhou[1], Qianzi Tang[8], Chris Stephen Jones[9], Rajeev K. Varshney[10,11], Rakesh K. Srivastava[10], Min He[12], Zheni Xie[1,13], Xiaoshan Wang[1], Guangyan Feng[1], Gang Nie[1], Dejun Huang[14], Xinquan Zhang[1], Fangjie Zhu[4,*] and Linkai Huang[1,12,*]

[1]*College of Grassland Science and Technology, Sichuan Agricultural University, Chengdu, China*
[2]*School of Plant and Environmental Sciences, Virginia Tech, Blacksburg, Virginia, USA*
[3]*Department of Genetics, University of Georgia, Athens, Georgia, USA*
[4]*College of Life Sciences, Fujian Agriculture and Forestry University, Fujian, China*
[5]*Novogene Bioinformatics Institute, Beijing, China*
[6]*Department of Ecology, Hubei Key Laboratory of Cell Homeostasis, College of Life Sciences, Wuhan University, Wuhan, China*
[7]*College of Grassland Science and Technology, China Agricultural University, Beijing, China*
[8]*College of Animal Science and Technology, Sichuan Agricultural University, Chengdu, China*
[9]*Feed and Forage Development, International Livestock Research Institute, Nairobi, Kenya*
[10]*Center of Excellence in Genomics and Systems Biology (CEGSB), International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, India*
[11]*Murdoch's Centre for Crop and Food Innovation, Food Futures Institute, Murdoch University, Murdoch, Western Australia, Australia*
[12]*State Key Laboratory of Crop Gene Exploration and Utilization in Southwest China, Sichuan Agricultural University, Chengdu, Sichuan, China*
[13]*College of Agro-Grassland Science, Nanjing Agricultural University, Nanjing, China*
[14]*Herbivorous Livestock Research Institute, Chongqing Academy of Animal Sciences, Chongqing, China*

Received 7 February 2023;
revised 1 July 2023;
accepted 17 July 2023.
**Correspondence* (Tel 18783575058; fax 2886080; email huanglinkai@sicau.edu.cn (LH); Tel 059-86395360; fax 059-86395360; email fjzhu@fafu.edu.cn (FZ))
[†]These authors contributed equally to this work.

Keywords: millet, database, functional genomics, multi-omics, abiotic stress.

Summary

Millets are a class of nutrient-rich coarse cereals with high resistance to abiotic stress; thus, they guarantee food security for people living in areas with extreme climatic conditions and provide stress-related genetic resources for other crops. However, no platform is available to provide a comprehensive and systematic multi-omics analysis for millets, which seriously hinders the mining of stress-related genes and the molecular breeding of millets. Here, a free, web-accessible, user-friendly millets multi-omics database platform (Milletdb, http://milletdb.novogene.com) has been developed. The Milletdb contains six millets and their one related species genomes, graph-based pan-genomics of pearl millet, and stress-related multi-omics data, which enable Milletdb to be the most complete millets multi-omics database available. We stored GWAS (genome-wide association study) results of 20 yield-related trait data obtained under three environmental conditions [field (no stress), early drought and late drought] for 2 years in the database, allowing users to identify stress-related genes that support yield improvement. Milletdb can simplify the functional genomics analysis of millets by providing users with 20 different tools (e.g., 'Gene mapping', 'Co-expression', 'KEGG/GO Enrichment' analysis, etc.). On the Milletdb platform, a gene *PMA1G03779.1* was identified through 'GWAS', which has the potential to modulate yield and respond to different environmental stresses. Using the tools provided by Milletdb, we found that the stress-related PLATZs TFs (transcription factors) family expands in 87.5% of millet accessions and contributes to vegetative growth and abiotic stress responses. Milletdb can effectively serve researchers in the mining of key genes, genome editing and molecular breeding of millets.

Introduction

Global hunger has been increasing since 2014 (Molotoks *et al*., 2021). According to the Food and Agriculture Organization of the United Nations (FAO, https://www.fao.org/state-of-food-security-nutrition/en/), approximately 720 to 811 million people faced hunger in 2020, with those from Asia and Africa being the worst affected. Global climate change and frequent extreme weather events are important factors leading to global hunger (FAO, 2018), with climate-affected cereal production expected to be 1%–7% by 2060 (Juma and Kelonye, 2016). Consequently, cultivating crops with high-abiotic stress tolerance is essential to ensure an adequate food supply in the future (Varshney *et al*., 2021a,b).

Millets are a collective name for coarse grains (McSteen and Kellogg, 2022), with the characteristics of small grains

© 2023 The Authors. *Plant Biotechnology Journal* published by Society for Experimental Biology and The Association of Applied Biologists and John Wiley & Sons Ltd. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

and ranking sixth in the world in terms of yield (Shahidi and Chandrasekara, 2013). In 2018, the world's total millet production was estimated at 31 million tones, with more than 96% of millet crops grown in regions with poor soil fertility and limited rainfall in Africa and Asia (Muthamilarasan and Prasad, 2021; Yousaf *et al*., 2021). Millets have evolved sophisticated regulatory mechanisms to improve tolerance to various stresses, in adaptation to different environmental conditions and become climate–smart crops (Ceasar and Maharajan, 2022). Consequently, millets can curb food insecurity caused by climate change (Adebiyi *et al*., 2018).

Millets include 11 genera, the more famous are pearl millet (*Pennisetum glaucum* (L.) R. Br., syn. *Cenchrus americanus* (L.) Morrone) and the small millets, finger millet (*Eleusine coracana*), foxtail millet (*Setaria italica*), proso millet (*Panicum miliaceum*), barnyard millet (*Echinochloa crus-galli*), tef (*Eragrostis tef*), fonio (*Digitaria exilis*) and Job's tears (*Coix lacryma-jobi*) (Goron and Raizada, 2015; Muthamilarasan *et al*., 2019; Yousaf *et al*., 2021). Some millets, such as pearl millet, have a close phylogenetic relationship with other major Poaceae crops such as sorghum (*Sorghum bicolor*), maize (*Zea mays*) and rice (*Oryza sativa*), which could enable the easy transfer of its abiotic-stress resistance genes to these crops (Desai *et al*., 2006; Islam *et al*., 2010; Verma *et al*., 2007). For example, a pearl millet glutathione peroxidase (*PgGPX*) gene transformed into rice effectively improved both salt tolerance and drought tolerance (Islam *et al*., 2015). Thus, millets provide new insights into understanding plant abiotic stress tolerance and genetic resources for improving stress tolerance in major crops.

The broad application of high-throughput sequencing technologies has enabled the genomes of millets to be deciphered (Varshney *et al*., 2017) and a large number of multi-omics data from millets have been reported. For instance, intensive research on stress-related genetic resources of millets has produced enormous data covering transcriptome sequencing (Awan *et al*., 2022; Huang *et al*., 2021; Ji *et al*., 2021; Sun *et al*., 2021; Wu *et al*., 2021; Zhang *et al*., 2021), whole-genome resequencing (Varshney *et al*., 2017) and phenomics (Varshney *et al*., 2017). These offer new opportunities to innovate knowledge regarding plant abiotic stress tolerance and to advance the genetic improvement of stress tolerance in major crops. However, the collection and analysis of these data are time-consuming, especially for researchers who lack bioinformatics experience and computing resources, resulting in data mining of key stress-tolerance genes still being a challenging task.

To solve the data mining problem, we have developed a comprehensive and user-friendly millets multi-omics database (Milletdb, http://milletdb.novogene.com), comprising voluminous data with browsing and analytical tools. The Milletdb contains genomes of seven genera and 1800 sets of diverse omics data including graph-based pan-genomics, transcriptomics, epigenomics, variomics and phenomics data. Various practical tools integrated by Milletdb can help users quickly identify a single gene from multiple perspectives (homologous search, blast, traits), build regulatory networks from multiple levels [TE (transposable element) distribution, TF (transcription factor) binding sites, expression level, protein level] and characterize gene sets (sequence characteristics, expression patterns and functional enrichment). The database platform can provide effective services for the entire scientific community in millets' functional genomics and population genetic studies.

