

RESEARCH

Open Access



MDSi: Multi-omics Database for *Setaria italica*

Xukai Li^{1,2†}, Siyu Hou^{1,3†}, Mengmeng Feng², Rui Xia⁴, Jiawei Li⁴, Sha Tang⁵, Yuanhuai Han^{1,3}, Jianhua Gao^{1,2*} and Xingchun Wang^{1,2*}

Abstract

Background Foxtail millet (*Setaria italica*) harbors the small diploid genome (~450 Mb) and shows the high inbreeding rate and close relationship to several major foods, feed, fuel and bioenergy grasses. Previously, we created a mini foxtail millet, *xiaomi*, with an Arabidopsis-like life cycle. The de novo assembled genome data with high-quality and an efficient *Agrobacterium*-mediated genetic transformation system made *xiaomi* an ideal C₄ model system. The mini foxtail millet has been widely shared in the research community and as a result there is a growing need for a user-friendly portal and intuitive interface to perform exploratory analysis of the data.

Results Here, we built a Multi-omics Database for *Setaria italica* (MDSi, <http://sky.sxau.edu.cn/MDSi.htm>), that contains *xiaomi* genome of 161,844 annotations, 34,436 protein-coding genes and their expression information in 29 different tissues of *xiaomi* (6) and JG21 (23) samples that can be showed as an Electronic Fluorescent Pictograph (xEPF) in-situ. Moreover, the whole-genome resequencing (WGS) data of 398 germplasm, including 360 foxtail millets and 38 green foxtails and the corresponding metabolic data were available in MDSi. The SNPs and Indels of these germplasm were called in advance and can be searched and compared in an interactive manner. Common tools including BLAST, GBrowse, JBrowse, map viewer, and data downloads were implemented in MDSi.

Conclusion The MDSi constructed in this study integrated and visualized data from three levels of genomics, transcriptomics and metabolomics, and also provides information on the variation of hundreds of germplasm resources that can satisfies the mainstream requirements and supports the corresponding research community.

Keywords Database, Multi-omics, Foxtail millet, *Setaria italica*, Tools

Introduction

Foxtail millet (*Setaria italica*) ($2n=2x=18$) is thought to have been domesticated from the wild species green foxtail (*Setaria viridis*) more than 11,000 years ago in northern China [1]. With excellent drought tolerance and water-use efficiency, foxtail millet is still widely cultivated in the arid and semiarid regions in the world, particularly in China and India [2]. Foxtail millet are rich in essential amino acids, fatty acids and minerals, and easily digestible without allergenic potential [3]. It is being developed as a novel model system due to the advantages over other C₄ plants, including small diploid genome (~450 Mb), high inbreeding rate and close relationship to several major foods, feed, fuel and bioenergy grasses [4–7].

*Correspondence:

Jianhua Gao

jhgao@sxau.edu.cn

Xingchun Wang

wxingchun@sxau.edu.cn

¹ Hou Ji Laboratory in Shanxi Province, Shanxi Agricultural University, Taiyuan, Shanxi 030031, China

² College of Life Sciences, Shanxi Agricultural University, Taigu, Shanxi 030801, China

³ College of Agriculture, Shanxi Agricultural University, Taigu, Shanxi 030801, China

⁴ South China Agricultural University, Guangzhou, Guangdong 510640, China

⁵ Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Due to the rapid progress of high-throughput techniques (including genotyping by the array, RNA sequencing, metabolite identification, and phenotypic measurement), there has been rapid growth in related foxtail millet multidimensional omics data, and obtaining a high-quality genome assembly is no longer a bottleneck for most species. Even some species have multiple sets of genomes, such as rice (RAP-DB, <https://rapdb.dna.affrc.go.jp>; RGAP7, <http://rice.uga.edu>; RPAN, <https://cgm.sjtu.edu.cn/3kricedb>; IC4R, <http://ic4r.org>; RiceVarMap v2.0, <http://ricevarmap.ncpgr.cn>; RGI, <https://riceome.hzau.edu.cn>), and maize (MaizeGDB, <https://maizegdb.org>; MGR, <http://maize.uga.edu>; ZEAMAP, <http://www.zeamap.com>; MaizeMine, <https://maizemine.rnet.missouri.edu>; Gramene, <https://www.gramene.org>). A model crop system requires support of sufficient multi-omics data, such as genomics, RNA-seq and metabolic data. At present, the genome of four foxtail millet elite varieties or mutant, including Yugu1 [8], Zhanggu [9], TT8 [10], *xiaomi* [11] and Huagu11 [12] were assembled individually. The genome data of Yugu1 can be accessed in Phytozome [13], the data of Zhanggu was uploaded in BGI Foxtail millet Database (<http://foxtailmillet.genomics.org.cn>) whereas the information of Huagu11 was stored in China National GeneBank DataBase (CNCBdb) (<https://db.cngb.org/search/project/CNP0000993/>). Previously, we assembled the genome of the *xiaomi*, a mutant with a short life cycle like *Arabidopsis thaliana* of elite variety Jingu21 (JG21), to the scale of full chromosomes using single-molecule real-time sequencing (SMRT) and high-throughput chromosome conformation capture (Hi-C) [11]. We have also generated other types of omics data including the transcriptomic profiles of different tissues of *xiaomi* and its mutagenic parent JG21, the gene co-expression networks, as well as the single-nucleotide polymorphisms (SNPs) and insertion-deletion polymorphisms (InDels) of 398 foxtail millet and green foxtail germplasms and so on. Recently, the metabolomes of these germplasms were detected and used to multi-omics analysis [14]. An *Arabidopsis*-like life of *xiaomi* along with the plenty of multi-omics data greatly facilitate the application of *xiaomi*, as well as the foxtail millet in the C_4 model system. A transversion from G to T occurred in the second exon of the *PHYC* (*Si9g09200*) gene, leading to the creation of a stop codon. It is the causal factor behind the *xiaomi* mini phenotype [11]. At present, *xiaomi* has been directly shared more than 40 research groups from 32 departments.