Results

Database content

Milletdb contains 824 198 entries of genes from 18 genomes viz 11 pearl millet, two elephant grass (*Cenchrus purpureus*), one foxtail millet (v 2.0) one proso millet (Pm_0390_v2), one finger millet (Ragi_PR202_v._2.0), one fonio (DiExil) and one barnyard millet (ec_v3), with information on 7184 biological pathways (939 pathways related to abiotic stress) and one graph-based pan-genome with information on 30 050 483 SNPs (single nucleotide variants), 424 085 SVs (structural variants), and 692 transcriptome data. Approximately 147 of the transcriptome data are derived from different developmental stages, 527 and 18 are from abiotic and biotic stress, respectively. The database also holds four histones ChIP–seq data (H3K4me3 and H3K36me3 modification of root and panicle), 400 basic information data (such as biological status, seed source, etc.) 242 with phenotypic data, 378 resequencing data, 1 455 924 pearl millet population SNPs and 124 532 SVs among pearl millet accessions (Figure 1a, Table S1). Notably, the pearl millet materials for generating stress-related transcriptome data were uniformly grown and processed. It can avoid errors due to different growth conditions of the materials. The pan-genome browser in Milletdb displays all SVs, SNPs, indels and histone modification sites of pearl millet. Thus, Milletdb contains the largest amount of millet-related data up to now, solving the dilemma associated with collecting and interrogating all of these data.

Gene identification

Reverse-genetics strategy

Milletdb shown as follows provides three options when searching for genes. (A) Gene ID search: where users can obtain the target gene in 'Gene' through gene ID, KO/GO number, Pfam ID, keywords and corresponding gene ID in Ensemble and Phytozome databases (Figure 1a,b); (B) Homology search: whereby users can directly input the gene ID of other species (e.g., *Arabidopsis*, rice and maize) in the 'Homologous gene search' bar to identify candidate genes (Figure 1a,b). The 'Homologous gene search' supports a two-way search, enabling users to identify homologous candidate genes across species that could benefit crop breeding. In addition, the sequence alignment of protein sequences between homologous genes is shown; (C) Search by sequence, which allows users to employ 'Blast' (basic local alignment search tool) to quickly obtain specific genes using sequence information (Figure 1a,b). The hyperlinks provided by 'Blast' helps users obtain full information on target genes.

Forward-genetics strategy

Trait-based search allows users to screen key genes potentially controlling the target trait via the 'Significant SNPs & Gene'/'Significant SVs & Gene' part of the 'GWAS' (genome-wide association study) module under 'Variation' (Figures 1a and 2). For example, we identified an SV associated with the vegetative growth index [GI (kg/ha/d)] through the 'Significant SVs & Gene' page in the 'GWAS' section of the 'Variation' module (Figure 2a, Figure S1A). Navigating to the results section of this web page showed that this SV was located within 4.89 kb of the gene

14677652, 2023, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/pbi.14136 by International Crops Research Institute for Semi Arid Tropics, Wiley Online Library on [21/11/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

PMA1G03779.1, which can be identified from 122 (31%) pearl millet accessions (Figure 2a, Figure S1B). According to the 'Individual Alleles' page, we observed that the GI value of the accessions (122) with this SV was significantly lower than that of the accessions (256) without this SV (Figure 2b, Figure S1B). Further analysis of the potentially associated gene, *PMA1G03779.1*, revealed that it is likely to be involved in the G-protein coupled receptor signalling pathway, which has a potential role in plant growth (Figure S1C; Colucci *et al*., 2002). Utilization of the 'Heatmap' feature on the webpage shows that this gene is not only expressed during the vegetative growth stage of pearl millet but also in the leaves under high temperature, drought and salt stress (Figure 2c, Figure S1C). Through 'Homologous gene search', we found that this gene has no orthologous gene in maize, rice or *Arabidopsis* and it is presumed to be a new gene that controls GI (Figure S1D).

The landscape of gene information

Milletdb provides comprehensive information on millet genes, including their annotation, location, homology and expression (Figure S2), which is helpful for in-depth functional studies of the target. The gene information mainly contains three parts: basic gene information, homologous genes and heat maps. The basic gene information shows the millet accessions to which the gene belongs, the annotation information of the gene in five databases [Swissport (Swiss-Prot Protein Sequence Database), NR (Non-Redundant Protein Sequence Database), KEGG (Kyoto Encyclopedia of Genes and Genomes, https://www.kegg.jp/), GO (Gene Ontology, http://geneontology.org/), Interpro (https://www.ebi.ac.uk/interpro/)], the sequence information of the gene (nucleic acid and protein sequence) and the chromosomal location. Clicking on the location hyperlink opens the pan-genome browser (the pan-genome browser supports genes from the reference PI537069 and the others are supported by the genome browser), which shows the homologous genes and their structures among pearl millet accessions. The browser also displays information on SVs and SNPs adjacent to the genes, enabling users to design markers for exploring the genetic differences between pearl millet accessions (Figure 1b; Figure S3). Meanwhile, the homologous genes show those genes with similar homologous genes among different millets, *Arabidopsis*, rice and maize. The heat map shows the expression of genes in three categories, viz multiple stresses (i.e., heat stress, drought stress and salt stress), multiple tissues and multiple materials. For convenience, users can extract the gene information through the Gene ID hyperlinks in 'Gene', 'Variation' and 'Tools', etc., which prompts users to obtain information on key genes more efficiently and comprehensively.

Identification of upstream regulatory elements of genes

Milletdb contains two tools, the 'Transposable Elements Identification' and 'Motif binding site prediction' for analysing regulatory elements adjacent to the genes (Figure 1b). This facilitates users to find the upstream regulatory elements of the target gene. 'Transposable Elements Identification' can be used to summarize the TE statistics around a gene set, the distribution of TEs and the TE information of each gene. 'Motif binding site prediction' provides a convenient method for users to analyse upstream *cis*-elements of genes by matching multiple expectation maximizations for motif elicitation (MEME) motifs. Users can obtain a list of genes containing the binding sites of the TF after entering the motif of a target TF and selecting the length of the upstream sequence of the genes (Figure 1b; Figure S4). Milletdb also supports directly downloading complete TE information via the 'Transposable Elements' tool. In brief, the Milletdb is helpful for users to understand the potential regulation of gene expression in a simplified way.

Gene network construction

Millets have excellent resistance to abiotic stresses through complex gene regulation networks. Therefore, we developed modules to enable prediction on the gene regulatory networks. All the gene information involved in the regulatory pathway can be quickly retrieved using a keyword or pathway number of the target gene (Figure 1). The main module 'Pathway' contains 422 non-redundant KEGG pathways. The 'Accessions' dropdown menu can be used to select the millet accessions. Moreover, the 'Map_ID' contains a hyperlink for displaying detailed pathway information, while 'Map Info' shows the schematic diagram of the pathway, where the green fill indicates its existence in millet. The genes involved in a specific pathway are displayed in a table format at the bottom of the schematic diagram and can be downloaded in CSV or excel file formats. Furthermore, the 'Co-expression' tool is constructed based on the expression patterns of genes under various stresses (heat, drought, salt), multiple materials and multiple tissues (Figure 1, Figure S4). This paves the way for screening a large number of genes potentially interacting with a target gene. Also, Milletdb provides 'PPI' (protein–protein interaction) analysis for narrowing down gene sets obtained by 'Co-expression'.

Analysis of *PLATZ* genes in millets based on tools provided by Milletdb

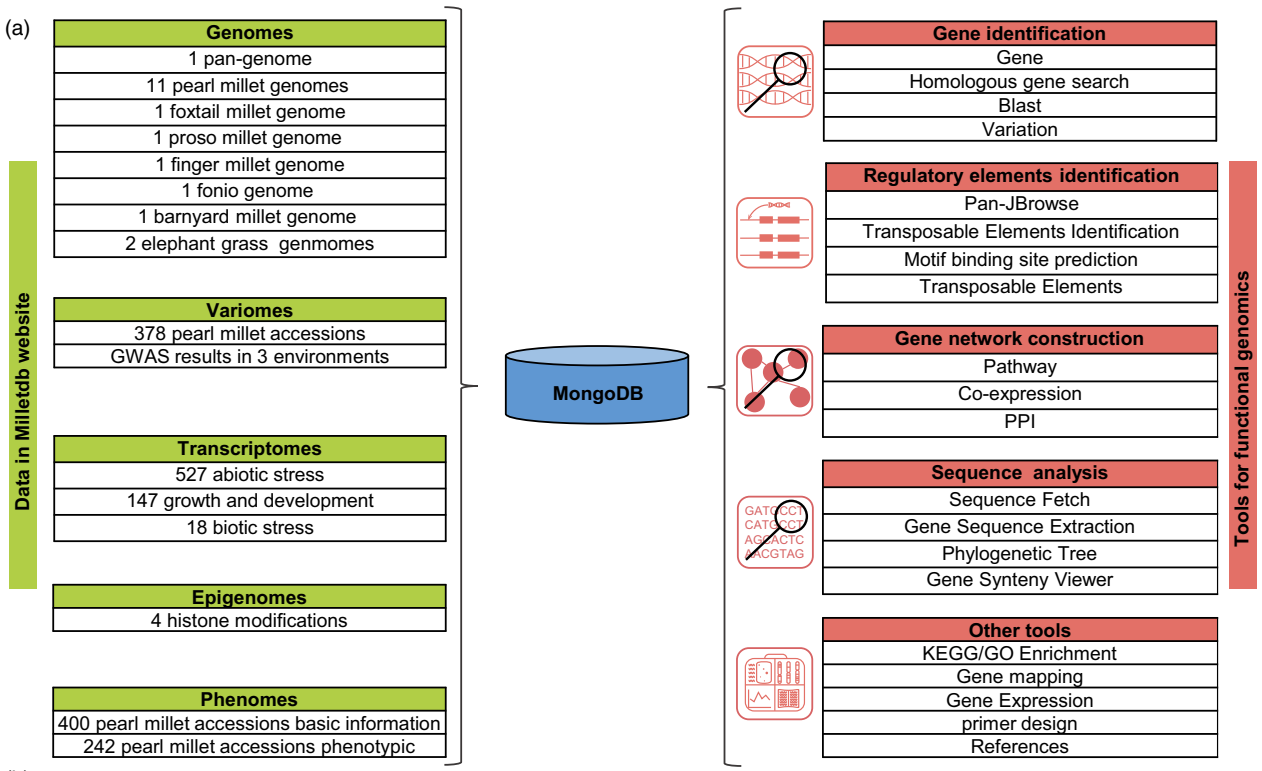
Milletdb provides practical tools for gene sequence analysis and gene function prediction (Figure 1). For example, users can extract sequence fragments, coding sequences (CDSs), protein sequences and upstream or downstream sequences using gene location or ID through 'Sequence Fetch' and 'Gene Sequence Extraction' tools' (Figure 1a). Subsequently, the polymorphisms between gene sequences of different millet accessions can be determined via 'Gene Synteny Viewer' and gene sets clustered according to sequence features using the 'Phylogenetic Tree' tool (Figure 1a, Figure S4). The functional enrichment analysis of the gene clusters is then conducted using the 'KEGG/GO Enrichment' tool, which also supports the online adjustment of pictures (Figure 1). The 'Gene Expression' tool provided by Milletdb allows batch extraction of expression of the genes under different catalogues. The distribution of specific gene sets in millet chromosomes is displayed via the 'Gene mapping' tool (Figure 1). Milletdb also provides the 'References' and 'Primer design' tools for searching millet-related publications and primers, respectively (Figure 1).

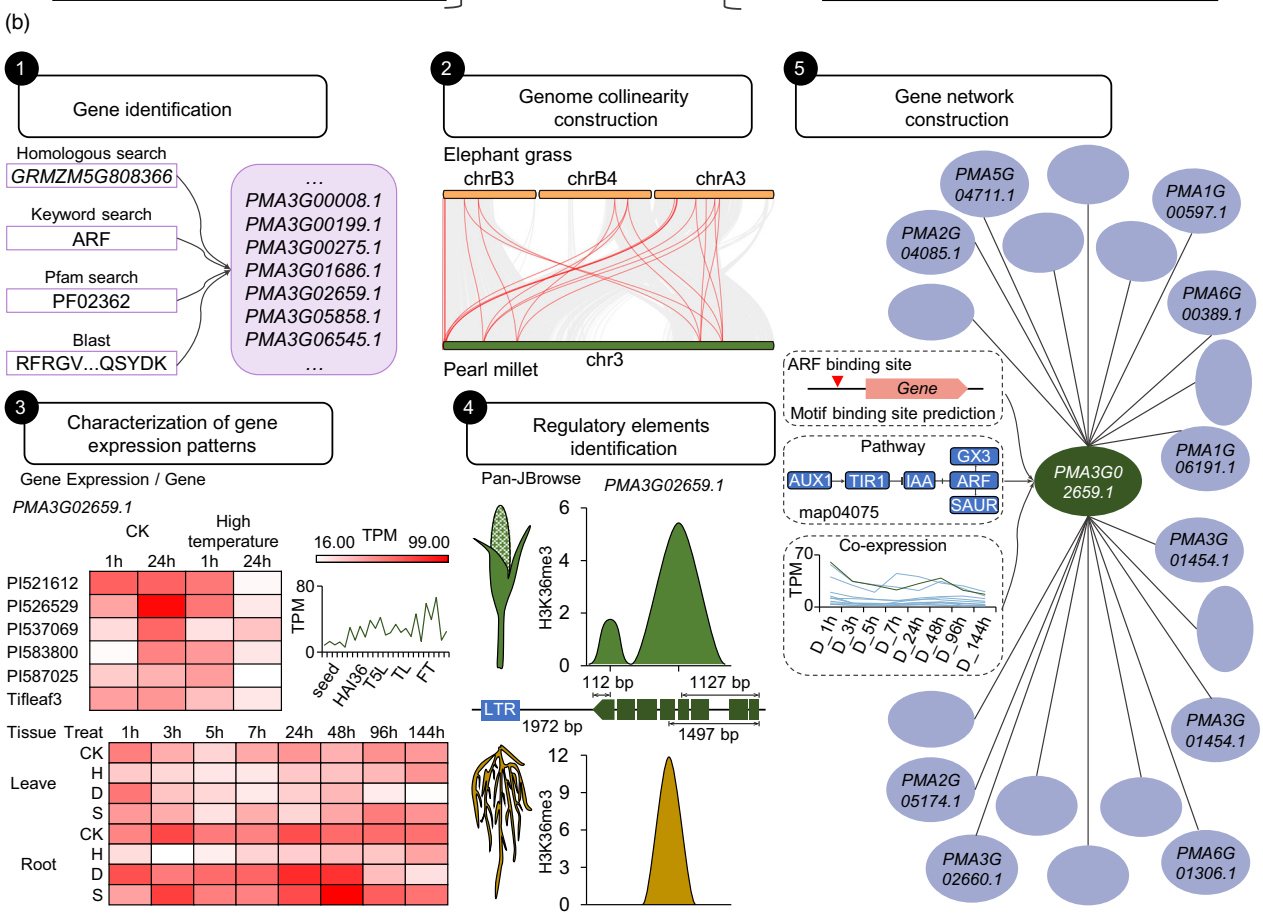
A typical case of a user using the web is shown in Figure 3. Millets are generally cereal crops that are extremely tolerant to environmental stresses (Muthamilarasan and Prasad, 2021). PLATZs (plant AT-rich sequence and zinc-binding proteins) are plant-specific TFs involved in plant responses to abiotic stress (Fu *et al*., 2020; González-Morales *et al*., 2016). However, the PLATZ gene family in millets has not been reported so far. Based on the Milletdb platform, we identified that millets contain 17–59 genes encoding PLATZs depending on the accession. Of the millet accessions, 87.5% (14/17) contain more PLATZs than maize, rice and barley (*Hordeum vulgare*) (Figure 3a, Figure S4A).

© 2023 The Authors. *Plant Biotechnology Journal* published by Society for Experimental Biology and The Association of Applied Biologists and John Wiley & Sons Ltd.,

14677652, 2023, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/pbi.14136 by International Crops Research Institute for Semi Arid Tropics, Wiley Online Library on [21/11/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The 'Gene Sequence Extraction' tools were used to extract the protein sequences of PLATZ genes in millets. According to the analysis results of the 'Phylogenetic Tree' tool, PLATZ genes are divided into four clades. Of these, two clades (clade 2 and clade 3) only contain PLATZ genes from millets (Figure 3b, Figure S4B,C). Further analysis of the protein sequences of the