Phytozome [13], (yugu1: https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Sitalica) and BGI Foxtail millet Database (zhanggu: <http://foxtailmillet.genomics.org.cn>) hold sections for the foxtail millet, in which the genome sequence of the yugu1 and zhanggu

are retrievable. In addition, their genes ID are totally different, e.g. Seita.xGxxxxxx for yugu1 and Millet_GLEAN_xxxxxxx for zhanggu, which greatly restrict the application of genome data. The lack of essential information in molecular breeding, such as phenotypic information, molecular markers, and gene expression data, impedes progress in the field. To overcome this challenge, there is an urgent need for a multi-dimensional database that encompasses multiple levels of omics data for the same panel. This comprehensive database should comprise high-quality data on genomic variations, reliable quantitation of gene expression, detailed measurement of cellular metabolites, and high-throughput characterization of morphological phenotypes for large, representative, and diverse populations. Furthermore, mapping results based on these data and generated using different methods will directly contribute to further scientific studies. The resource should also be equipped with a user-friendly mechanism for data retrieval, convenient visualization tools, and comprehensive annotation information.

Advances in high-throughput technologies have given rise to huge amounts of omics data for many species. However, wet-lab biologists prefer a user-friendly portal and intuitive interface to perform exploratory analysis of the data. The data of the model plants and major crops have been organized into the corresponding databases with user-friendly interface, including *A. thaliana* [15], rice (*Oriza sativa*) [16–18], maize (*Zea mays*) [19, 20], soybean (*Glycine max*) [21], sorghum (*Sorghum bicolor*) [22] and cotton (*Gossypium* sp.) [23, 24], etc. To facilitate the cross-species comparison, several comprehensive databases have been constructed, such as Phytozome [13], PlantGDB [25], Gramene [26]. Although these comprehensive databases already encompass the foxtail millet genomic data (Yugu1) [8], a dedicated database including the information of different varieties along with flexible analysis tools, are not yet available.

In this study, the MDSi (Multi-omics Database for *Setaria italica*, <http://sky.sxau.edu.cn/MDSi.htm>) was developed that contains two annotated genomes (*xiaomi* and Yugu1), the expression atlas from various tissues of *xiaomi* and JG21, transcription factors, transposons and ncRNAs datasets of *xiaomi*, molecular markers (SNPs and InDels) among 398 foxtail millet and green foxtail germplasms, as well as their metabolite traits. The database also provides intuitive search portal and common analysis tools, such as Blast, GBrowse and JBrowse. The Electronic Fluorescent Pictograph (xEFP) browser (Yang, Zhang et al. 2020), which can display the gene-expression data in-situ, is also another feature of MDSi. This large dataset will allow association mapping that will advance understanding of the genetic architecture of specific

phenotypes, which, in turn, will allow deep annotation of the foxtail millet genome. While genetic and genomic research is progressing in foxtail millet, the related database resources have lagged behind and are limited for next level analyses—functional genome studies such as QTL mapping.

Results

Gene search

We built a MDSi (Multi-omics Database for *Setaria italica*, <http://sky.sxau.edu.cn/MDSi.htm>, Fig. S1), and the gene data of both *xiaomi* and *Yugu1* were deployed in our database. The gene information of interest was able to be obtained either by Gene Search in Search menu or according to the corresponding annotation (Browse→Annotation) or known functions (Browse→Functional Category). At present, 161,844 annotations of 33,785 genes and 4,944 functional categories were collected depending on the databases of non-redundant proteins (Nr), Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO), respectively. The predicted transcription factors and

transposons were also classified in Browse menu, respectively. All the searching results guide users to the gene information (Gene info) page of the target gene (Fig.S2) in which several modules were integrated including Gene Structure (GBrowse, JBrowse), Gene Sequence, Gene Expression (the updated xEFP, Fig. 1), Protein Homologs and Gene Annotation.