14677652, 2023, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/pbi.14136 by International Crops Research Institute for Semi Arid Tropics, Wiley Online Library on [21/11/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Figure 1 Overview of Milletdb and its application in millets functional genomics. (a) Milletdb data content and functions. The left panel illustrates the multi-omics data stored in Milletdb and the right panel shows the utilities in Milletdb and their purposes. 'Gene' is used to search for target genes based on gene identifier (ID), KO/GO/Pfam ID and keywords. 'Homologous gene search' is used to search for genes of interest across species. The Basic Local Alignment Search Tool 'Blast' is used to retrieve target genes based on sequence similarity. 'Variation' is used to search for genes of interest based on associated traits. 'Pan-JBrowse' is used to view homologous genes and TEs (transposable elements), SNPs, SVs and nearby genes. 'Transposable Elements Identification' is used to view TEs near the target gene. 'Motif binding site prediction' is used to find genes containing specific motifs. 'Transposable Elements' is used to view and download whole-genome TE information. 'Pathway' is used to search for genes involved in a specific pathway. 'Co-expression' searches for gene sets that are co-expressed with the specified genes. 'PPI' (protein–protein interaction) searches for proteins that interact with the specified protein. 'Sequence Fetch' is used to extract sequence fragments based on their position. 'Gene Sequence Extraction' is used to retrieve coding sequences (CDS), protein sequences or upstream and downstream sequences based on gene IDs. 'Phylogenetic Tree' is used to build a phylogenetic tree of a specific gene set. 'Gene Synteny Viewer' is used to examine collinearity among multiple genes. 'KEGG/GO Enrichment' is used for functional enrichment analysis of specified gene sets. 'Gene mapping' is used to display the chromosomal distribution of the specified gene set. 'Gene expression' is used to extract the expression of a gene set. 'primer design' is used to design primers. 'References' is used to search for references in the literature. (b) Shows the tools in Milletdb used for the complete analysis process of the auxin response factors (ARFs) gene family. Firstly, the homologous gene ID, keywords, Pfam and ARF protein sequence information are used to search for ARF genes on the Milletdb platform (1); According to the genome information provided by Milletdb, genome collinearity analysis is conducted (2); The expression of ARF genes is characterized by 'Gene expression' and 'Gene' in Milletdb (3); Histone modified regions are identified based on 'Pan-JBrowse' (4); Finally, 'Motif binding site prediction', 'Pathway', 'Co-expression', 'Pan-JBrowse', or 'Transposable Elements Identification' are used to build the regulatory network (5). CK, control group; D, drought stress; S, salt stress; H, heat stress; seed, seeds in the ripening stage; HAI36, imbibition after 36 h; T5L, five-leaf stage; TL, tillering stage; FT, flowering stage.

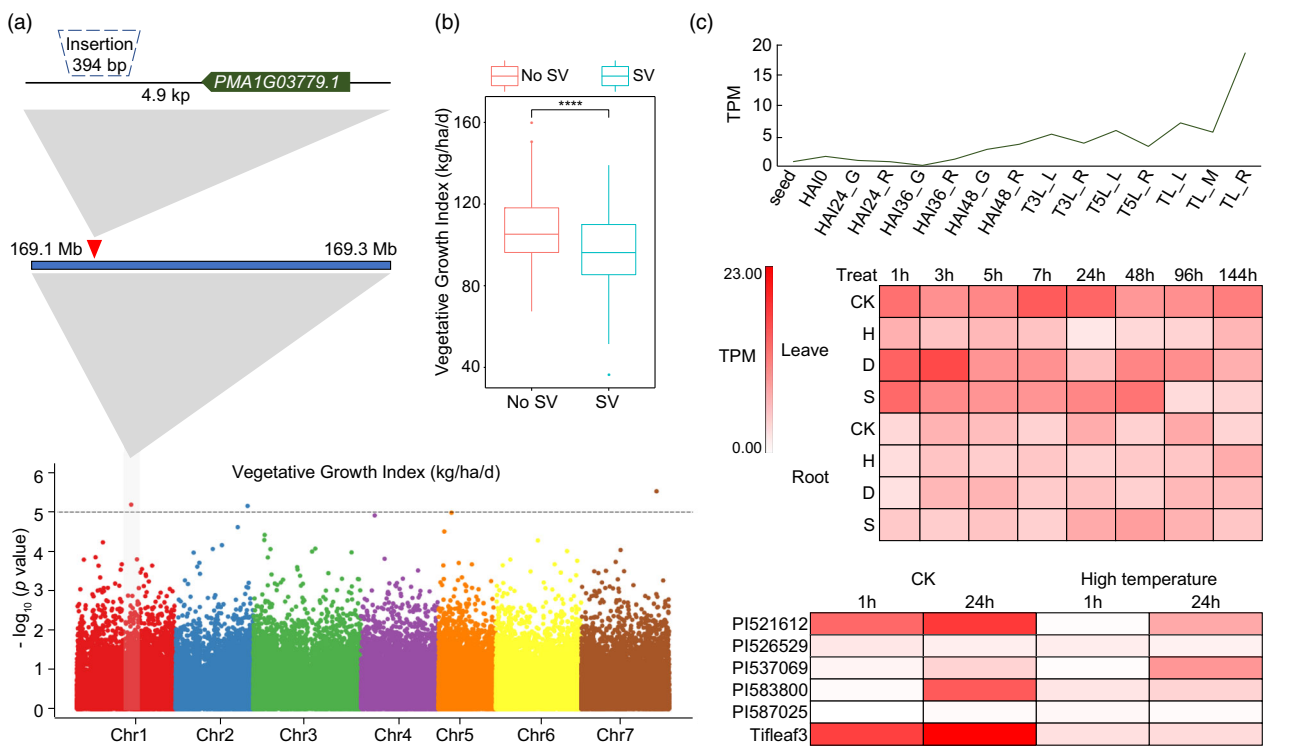
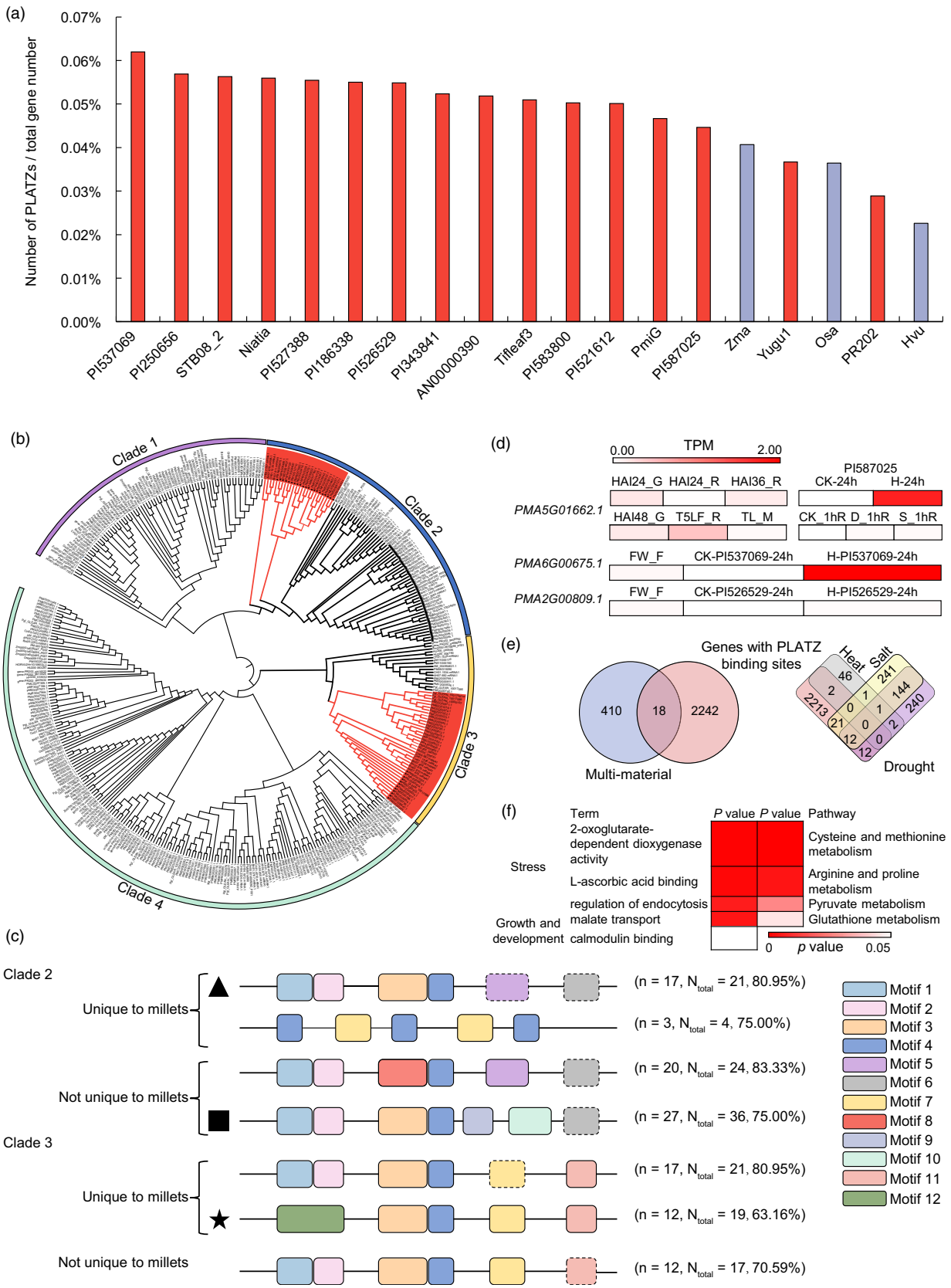