Annotation

The gene annotations (161,844) were listed as a table in the Annotation module in Browse menu (Fig.S3). The genes of interest can be screened in this module by the Gene ID or annotated keywords. The planteome ontology information of the corresponding gene was indicated in the last column linking the Planteome ontology module in Search menu. The predicted transcription factors and transposons were independently listed in corresponding modules in Browse menu. There were 1848 transcript factors belong to 55 classes and 394,434 transposon related sequences within 40 classes. Moreover, 1120 trait ontologies of foxtail millet were screened referred to the

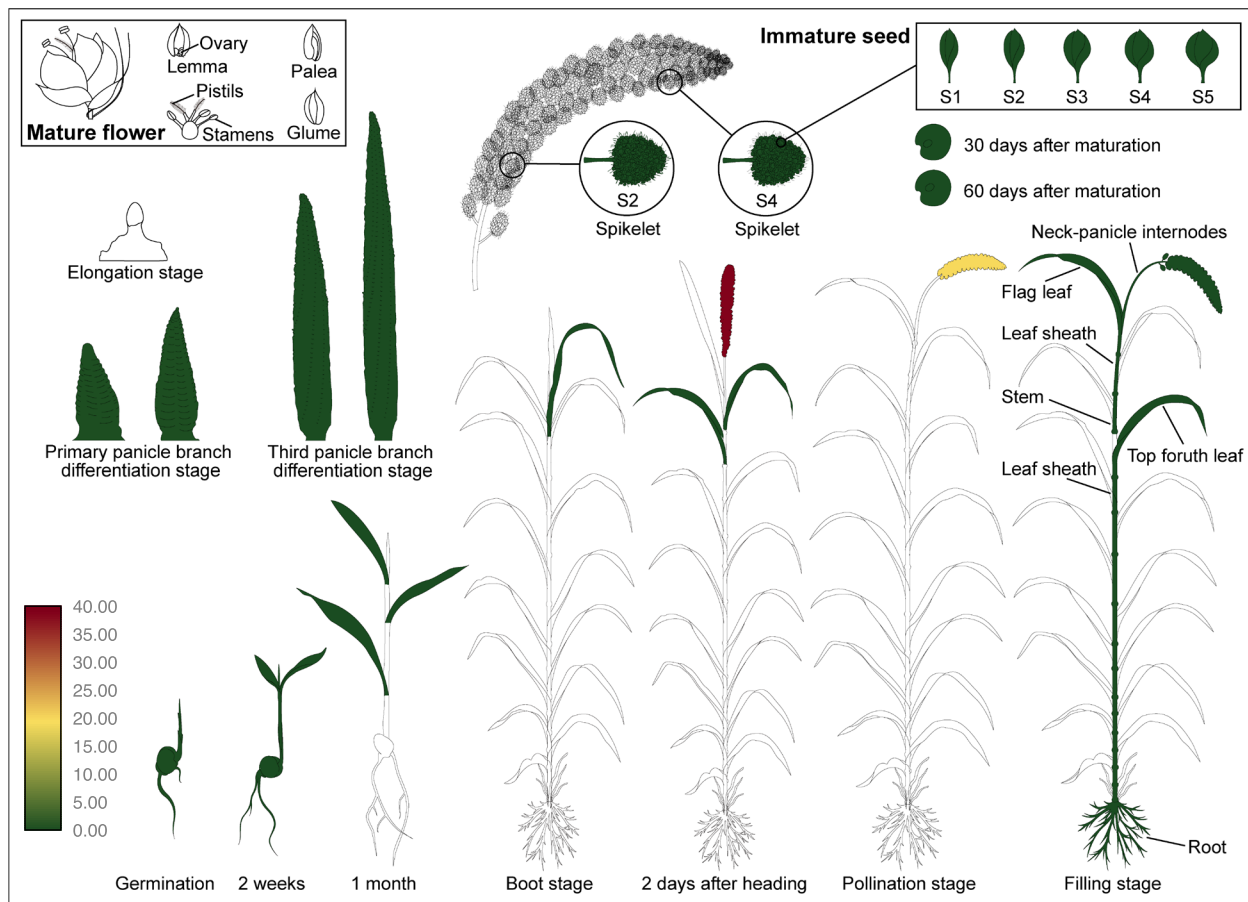


Fig.1 The visualization of the *Sitg01230* gene expression in different tissues

data in China rice data center (<https://www.ricedata.cn/ontology/>) that can be obtained in the Trait Ontology module in Browse menu or Trait Ontology Search module in Search menu.