Figure 2 Identification of new genes regulating GI by Milletdb. (a) PAV–GWAS (presence and absence variations genome-wide association study) of SVs associated with the GI trait based on the graph-based pan-genome. (b) Identification of new genes regulating GI by Milletdb. ****indicates *P*-values at significance levels of <0.05. (c) The expression pattern of *PMA1G03779.1* under different stresses (heat, drought and salt stress), different growth stages and different accessions. CK, control group; D, drought stress; S, salt stress; H, heat stress; seed, seeds in ripening stage; HAI0, imbibition; HAI24, imbibition after 24 h; HAI36, imbibition after 36 h; HAI48, imbibition after 48 h; T3L, three leaf stage; T5L, five-leaf stage; TL, tillering stage; G, germ; R, root; L, leaf; M, tillering tissue.

above two clades found that ~~C~~clade 2 contains motifs 1, 2, 3 and 4 and clade 3 contains motifs 12, 3, 4, 7 and 11 structures which are uniquely present in pearl millet. With the 'Gene mapping' tool, we found that these genes are mainly distributed at both ends of the chromosomes (Figure S4D). The 'Gene Expression' tool used to characterize the above genes expression patterns showed that the gene *PMA5G01662.1*, containing motifs 1, 2, 3, 4, 5 and 6, was involved in seed germination, seedling growth and tiller tissue growth of pearl millet and was up-regulated in response to heat, drought and salt stresses (Figure 3d,

14677652, 2023, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/pbi.14136 by International Crops Research Institute for Semi Arid Tropics, Wiley Online Library on [21/11/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Figure S4E). The gene *PMA6G00675.1*, which only contains motifs 3, 4 and 5, plays a role in flowering and high-temperature stress (Figure 3d). The protein sequence of gene *PMA2G00809.1*, which regulates the flowering in pearl millet and participates in the response to high-temperature stress, contains the motifs 12, 3, 4, 7 and 11 (Figure 3d). The above results suggest that motifs



14677652, 2023, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/pbi.14136 by International Crops Research Institute for Semi Arid Tropics, Wiley Online Library on [21/11/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Figure 3 The Milletdb platform is used to identify the PLATZ gene family. (a) The proportion of PLATZ gene family members among all genes in millets, maize, rice and barley. (b) The phylogenetic tree of PLATZ genes. (c) MEME motif structure of PLATZ genes in Clade2 and Clade3. The solid line and dotted line represent the conservative motif and alternative motif, respectively. (d) Expression pattern of genes *PMA6G00675.1*, *PMA2G00809.1* and *PMA5G01662.1*. (e) Venn diagram of gene sets interacting with *PMA5G01662.1*. (f) Functional enrichment of genes interacting with *PMA5G01662.1*. The solid line and dotted line represent the conservative motif and alterable motif, respectively. The black triangle and star represent the pearl millet-specific protein structure. CK, control group; D, drought stress; S, salt stress; H, heat stress; HAI24, imbibition after 24 h; HAI36, imbibition after 36 h; HAI48, imbibition after 48 h; T5L, five-leaf stage; TL, tillering stage; FW, flowering stage; R, root; M, tillering tissue; F, spike.

1, 2 and 6 may be important in plant vegetative growth and multiple abiotic stress responses, although subsequent functional analysis is still needed. Based on 'Co-expression' and 'Motif binding site prediction', 165 genes and 18 genes were identified as potentially interacting with *PMA5G01662.1* at levels of stress (heat, drought and salinity) and growth, respectively (Figure 3e, Figure S4F,G).

The 'KEGG/GO Enrichment' tool showed that 165 genes were enriched in the pathway or terms related to stress, such as the 'Cysteine and methionine metabolism' pathway (Romero *et al.*, 2014), 'Arginine and proline metabolism pathway' (Dar *et al.*, 2016), 'Pyruvate metabolism' pathway (Kato-Noguchi, 2006), '2-oxoglutarate-dependent dioxygenase activity' term (Vigani *et al.*, 2013), 'L-ascorbic acid binding' term (Gallie, 2013), 'regulation of endocytosis' term (Fan *et al.*, 2015) (Figure 3f, Figure S4H). Eighteen genes were enriched in the 'Glutathione metabolism' pathway (May *et al.*, 1998) and 'calmodulin binding' term (Bouché *et al.*, 2005), which are associated with plant growth and development (Figure 3f). In summary, the Milletdb platform contains a large amount of data and practical tools to meet the needs of researchers to quickly mine key genes, identify interacting genes and build regulatory networks using forward or reverse genetics strategies.

Discussion

In summary, Milletdb is the most comprehensive millet database produced so far. It comprises and visualizes a large amount of multi-omics data for users to extend their studies from individual genes to the level of millet genetic networks. Milletdb provides plenty of candidate stress-related genes orthologous to genes of other major crops, which may help breeders easily identify important genes that improve crop yields and positively respond to stress. The interface is user-friendly and bilingual and provides operation manuals (http://milletdb.novogene.com/home; http://milletdb.novogene.com/document) for all tools. Moreover, the database is an open platform where users can extract and share data through the contact information provided on the contact page (http://milletdb.novogene.com/contact). It will be continuously updated to provide long-term support for scientists working on millets and developing stress-tolerant crops.

Materials and methods

Genomics and pan-genome data

We collected whole genome sequence data from eleven pearl millet accessions, including PI186338 (SAMN28616529), PI250656 (SAMN28613898), PI343841 (SAMN28616536), PI521612 (SAMN20372179), PI526529 (SAMN20372180), PI527388 (SAMN28614406), PI537069 (SAMN20372178), PI587025 (SAMN20372182), PI583800 (SAMN20372181), Tifleaf 3 (SAMN20372183), Tifleaf 3 (SAMN20372183) (Yan *et al.*, 2023) and PmiG (Varshney *et al.*, 2017) from NCBI (https://www.ncbi.nlm.nih.gov/assembly). Genome sequence information of other millet accessions and related species, including foxtail millet (Setaria_italica_v2.0) (Bennetzen *et al.*, 2012), proso millet (Pm_0390_v2) (Zou *et al.*, 2019), finger millet (Ragi_PR202_v._2.0) (Hatakeyama *et al.*, 2017), fonio (DiExil) (Wang *et al.*, 2021), barnyard millet (ec_v3) (Wu *et al.*, 2022) and elephant grass (Yan *et al.*, 2021; Zhang *et al.*, 2022), was downloaded from NCBI. The software package EggNOG v5 (http://eggnog5.embl.de/#/app/home) (Huerta-Cepas *et al.*, 2018) was used for functional annotation. TE annotation is done by the software DeepTE (Yan *et al.*, 2020).

The OrthoMCL10 (v2.0.9) (http://orthomcl.org/orthomcl/) and MUMmer (v4.0.0) (Delcher *et al.*, 2003) software packages were used to identify the gene atlas of the eleven genomes (core, dispensable and private genes), generate a genetic variation atlas and construct a graph-based pearl millet pan-genome.

Resequencing data

The 378 whole-genome resequencing data of pearl millet were derived from SRP063925 (Varshney *et al.*, 2017) and mapped to the graph-based pan-genome using vg tools (Garrison *et al.*, 2018). Thereafter, PAV–GWAS (presence and absence variations genome-wide association study) and SNP–GWAS were performed using GEMMA (v0.94.1) (Zhou and Stephens, 2012) and the results were made available in Milletdb.

RNA-Seq data

We collected a total of 192 transcriptome data of accession Tifleaf 3 grown under abiotic stress conditions, including heat (Huang *et al.*, 2021; Sun *et al.*, 2021), drought (Ji *et al.*, 2021; Zhang *et al.*, 2021) and salt (Awan *et al.*, 2022) (Table S1). The 13-day-old pearl millet seedlings were grouped into the normal culture group (CK), heat treatment group (40 °C/35 °C), drought treatment group (20% PEG) and salt treatment group (100 mM/L). The treatments were performed simultaneously, and fresh leaves and roots were collected at 1, 3, 5, 7, 24, 48, 96 and 144 h (h) after the treatments. The raw data was filtered by fastq (Version 0.11.9, Default setting) using the default parameters (Andrews, 2014) and Kallisto (v0.46.2, −b 100) (Bray *et al.*, 2016) was used to assess expression levels. Moreover, we collected transcriptome data of PI537069, PI521612, PI587025, PI583800, PI526529 and Tifleaf 3 from normal culture and high-temperature treatment (45 °C/40 °C).

The transcriptome data for pearl millet seed germination were also obtained from a previous study (Wu *et al.*, 2021) (Table S1). We conducted transcriptome sequencing using different tissues of accession Tifleaf3 under normal culture conditions at different developmental stages: root and leaf samples at the three-leaf and five-leaf stages; roots, stems, leaves and tiller tissues samples at the tillering stage; panicles, leaves and stem samples at the

© 2023 The Authors. *Plant Biotechnology Journal* published by Society for Experimental Biology and The Association of Applied Biologists and John Wiley & Sons Ltd.,

14677652, 2023, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/pbi.14136 by International Crops Research Institute for Semi Arid Tropics, Wiley Online Library on [21/11/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

heading, flowering and dough stage. Fastq (Version 0.11.9) (Andrews, 2014) and Kallisto (Bray *et al.*, 2016) were used to analyse the transcriptome data.

Raw transcriptomic data of other millets was sourced from the SRA database (Bandyopadhyay *et al.*, 2020; Bennetzen *et al.*, 2012; Cannarozzi *et al.*, 2014; Fang *et al.*, 2019; Guo *et al.*, 2017; Hatakeyama *et al.*, 2017; Jin *et al.*, 2021; Lai *et al.*, 2021; Pan *et al.*, 2020; Qin *et al.*, 2020; Ramadoss, 2014; Shen *et al.*, 2020; Sun *et al.*, 2022; Wang *et al.*, 2020, 2021; Watson-Lazowski *et al.*, 2019; Wu *et al.*, 2022; Yan *et al.*, 2021, 2022, 2023; Yu *et al.*, 2020; Yuan *et al.*, 2021, 2022; Zhang *et al.*, 2022; Zou *et al.*, 2019) (Table S1). Fastq (Version 0.11.9) was used for raw data filtering. Kallisto was used to align with the reference genome (AN00000390, Yugu1 and PR202) and to calculate the count value.

Co-expression analysis

We grouped the transcriptome data into five catalogues: heat stress, drought stress, salt stress, multi-tissue (growth and development) and multi-accession (PI537069, PI521612, PI587025, PI583800, PI526529 and Tifleaf3). After that, the Pearson correlation coefficient was used to analyse the correlation between gene expression within each catalogue based on the transcript per million (TPM) value of the genes using Hmisc (Jr and Dupont, 2015) software. We obtained the correlation index and *P*-values between genes within each class and saved the correlation results in Milletdb.

ChIP–seq data

Spikelets and mature roots of pearl millet grown under normal field conditions were collected separately and treated according to the method described previously (Wedel and Siegel, 2017). Briefly, the samples were homogenized in liquid nitrogen and digested with micrococcal nuclease (MNase) for 8 min to achieve chromatin shearing. Following this, anti-H3K4me3 (Millipore, 05-745R) and anti-H3K36me3 (Abcam, ab9050) were added to the immunoprecipitated sheared chromatin, whose fragments were then captured using protein A/G magnetic beads (Thermo, cat 88 802). All libraries were prepared with the VAHTS universal DNA library prep kit for Illumina Library Systems (Vazyme) according to the manufacturer's instructions. The obtained library was then sequenced using Illumina Hiseq-Xten and Fastp software (Chen *et al.*, 2018) was used to filter reads. Alignment was performed using Bowtie 2.4.5 (Langmead and Salzberg, 2012) with PI537069 as the reference genome sequence. The software MACS 3.0.0a7 (Feng *et al.*, 2012) was used to call peaks (Table S1).

Pearl millet accession details

We collected basic information on 400 pearl millet accessions (Varshney *et al.*, 2017), of which 242 materials contained phenotypic data while growing under three conditions (field, early drought stress and late drought stress) (Varshney *et al.*, 2017). The obtained information was then deposited in Milletdb.

References

We searched the Web of Science website (http://webofknowledge.com) for pearl millet-related articles and exported the search results in batches. The Citavi software (Com, 2006) was used to generate hyperlinks to the articles.

Data integration

The genomic, transcriptomic, epigenomics, phenotypic, co-expression analysis, SNP–GWAS and PAV–GWAS data of millets were stored in MongoDB.

Database construction

Milletdb is implemented by the Linux-operating system and Nginx. Ant Design and Django were used for interactive front-end and back-end queries. Genomic and pan-genomic features were displayed by JBrowse (Buels *et al.*, 2016) and its plugins. The variation in the pearl millet population is displayed via Ant Design Charts (https://charts.ant.design/), igv (Thorvaldsdottir *et al.*, 2012) and ant-design/icons (https://ant.design/components/icon-cn/). BLAST 2.2.31+ (Tatusova and Madden, 1999) was used for 'Blast' construction. SYNVISIO (https://github.com/kiranbandi/synvisio) was used to visualize results based on inter-genome alignments (blastp). The 'Sequence Fetch', 'Gene Sequence Extraction', 'Transposable Elements Identification', 'Transcription Factor identification' and 'Co-expression' tools were built using Python 3.9.

Acknowledgements

This work was supported by the Modern Agricultural Industry System Sichuan Forage Innovation Team (SCCXTD-2021-16), the earmarked fund for CARS (CARS-34), the Sichuan Province Research Grant (2021YFYZ0013) and the National Natural Science Foundation of China (Nos. 31771866 and 32071867). We thank John Pablo Mendieta (Department of Genetics, University of Georgia, Athens, GA, 30602, USA) for providing valuable suggestions for the language improvement of the manuscript.

Author contributions

L.H., H.Y. and M.S. designed and managed the project. M.S., A.Z., Y.J., C.L., L.L., R.K.S., R.K.V. and B.W. participated in material collecting and processing. S.T., M.S, Y.F., H.Y. and X.C. performed bioinformatics analyses. L.H., H.Y. and M.S. wrote the manuscript. F.Z., L.H., X.Z., X.W., D.H., G.N., G.F., Z.X., M.H., R.K.S., R.K.V., C.S.J., Q.T., P.Z., J.J., Y.Y., Z.W. and J.L. revised the article.

Competing interests

The authors declare no competing interests.

References

Adebiyi, J.A., Obadina, A.O., Adebo, O.A. and Kayitesi, E. (2018) Fermented and malted millet products in Africa: Expedition from traditional/ethnic foods to industrial value-added products. *Crit. Rev. Food Sci. Nutr.* **58**, 463–474.

Andrews, S. (2014) *FastQC A quality control tool for high throughput sequence data*.

Awan, S.A., Khan, I., Tariq, R., Rizwan, M., Wang, X., Zhang, X. and Huang, L. (2022) Genome-wide expression and physiological profiling of pearl millet genotype reveal the biological pathways and various gene clusters underlying salt resistance. *Front. Plant Sci.* **13**, 849618.

Bandyopadhyay, T., Swarbreck, S.M., Jaiswal, V., Gupta, R., Bentley, A.R., Griffiths, H. and Prasad, M. (2020) *Grain number and genotype drive nitrogen-dependent yield response in the C4 model Setaria italica* (L.) P. Beauv. https://doi.org/10.1101/2020.03.23.003004

14677652, 2023, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/pbi.14136 by International Crops Research Institute for Semi Arid Tropics, Wiley Online Library on [21/11/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Bennetzen, J.L., Schmutz, J., Wang, H., Percifield, R., Hawkins, J., Pontaroli, A.C., Estep, M. *et al.* (2012) Reference genome sequence of the model plant Setaria. *Nature Biotechnology* **30**(6), 555–561.

Bouché, N., Yellin, A., Snedden, W.A. and Fromm, H. (2005) Plant-specific calmodulin-binding proteins. *Annu. Rev. Plant Biol.* **56**, 435–466.

Bray, N.L., Pimentel, H., Melsted, P. and Pachter, L. (2016) Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527.

Buels, R., Yao, E., Diesh, C.M., Hayes, R.D., Munoz-Torres, M., Helt, G., Goodstein, D.M. *et al.* (2016) JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.* **17**, 66.