Marker search

The 4,158,075 SNPs and InDels among 398 foxtail millet (360) and green foxtail (38) germplasm were collected and deployed. The information can be screened by region in chromosome, Gene ID or Variation ID in Variation Search page (Search menu, Fig.S4). JBrowse allows for the visualization of gene mutations and their impact on genes, including both synonymous, non-synonymous, stopgain, stoplost, startlost, splice, and so on (Tools → JBrowse, Fig.S5). Notably, the locations of the variations were also marked by the common features such as untranslated regions (UTR), exon and intron as so on. Additionally, the variation distributions of each germplasm were summarized in SNPs Variation and InDels Variation page (Browse menu, Fig.S6 and Fig.S7). The chromosome 8 shows highest variation frequency (the ratio of SNPs and InDels numbers to the length of corresponding region) in different regions including intron, CDS, UTR5, UTR3 and intergenic sequences (Fig. 2). Interestingly, the regions of CDS and intron take less variations, especially in latter sequences.

Online tools

BLAST was implemented using ViroBlast [27] for identifying sequence homology in genomic scaffolds, coding sequences, or proteins levels. The Gbrowse [28] and

JBrowse [29] were included simultaneously for gene structure visualization. The MDSi also provided a special tool for displaying the network of a gene of interest (Gene network in Tools menu) based on the context likelihood of relatedness (CLR) algorithm. For instance, total 20 genes would be related to the trithorax group protein osa-like (Si1g01230) with the p-value more than 0.98 (Fig. 3 and Fig.S8). The expression level of the screened or other genes of interest can be visualized in Expression Visualization module of Tools menu (Fig.S9). The corresponding raw data were also provided for further analysis. Both the tool and the aforementioned xEFP provided users with intuitive graphs of the gene expression patterns in different tissues at multiple- or single-gene level, respectively. The Chromosome Map can visualize the location of these genes (Fig.S10). The SiTRAIT and SiMetabolite were used to display data. The SiTRAIT (Tools → SiTRAIT) contains 18 agronomic traits, such as plant height, thousand seed weight, tiller number, length and width of leaf, etc. And the metabolic data of grains were also available in SiMetabolite (Tools → SiMetabolite).

Data download

The Download page (Fig.S11) was provided for users to obtain entire datasets as needed, including genomic scaffolds, predicted gene sequences and the corresponding proteins in the FASTA format. The germplasm information can also be collected directly here.

Chr1	0.67	0.70	0.30	0.16	0.59
Chr2	0.77	0.80	0.38	0.25	0.71
Chr3	0.75	0.79	0.36	0.22	0.66
Chr4	0.61	0.66	0.36	0.27	0.60
Chr5	0.63	0.67	0.31	0.19	0.60
Chr6	0.92	1.07	0.50	0.33	0.85
Chr7	0.83	0.86	0.40	0.25	0.82
Chr8	1.56	1.54	0.86	0.68	1.23
Chr9	0.52	0.56	0.24	0.14	0.53
	UTR5	UTR3	CDS	Intron	Intergenic

Fig.2 The variation frequency distribution of the SNPs and InDels

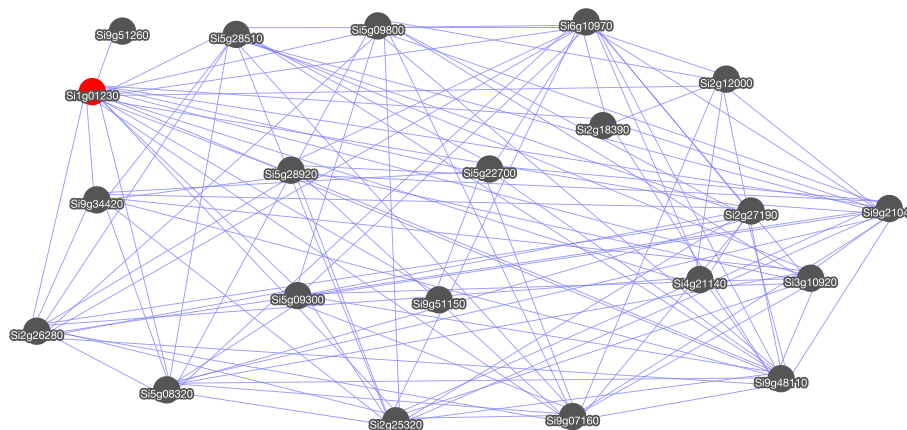


Fig.3 The gene network of *Si1g01230* gene

Moreover, the download functions also distributed in some modules described above if needed.

Discussion

Foxtail millet is being developed as a novel C_4 model which demands the high-throughput multi-omics data. At present, there were 5 de novo assembled genomic data of 4 elite cultivars [8–10, 12] and a mutant *xiaomi* with shorter life cycle [11]. We also published the RNA-seq data of 11 tissues of *xiaomi* at different stages [11]. Additionally, twenty-three RNA-seq data of the elite varieties JG21, a benchmark for high-quality foxtail millet [30] were also prepared. The whole-genome resequencing data of 398 germplasms, including 360 foxtail millets and 38 green foxtails, were reported for reveal genomic regions associated with many traits including 2557 metabolites identified by UPLC–MS/MS [14]. Recently, the restriction site associated DNA sequencing (RAD sequencing) data of a total of 1127 foxtail millet cultivars were published [31]. Therefore, the construction of a flexible and dedicated multi-omics database for foxtail millet is therefore an urgent priority.

At present, there is still no dedicated bioinformatics database for foxtail millet. The genome data of Yugu1 were implemented in Phytozome that provides a remarkably user-friendly interface for the cross-species comparison and analysis among multiple plants. However, for the time being, Phytozome does not provide in-depth and multidimensional species-specific data and analysis tools. Therefore, many dedicated databases for the important plants were constructed independently. FmMDB (Foxtail millet Marker Database) is the first database that organized and visualized large-scale marker datasets of Foxtail millet [32, 33], but unfortunately, it is no longer accessible. SIFGD (*S. italica* Functional Genomics Database) [34] was established for bioinformatics analyses of

gene function or regulatory modules using integrated the genome sequences of Yugu1 [8] and Zhanggu [9]. To this date, SIFGD remains an important foxtail millet genomics database. However, it is clear that, due to the progress of research at the time, it lacks relevant information at the transcriptional and metabolic levels.

The construction of a dedicated database for foxtail millet is necessary to facilitate effective analysis and utilization of the available genomic datasets. Such a database could provide comprehensive and multidimensional species-specific data, as well as sophisticated analysis tools, thereby promoting better understanding of the plant's molecular biology and accelerating the development of novel genomics-assisted breeding strategies. In comparison to the currently published foxtail millet genomes, *xiaomi* has been found to be the most complete. Similarly, when compared to other available foxtail millet genome databases, our MDSi ranks as the most comprehensive and reliable (Table 1).

Conclusion

MDSi constructed in this study follows the frontiers of cereal is the only multi-omics database that is currently operational and maintained. The MDSi constructed in this study integrated and visualized data from three levels of genomics, transcriptomics and metabolomics, and also provides information on the variation of hundreds of germplasm resources that can satisfies the mainstream requirements and supports the corresponding research community.

Materials and methods

Plant materials

xiaomi was identified from an EMS-mutagenized M2 population of Jingu21, a variety of foxtail millet widely cultivated in North China for its good grain quality and

Table 1 Comparison among the database of *xiaomi*, TT8, Yugu1, and Zhanggu

Statistic	<i>xiaomi</i>	Yugu1	Zhanggu	TT8
Estimated size (Mb)	438	510	485	NA
Assembly size (Mb)	429.9	405.7	423.4	477.6
Genome coverage	98%	80%	~86%	97%
Contigs	419	6,791	71,571	NA
Contig N50 (kb)	18,838.5	126.3	25.4	NA
Scaffolds	366	336	37,727	2,688
Scaffold N50 (Mb)	42.4	47.3	1.0	53.2
Number of gaps	48	6,455	34,282	7,136
Number of N's (bp)	NA	4,826,887	26,194,380	3,381,668
Genome wide annotation	org.Si.eg.db (R package)	NA	NA	NA
Database data sets	Genome, Transcriptome, Metabolome, Variations	Genome	Genome	NA
Database search	Gene, Variation, Planteome Ontologies, Trait Ontology	Keyword	Error: Access denied	NA
Database tools	Blast, Gene network, Jbrowse, Motif, Chromosome Map, SiTRAIT, SiMetabolite	Blast, Jbrowse, Motif, Synteny, GSEA	Error: Access denied	NA
Genetic transformation	Yes. The average transformation efficiency for NPTII is 23.28%	NA	NA	NA

high yield. The *xiaomi* mutant shows short life-cycle like *A. thaliana*. The *xiaomi* mutant was maintained by self-pollination in the laboratory for ten generations, leading to a very low level of heterozygosity. A total of 398 accessions, including 360 foxtail millets and 38 *S. viridis*, were also used in this study [14].

Architecture and implementation of MDSi

The MDSi was implemented in the linux system (CentOS 7), including PHP, JAVA, Apache web server, MySQL database management, and Perl CGI. An interactive Web interface was designed for users (Fig.S1). User query

information is transferred into PHP script and explained by corresponding interpreter invoked by apache web server (Fig. 4). The MDSi comprises a set of relational databases that store processed data in MySQL. To facilitate effortless access to the MDSi for basic research applications or biological analysis, an interactive web interface has been created. The interface utilizes PHP script to rapidly transmit user query information and extract data from the MySQL database management system to generate report pages. Additionally, the genome visualization tool has been integrated using the genetic genome browser, comprising Gbrowse [28] and Jbrowse [29],