Cannarozzi, G., Plaza-Wüthrich, S., Esfeld, K., Larti, S., Wilson, Y.S., Girma, D., de Castro, E. *et al.* (2014) Genome and transcriptome sequencing identifies breeding targets in the orphan crop tef (Eragrostis tef). *BMC Genomics* **15**(1). https://doi.org/10.1186/1471-2164-15-581

Ceasar, S.A. and Maharajan, T. (2022) The role of millets in attaining United Nation's sustainable developmental goals. *Plants People Planet*, **4**, 345–349.

Chen, S., Zhou, Y., Chen, Y. and Gu, J. (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics (Oxford, England)*, **34**, i884–i890.

Colucci, G., Apone, F., Alyeshmerni, N., Chalmers, D. and Chrispeels, M.J. (2002) GCR1, the putative Arabidopsis G protein-coupled receptor gene is cell cycle-regulated, and its overexpression abolishes seed dormancy and shortens time to flowering. *Proc. Natl. Acad. Sci. USA*, **99**, 4736–4741.

Com, I. (2006) *Citavi – literaturverwaltung und wissensorganisation*.

Dar, M.I., Naikoo, M.I., Rehman, F., Naushin, F. and Khan, F.A. (2016) Proline accumulation in plants: roles in stress tolerance and plant development. In *Osmolytes and Plants Acclimation to Changing Environment: Emerging Omics Technologies*(Iqbal, N., Nazar, R. and Khan, N.A., eds), pp. 155–166. New Delhi: Springer India.

Delcher, A.L., Salzberg, S.L. and Phillippy, A.M. (2003) Using MUMmer to identify similar regions in large sequence sets. *Curr. Protoc. Bioinform.* (29). https://doi.org/10.1002/0471250953.bi1003s00

Desai, M.K., Mishra, R.N., Verma, D., Nair, S., Sopory, S.K. and Reddy, M.K. (2006) Structural and functional analysis of a salt stress inducible gene encoding voltage dependent anion channel (VDAC) from pearl millet (*Pennisetum glaucum*). *Plant Physiol. Biochem.* **44**, 483–493.

Fan, L., Li, R., Pan, J., Ding, Z. and Lin, J. (2015) Endocytosis and its regulation in plants. *Trends Plant Sci.* **20**, 388–397.

Fang, J., Zhang, Y., Liu, T., Yan, B., Li, J. and Dong, L. (2019) Target-Site and Metabolic Resistance Mechanisms to Penoxsulam in Barnyardgrass (Echinochloa crus-galli (L.) P. Beauv). *Journal of Agricultural and Food Chemistry* **67**(29), 8085–8095.

FAO, F. (2018) *Food and agriculture organization of the United Nations*. Rome. http://faostat.fao.org.

Feng, J., Liu, T., Qin, B., Zhang, Y. and Liu, X.S. (2012) Identifying ChIP-seq enrichment using MACS. *Nat. Protoc.* **7**, 1728–1740.

Fu, Y., Cheng, M., Li, M., Guo, X., Wu, Y. and Wang, J. (2020) Identification and characterization of PLATZ transcription factors in wheat. *Int. J. Mol. Sci.* **21**, 8934.

Gallie, D.R. (2013) The role of L-ascorbic acid recycling in responding to environmental stress and in promoting plant growth. *J. Exp. Bot.* **64**, 433–443.

Garrison, E., Sirén, J., Novak, A.M., Hickey, G., Eizenga, J.M., Dawson, E.T., Jones, W. *et al.* (2018) Variation graph toolkit improves read mapping by representing genetic variation in the reference. *Nat. Biotechnol.* **36**, 875–879.

González-Morales, S.I., Chávez-Montes, R.A., Hayano-Kanashiro, C., Alejo-Jacuinde, G., Rico-Cambron, T.Y., de Folter, S. and Herrera-Estrella, L. (2016) Regulatory network analysis reveals novel regulators of seed desiccation tolerance in Arabidopsis thaliana. *Proc. Natl. Acad. Sci. USA*, **113**, E5232–E5241.

Goron, T.L. and Raizada, M.N. (2015) Genetic diversity and genomic resources available for the small millet crops to accelerate a New Green Revolution. *Front. Plant Sci.* **6**, 157.

Guo, L., Qiu, J., Ye, C., Jin, G., Mao, L., Zhang, H., Yang, X. *et al.* (2017) Echinochloa crus-galli genome analysis provides insight into its adaptation and invasiveness as a weed. *Nature Communications* **8**(1). https://doi.org/10.1038/s41467-017-01067-5

Hatakeyama, M., Aluri, S., Balachadran, M.T., Sivarajan, S.R., Patrignani, A., Grüter, S., Poveda, L. *et al.* (2017) Multiple hybrid de novo genome assembly of finger millet, an orphan allotetraploid crop. *DNA Research* **25**(1), 39–47.

Huang, D., Sun, M., Zhang, A., Chen, J., Zhang, J., Lin, C., Zhang, H. *et al.* (2021) Transcriptional changes in pearl millet leaves under heat stress. *Genes*, **12**, 1716.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R. *et al.* (2018) eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research* **47**(D1), D309–D314.

Islam, S.M.T., Tammi, R.S., Singla-Pareek, S.L. and Seraj, Z.I. (2010) Enhanced salinity tolerance and improved yield properties in Bangladeshi rice Binnatoa through Agrobacterium-mediated transformation of PgNHX1 from *Pennisetum glaucum*. *Acta Physiol. Plant*, **32**, 657–663.

Islam, T., Manna, M. and Reddy, M.K. (2015) Glutathione peroxidase of *Pennisetum glaucum* (PgGPx) is a functional Cd2+ dependent peroxiredoxin that enhances tolerance against salinity and drought stress. *PloS One*, **10**, e0143344.

Ji, Y., Lu, X., Zhang, H., Luo, D., Zhang, A., Sun, M., Wu, Q. *et al.* (2021) Transcriptome reveals the dynamic response mechanism of pearl millet roots under drought stress. *Genes*, **12**, 1988.

Jin, F., Liu, J., Wu, E., Yang, P., Gao, J., Gao, X. and Feng, B. (2021) Leaf transcriptome analysis of broomcorn millet uncovers key genes and pathways in response to *Sporisorium destruens*. *International Journal of Molecular Sciences* **22**(17), 9542.

Jr, F.H. and Dupont, C. (2015) *Hmisc: Harrell Miscellaneous*. R package version 3.17-1. Accessed on, 12.

Juma, S.G. and Kelonye, F. (2016) Projected rainfall and temperature changes over Bungoma county in western Kenya by the year 2050 based precis modeling system. *Ethiop. J. Env. Stud. Manag.* **9**, 625.

Kato-Noguchi, H. (2006) Pyruvate metabolism in rice coleoptiles under anaerobiosis. *Plant Growth Regul.* **50**, 41–46.

Lai, X., Bendix, C., Zhang, Y., Schnable, J.C. and Harmon, F.G. (2021) 72-h diurnal RNA-seq analysis of fully expanded third leaves from maize, sorghum, and foxtail millet at 3-h resolution. *BMC Research Notes* **14**(1). https://doi.org/10.1186/s13104-020-05431-5

Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.

May, M.J., Vernoux, T., Leaver, C., Montagu, M.V. and Inze, D. (1998) Glutathione homeostasis in plants: implications for environmental sensing and plant development. *J. Exp. Bot.* **49**, 649–667.

Molotoks, A., Smith, P. and Dawson, T.P. (2021) Impacts of land use, population, and climate change on global food security. *Food Energy Secur.* **10**, 153.

Muthamilarasan, M. and Prasad, M. (2021) Small millets for enduring food security amidst pandemics. *Trends Plant Sci.* **26**, 33–40.

Muthamilarasan, M., Singh, N.K. and Prasad, M. (2019) Multi-omics approaches for strategic improvement of stress tolerance in underutilized crop species: a climate change perspective. *Adv. Genet.* **103**, 1–38.

Pan, J., Li, Z., Dai, S., Ding, H., Wang, Q., Li, X., Ding, G. *et al.* (2020) Integrative analyses of transcriptomics and metabolomics upon seed germination of foxtail millet in response to salinity. *Scientific Reports* **10**(1). https://doi.org/10.1038/s41598-020-70520-1

Qin, L., Chen, E., Li, F., Yu, X., Liu, Z., Yang, Y., Wang, R. *et al.* (2020) Genome-wide gene expression profiles analysis reveal novel insights into drought stress in foxtail millet (*Setaria italica* L.). *International Journal of Molecular Sciences* **21**(22), 8520.