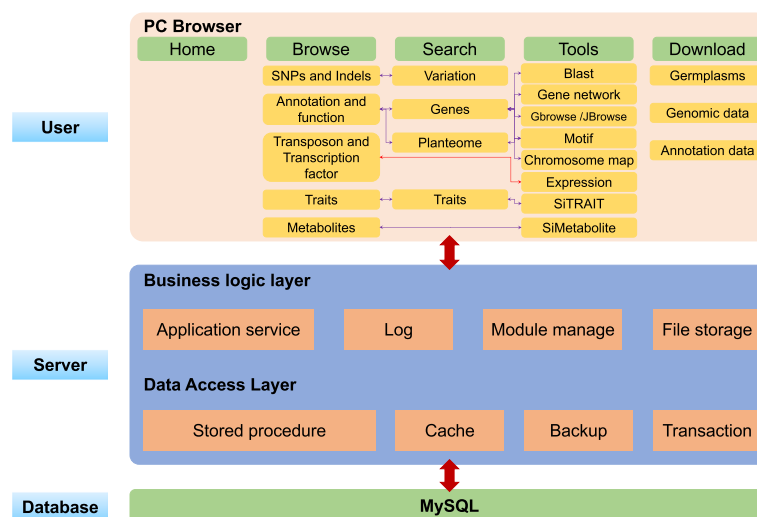


Fig.4 The schema architecture of MDSi

which provides improved performance and interface enhancements, including a robust faceted track selector. For interactive alignment of genome sequences, BLAST was performed by ViroBLAST, an independent web server for flexible queries of similar nucleotide and amino acid sequences. The browser finally displays the data by graphical manner.

Genomic data of xiaomi

We assembled reference-grade genome sequence (~429.9 Mb) de novo previously using three approaches, including the short read paired-end sequencing (Illumina HiSeq X Ten), the single molecule real-time (SMRT) sequencing (PacBio RSII and Sequel) and the chromosome conformation capture sequencing (Hi-C), for promoting its application in C_4 model system [11]. Compared with the previous Yugu1 [8] genome (~400.9 Mb) (9), Zhanggu [9] genome (~423.4 Mb), the *xiaomi* genome in MDSi is a high-quality, chromosome-level assembly that is more intact. The data were the core foundation of the MDSi.

Genomic data of germplasm

The genome of Yugu1 (~400.9 Mb), which was well visualized in Phytozome, was also added to the MDSi. Moreover, the MDSi also provides the whole-genome resequencing (WGS) results of 398 germplasm, including 360 foxtail millets and 38 green foxtails with an average of depth of 10 times [14]. The paired-end resequencing reads were mapped to the *xiaomi* reference genome with BWA using the default parameters. The SAMtools, Picard package and Genome Analysis Toolkit (GATK) and VCFtools were performed to filter duplicated reads, detect SNPs and filter data (missing data < 0.9, minor allele frequency ≥ 0.05 and depth ≥ 5). We identified 3,826,381 SNPs and 331,694 indels.

Transcriptomics data

To quantify the expression of known genes, we measured 29 RNA-seq in diverse tissues representing the major organ systems and varying developmental stages of *xiaomi* and JG21 with three biological repeats. RNA-seq data of 29 different tissues of *xiaomi* (6) and JG21 (23) including germinated-seed, seedling, leaf, leaf sheath, leaf-veins, mesophyll, internode, root, panicle, immature spikelet and seed at different stages or positions are encompassed in MDSi. The clean reads were obtained using Trimmomatic (0.38), and were subsequently aligned to *xiaomi* reference genomes using HISAT2 (v2.0.5). The Transcripts Per Kilobase per Million mapped reads (TPM) of the annotated genes were normalized using R with default parameters. A total of 25,949 informative genes with TPM more than 10 in at

least one tissue were obtained and gene pairs with the absolute similarity of expression correlation > 0.90 were used as the final dataset. On each gene page, tools for viewing temporal and spatial expression data are now embedded in the Expression table and render as soon as the page is selected. Mouse-overs display the read counts for individual data points. The values of genes expression transform set to log₂ to provide a more direct comparison.

Metabolic data

The metabolic data of grains of 398 germplasm obtained by ultra-performance liquid chromatography-tandem mass spectrometry (UPLC-MS/MS) were also available in the MDSi [14] (Browse → Metabolic Data).

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-023-04238-3>.

Additional file 1: Fig. S1. The home page of MDSi. **Fig. S2.** The gene info page of MDSi. **Fig. S3.** The annotation page of MDSi. **Fig. S4.** The variation search page of MDSi. **Fig. S5.** The JBrowse shows the gene mutations and their impact on Si9g49990 genes. **Fig. S6.** The SNP variation page of MDSi. **Fig. S7.** The InDel variation page of MDSi. **Fig. S8.** The network result page of Si1g01230 gene. **Fig. S9.** The network result of Si1g01230. **Fig. S10.** The chromosome location of genes related to Si1g01230. **Fig. S11.** The download page of MDSi.

Acknowledgements

We thank Dr. Yiwei Jiang (Purdue University) for their critical reading of the manuscript. We would like to thank the National Key R&D Program of China (nos. 2019YFD1000700, and 2019YFD1000704); the Central Government Guides the Local Science and Technology Development Fund Project (YDZJSX2021B010); and the National Natural Science Foundation of China (32001608) for their financial support.

Authors' contributions

Xingchun Wang and Jianhua Gao supervised the study. Yuanhui Han, Xukai Li and Xingchun Wang conceived of the study idea. Xukai Li, Siyu Hou and Xingchun Wang carried out the sampling process and participated in the material preparation. Xukai Li, Mengmeng Feng, Rui Xia, Jiawei Li, Sha Tang performed most of the experiments. Xukai Li and Jianhua Gao analyzed the data. Xukai Li and Jianhua Gao drafted the manuscript. All of the authors discussed the results and commented on the manuscript. The author(s) read and approved the final manuscript.

Funding

This work was supported by National Key R&D Program of China (nos. 2019YFD1000700, and 2019YFD1000704); the Central Government Guides the Local Science and Technology Development Fund Project (YDZJSX2021B010); and the National Natural Science Foundation of China (32001608).

Availability of data and materials

The genome assembly and annotation of *xiaomi* are also available at Genome Warehouse in the Beijing Institute of Genomics Data Center (<https://bigd.bigac.cn/>) under accession number GWHAZD00000000. The raw sequence data have been deposited in the Beijing Institute of Genomics Data Center with the following accession numbers: CRA001973 (Genome sequencing of *xiaomi* by PacBio), CRA001968 (Hi-C of *xiaomi*), CRA001972 (isoform sequencing of *xiaomi*), CRA001953 (RNA-seq of *xiaomi* tissues), CRA001954 (RNA-seq of Jingu21), CRA001974 (non-coding RNAs), and CRA002636 (genome resequencing of 398 accessions). The metabolomics data have been deposited

on the UCSD Metabolomics Workbench under the accession no. ST002168. The MDSi can be freely accessed at <http://sky.sxau.edu.cn/MDSi.htm> via the World Wide Web. A reliable data management system has been developed and all newly released information will be updated on this website. Inquiries concerning the database should be directed by email to xukai_li@sxau.edu.cn.

Declarations

Ethics approval and consent to participate

Study complied with local and national regulations for using plants.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 9 March 2023 Accepted: 20 April 2023