Ramadoss, A. (2014). *Effects of drought on Eleusine coracana* (L.) Gaertn. (finger millet) and identification of microsatellite markers. University of Alberta Libraries. https://doi.org/10.7939/R3HX16015

Romero, L.C., Aroca, M.Á., Laureano-Marín, A.M., Moreno, I., García, I. and Gotor, C. (2014) Cysteine and cysteine-related signaling pathways in *Arabidopsis thaliana*. *Mol. Plant*, **7**, 264–276.

Shahidi, F. and Chandrasekara, A. (2013) Millet grain phenolics and their role in disease risk reduction and health promotion: a review. *J. Funct. Foods*, **5**, 570–581.

Shen, J., Guo, M., Wang, Y., Yuan, X., Dong, S., Song, X. and Guo, P. (2020) An investigation into the beneficial effects and molecular mechanisms of humic acid on foxtail millet under drought conditions. *PLoS One* **15**(6), e0234029.

Sun, M., Lin, C., Zhang, A., Wang, X., Yan, H., Khan, I., Wu, B. *et al.* (2021) Transcriptome sequencing revealed the molecular mechanism of response of pearl millet root to heat stress. *J. Agro. Crop. Sci.* **207**, 768–773.

14677652, 2023, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/pbi.14136 by International Crops Research Institute for Semi Arid Tropics, Wiley Online Library on [21/11/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Sun, L., Liu, L., Wang, Y., Feng, Y., Yang, W., Wang, D., Gao, S. *et al*. (2022) Integration of Metabolomics and Transcriptomics for Investigating the Tolerance of Foxtail Millet (Setaria italica) to Atrazine Stress. *Frontiers in Plant Science* **13**. https://doi.org/10.3389/fpls.2022.890550

Tatusova, T.A. and Madden, T.L. (1999) BLAST 2 sequences, a new tool for comparing protein and nucleotide sequences. *Fems Microbiol. Lett.* **174**, 247–250.

Thorvaldsdottir, H., Robinson, J.T. and Mesirov, J.P. (2012) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics* **14**(2), 178–192.

Varshney, R.K., Shi, C., Thudi, M., Mariac, C., Wallace, J., Qi, P., Zhang, H. *et al*. (2017) Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. *Nat. Biotechnol.* **35**, 969–976.

Varshney, R.K., Bohra, A., Roorkiwal, M., Barmukh, R., Cowling, W.A., Chitikineni, A., Lam, H.M. *et al*. (2021a) Fast-forward breeding for a food-secure world. *Trends Genet.* **37**, 1124–1136.

Varshney, R.K., Bohra, A., Yu, J., Graner, A., Zhang, Q. and Sorrells, M.E. (2021b) Designing future crops: genomics-assisted breeding comes of age. *Trends Plant Sci.* **26**, 631–649.

Verma, D., Singla-Pareek, S.L., Rajagopal, D., Reddy, M.K. and Sopory, S.K. (2007) Functional validation of a novel isoform of Na+/H+ antiporter from *Pennisetum glaucum* for enhancing salinity tolerance in rice. *J. Biosci.* **32**, 621–628.

Vigani, G., Morandini, P. and Murgia, I. (2013) Searching iron sensors in plants by exploring the link among 2'-OG-dependent dioxygenases, the iron deficiency response and metabolic adjustments occurring under iron deficiency. *Front. Plant Sci.* **4**, 169.

Wang, X., Chen, S., Ma, X., Yssel, A.E.J., Chaluvadi, S.R., Johnson, M.S., Gangashetty, P. *et al*. (2021) Genome sequence and genetic diversity analysis of an under-domesticated orphan crop, white fonio (Digitaria exilis). *GigaScience* **10**(3). https://doi.org/10.1093/gigascience/giab01

Wang, T., Song, H., Li, P., Wei, Y., Hu, N., Chen, Z., Wang, W. *et al*. (2020) Transcriptome analysis provides insights into grain filling in foxtail millet (*Setaria italica* L.). *International Journal of Molecular Sciences* **21**(14), 5031.

Watson-Lazowski, A., Papanicolaou, A., Koller, F. and Ghannoum, O. (2019) The transcriptomic responses of C_4 grasses to subambient CO_2 and low light are largely species specific and only refined by photosynthetic subtype. *The Plant Journal* **101**(5), 1170–1184. Portico.

Wedel, C. and Siegel, T.N. (2017) Genome-wide analysis of chromatin structures in *Trypanosoma brucei* using high-resolution MNase-ChIP-seq. *Exp. Parasitol.* **180**, 2–12.

Wu, B., Sun, M., Zhang, H., Yang, D., Lin, C., Khan, I., Wang, X. *et al*. (2021) Transcriptome analysis revealed the regulation of gibberellin and the establishment of photosynthetic system promote rapid seed germination and early growth of seedling in pearl millet. *Biotechnol. Biofuels*, **14**, 94.

Wu, D., Shen, E., Jiang, B., Feng, Y., Tang, W., Lao, S., Jia, L. *et al*. (2022) Genomic insights into the evolution of Echinochloa species as weed and orphan crop. *Nat. Commun.* **13**, 689.

Yan, H., Bombarely, A. and Li, S. (2020) DeepTE: a computational method for de novo classification of transposons with convolutional neural network. *Bioinformatics (Oxford, England)*, **36**, 4269–4275.

Yan, Q., Li, J., Lu, L., Yi, X., Yao, N., Lai, Z. and Zhang, J. (2022) Comparative transcriptome study of the elongating internode in elephant grass (Cenchrus purpureus) seedlings in response to exogenous gibberellin applications. *Industrial Crops and Products* **178**, 114653.

Yan, Q., Wu, F., Xu, P., Sun, Z., Li, J., Gao, L., Lu, L. *et al*. (2021) The elephant grass (*Cenchrus purpureus*) genome provides insights into anthocyanidin accumulation and fast growth. *Mol. Ecol. Resour.* **21**, 526–542.

Yan, H., Sun, M., Zhang, Z., Jin, Y., Zhang, A., Lin, C., Wu, B. *et al*. (2023) Pangenomic analysis identifies structural variation associated with heat tolerance in pearl millet. *Nat. Genet.* **55**, 507–518.

Yousaf, L., Hou, D., Liaqat, H. and Shen, Q. (2021) Millet: a review of its nutritional and functional changes during processing. *Food Res. Int. (Ottawa, Ont.)* **142**, 110197.

Yu, A., Zhao, J., Wang, Z., Cheng, K., Zhang, P., Tian, G., Liu, X. *et al*. (2020) Transcriptome and metabolite analysis reveal the drought tolerance of foxtail millet significantly correlated with phenylpropanoids-related pathways during germination process under PEG stress. *BMC Plant Biology* **20**(1). https://doi.org/10.1186/s12870-020-02483-4

Yuan, Y., Liu, C., Gao, Y., Ma, Q., Yang, Q. and Feng, B. (2021) Proso millet (*Panicum miliaceum* L.): A potential crop to meet demand scenario for sustainable saline agriculture. *Journal of Environmental Management* **296**, 113216.

Yuan, Y., Liu, L., Gao, Y., Yang, Q., Dong, K., Liu, T. and Feng, B. (2022) Comparative analysis of drought-responsive physiological and transcriptome in broomcorn millet (*Panicum miliaceum* L.) genotypes with contrasting drought tolerance. *Industrial Crops and Products* **177**, 114498.

Zhang, A., Ji, Y., Sun, M., Lin, C., Zhou, P., Ren, J., Luo, D. *et al*. (2021) Research on the drought tolerance mechanism of *Pennisetum glaucum* (L.) in the root during the seedling stage. *BMC Genom.* **22**, 568.

Zhang, S., Xia, Z., Li, C., Wang, X., Lu, X., Zhang, W., Ma, H. *et al*. (2022) Chromosome-scale genome assembly provides insights into speciation of allotetraploid and massive biomass accumulation of elephant grass (*Pennisetum purpureum* Schum.). *Mol. Ecol. Resour.* **22**, 2363–2378.

Zhou, X. and Stephens, M. (2012) Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44**, 821–824.

Zou, C., Li, L., Miki, D., Li, D., Tang, Q., Xiao, L., Rajput, S. *et al*. (2019) The genome of broomcorn millet. *Nat. Commun.* **10**, 436.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1 Data used in Milletdb.
Material S1 Database usage example.
Figure S1 An example of mining new genes potentially regulating GI.
Figure S2 Gene information in Milletdb.
Figure S3 The JBrowse window in Milletdb.
Figure S4 An example of analysing the PLATZ TF family in Milletdb.

14677652, 2023, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/pbi.14136 by International Crops Research Institute for Semi Arid Tropics, Wiley Online Library on [21/11/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License