Published online: 27 April 2023

References

- Yang X, Wan Z, Perry L, Lu H, Wang Q, Zhao C, Li J, Xie F, Yu J, Cui T, et al. Early millet use in northern China. *Proc Natl Acad Sci USA*. 2012;109(10):3726–30.
- Bettinger RL, Barton L, Morgan C. The origins of food production in north China: a different kind of agricultural revolution. *Evol Anthropol*. 2010;19(1):9–21.
- Lu HE, Zhang B, Wang X, Hongying LI, Han Y. Foxtail millet: nutritional and eating quality, and prospects for genetic improvement. *Front Agric Sci Eng*. 2015;2(2):124–33.
- Doust AN, Kellogg EA, Devos KM, Bennetzen JL. Foxtail millet: a sequence-driven grass model system. *Plant Physiol*. 2009;149(1):137–41.
- Brutnell TP, Wang L, Swartwood K, Goldschmidt A, Jackson D, Zhu XG, Kellogg E, Van Eck J. *Setaria viridis*: a model for C4 photosynthesis. *Plant Cell*. 2010;22(8):2537–44.
- Li P, Brutnell TP. *Setaria viridis* and *Setaria italica*, model genetic systems for the Panicoid grasses. *J Exp Bot*. 2011;62(9):3031–7.
- Lata C, Gupta S, Prasad M. Foxtail millet: a model crop for genetic and genomic studies in bioenergy grasses. *Crit Rev Biotechnol*. 2013;33(3):328–43.
- Bennetzen JL, Schmutz J, Wang H, Percifield R, Hawkins J, Pontaroli AC, Estep M, Feng L, Vaughn JN, Grimwood J, et al. Reference genome sequence of the model plant *Setaria*. *Nat Biotechnol*. 2012;30(6):555–61.
- Zhang G, Liu X, Quan Z, Cheng S, Xu X, Pan S, Xie M, Zeng P, Yue Z, Wang W, et al. Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. *Nat Biotechnol*. 2012;30(6):549–54.
- Tsai KJ, Lu MJ, Yang KJ, Li M, Teng Y, Chen S, Ku MS, Li WH. Assembling the *Setaria italica* L. Beauv. genome into nine chromosomes and insights into regions affecting growth and drought tolerance. *Sci Rep*. 2016;6:35076.
- Yang Z, Zhang H, Li X, Shen H, Gao J, Hou S, Zhang B, Mayes S, Bennett M, Ma J, et al. A mini foxtail millet with an Arabidopsis-like life cycle as a C4 model system. *Nature Plants*. 2020;6(9):1167–78.
- Wang J, Li S, Lan L, Xie M, Cheng S, Gan X, Huang G, Du G, Yu K, Ni X, et al. *De novo* genome assembly of a foxtail millet cultivar Huagu11 uncovered the genetic difference to the cultivar Yugu1, and the genetic mechanism of imazethapyr tolerance. *BMC Plant Biol*. 2021;21(1):271.
- Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N et al. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res*. 2012;40(Database issue):D1178–1186.
- Li X, Gao J, Song J, Guo K, Hou S, Wang X, He Q, Zhang Y, Zhang Y, Yang Y, et al. Multi-omics analyses of 398 foxtail millet accessions reveal genomic regions associated with domestication, metabolite traits, and anti-inflammatory effects. *Mol Plant*. 2022;15(8):1367–83.
- Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, et al. The *Arabidopsis* Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res*. 2011;40(D1):D1202–10.
- Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, McCombie WR, Ouyang S, Schwartz DC, Tanaka T, Wu J, Zhou S, et al. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice*. 2013;6(1):4. <https://thericejournal.springeropen.com/articles/10.1186/1939-8433-6-4#citeas>.
- Sakai H, Lee SS, Tanaka T, Numa H, Kim J, Kawahara Y, Wakimoto H, Yang CC, Iwamoto M, Abe T, et al. Rice Annotation Project Database (RAP-DB): an integrative and interactive database for rice genomics. *Plant Cell Physiol*. 2013;54(2): e6.
- Sang J, Zou D, Wang Z, Wang F, Zhang Y, Xia L, Li Z, Ma L, Li M, Xu B, et al. IC4R-2.0: Rice genome reannotation using massive RNA-seq data. *Gen Proteomics Bioinform*. 2020;18(2):161–72.
- Woodhouse MR, Cannon EK, Portwood JL, Harper LC, Gardiner JM, Schaeffer ML, Andorf CM. A pan-genomic approach to genome databases using maize as a model system. *BMC Plant Biol*. 2021;21(1):385.
- Gui S, Yang L, Li J, Luo J, Xu X, Yuan J, Chen L, Li W, Yang X, Wu S, et al. ZEAMAP, a comprehensive database adapted to the maize multi-omics era. *iScience*. 2020;23(6):1–12.
- Grant D, Nelson RT, Cannon SB, Shoemaker RC. SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acids Res*. 2009;38(suppl_1):D843–6.
- Tian T, You Q, Zhang L, Yi X, Yan H, Xu W, Su Z. SorghumFDB: sorghum functional genomics database with multidimensional network analysis. *Database*. 2016;2016:1–16.
- Yu J, Jung S, Cheng CH, Ficklin SP, Lee T, Zheng P, Jones D, Percy RG, Main D. CottonGen: a genomics, genetics and breeding database for cotton research. *Nucleic Acids Res*. 2014;42(Database issue):D1229–1236.
- Zhu T, Liang C, Meng Z, Sun G, Meng Z, Guo S, Zhang R. CottonFGD: an integrated functional genomics database for cotton. *BMC Plant Biol*. 2017;17(1):1–9.
- Duvick J, Fu A, Muppirala U, Sabharwal M, Wilkerson MD, Lawrence CJ, Lushbough C, Brendel V. PlantGDB: a resource for comparative plant genomics. *Nucleic Acids Res*. 2007;36(suppl_1):D959–65.
- Tello-Ruiz MK, Naithani S, Gupta P, Olson A, Wei S, Preece J, Jiao Y, Wang B, Chougule K, Garg P, et al. Gramene 2021: harnessing the power of comparative genomics and pathways for plant research. *Nucleic Acids Res*. 2021;49(D1):D1452–63.
- Deng W, Nickle DC, Learn GH, Maust B, Mullins JI. ViroBLAST: a stand-alone BLAST web server for flexible queries of multiple databases and user's datasets. *Bioinformatics*. 2007;23(17):2334–6.
- Stein LD. Using GBrowse 2.0 to visualize and share next-generation sequence data. *Brief Bioinform*. 2013;14(2):162–71.
- Buels R, Yao E, Diesh CM, Hayes RD, Munoz-Torres M, Helt G, Goodstein DM, Elsik CG, Lewis SE, Stein L, et al. JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol*. 2016;17(1):66.
- Ma K, Zhao L, Zhao X, Li X, Dong S, Zhang L, Guo P, Yuan X, Diao X. The relationship between ecological factors and commercial quality of high-quality foxtail millet "Jingu 21." *Food Res Int*. 2023;163: 112225.
- Wang Y, Wang X, Sun S, Jin C, Su J, Wei J, Luo X, Wen J, Wei T, Sahu SK, et al. GWAS, MWAS and mGWAS provide insights into precision agriculture based on genotype-dependent microbial effects in foxtail millet. *Nat Commun*. 2022;13(1):5913.
- Bonthala VS, Muthamilarasan M, Roy R, Prasad M. FmTFDB: a foxtail millet transcription factors database for expediting functional genomics in millets. *Mol Biol Rep*. 2014;41(10):6343–8.
- Vs B. Muthamilarasan M, Misra G, Prasad M: FmMDB: a versatile database of foxtail millet markers for millets and bioenergy grasses research. *PLoS ONE*. 2013;8(8): e71418.
- You Q, Zhang L, Yi X, Zhang Z, Xu W, Su Z. SIFGD: *Setaria italica* Functional Genomics Database. *Mol Plant*. 2015;8(6):967–70.

Publisher's Note

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