

EfGD: the *Erianthus fulvus* genome database

Zhenfeng Qian^{1,2,3,†}, Xuzhen Li^{4,5,†}, Lilian He^{2,3}, Shujie Gu^{2,3}, Qingqing Shen^{2,3}, Xibing Rao^{2,3}, Rongqiong Zhang^{2,3}, Yining Di^{2,3}, Linyan Xie^{2,3}, Xianhong Wang^{1,2,3}, Shuying Chen^{2,3}, Yang Dong^{4,5,*} and Fusheng Li^{1,2,3,*}

¹The Key Laboratory for Crop Production and Intelligent Agriculture of Yunnan Province, Yunnan Agricultural University, No. 95 Jinhei Road, Kunming 650201, China

²College of Agronomy and Biotechnology, Yunnan Agricultural University, No. 95 Jinhei Road, Kunming 650201, China

³Sugarcane Research Institute, Yunnan Agricultural University, No. 95 Jinhei Road, Kunming 650201, China

⁴State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan, Yunnan Agricultural University, No. 95 Jinhei Road, Kunming 650201, China

⁵College of Biological Big Data, Yunnan Agriculture University, Kunming, No. 95 Jinhei Road, Yunnan 650201, China

*Corresponding author: Tel: +8613888191720; Fax: +0086-871-65227728; Email: Lfs810@sina.com

Correspondence may also be addressed to Yang Dong. Tel: +8613577016903; Fax: +0086-871-65220061; Email: loyalyang@163.com

[†]These authors contributed equally to this work.

Citation details: Qian, Z., Li, X., He, L. *et al.* EfGD: the *Erianthus fulvus* genome database. *Database* (2022) Vol. 2022: article ID baac076; DOI: <https://doi.org/10.1093/database/baac076>

Abstract

Erianthus fulvus (TaxID: 154759) is a valuable germplasm resource in sugarcane breeding and research and has excellent agronomic traits, such as drought resistance, cold resistance, barren tolerance and high brix. With a stable chromosome number ($2n = 20$) and a small genome (0.9 Gb), it is an ideal candidate for research on sugarcane. Next-generation sequencing technology has enabled a growing number of studies to focus on genomics. Due to the large amount of omics data available, a centralized platform is necessary for ensuring the consistency, independence and maintainability of these large-scale datasets through storage, analysis and integration. Here, we present a comprehensive database for the *E. fulvus* genome, EfGD. By using the new high-quality reference genome and its annotations, the EfGD provides the largest whole-genome sequencing reference dataset for *E. fulvus*, which archives 27 165 protein-coding genes and 55 564 488 SNPs from 202 newly resequenced genomes. Furthermore, we created a user-friendly graphical interface for visualizing genomic diversity, population structure and evolution and provided other tools on an open platform.

Database URL: <https://efgenome.ynau.edu.cn>

Introduction

Sugarcane produces 80% of the world's sugar and 40% of biofuels and is an important cash crop. Sugarcane requires a significant amount of water, fertilizer and heat for its growth and development, which is why it is primarily found in tropical and subtropical climates. Continuous population on a global scale, a limited amount of sugarcane planting areas and the frequent occurrence of extreme weather conditions are driving the demand to improve sugarcane yield, quality and adaptability (1–3). This can be addressed with the use of appropriate germplasm materials. To acquire new cytoplasmic and nuclear genomes in sugarcane hybrids, the wild species of *Erianthus fulvus* Ness (axID: 154759), *Saccharum spontaneum* L. (TaxID: 4547) and *Erianthus arundinaceus* Retz. (TaxID: 50346) have been used as hybrid parents to produce an array of advanced generation hybrids (4, 5).

Erianthus fulvus Ness is a wild species of *Erianthus* that is close to *Saccharum*. It is mainly distributed in E 97.83 to 106.40 latitude and N 24.38 to 28.27 longitude and at

altitudes between 480 and 3003 m (6, 7). *Erianthus fulvus* has many characteristics that are very useful for improving sugarcane varieties, including early maturity, high brix, drought resistance, cold resistance and barren resistance (8). Furthermore, *E. fulvus* has a stable chromosome number, $2n = 20$ (9), and there is genetic diversity among all clones (8, 10). Our team investigated and collected *E. fulvus* resources in China as early as 1985 (6) and used the resources for germplasm resource research and genetic improvement of sugarcane varieties (11).

Although *E. fulvus* has great value in sugarcane variety improvement, little research has been conducted on its heredity, variation and functional genes, especially when compared to sugarcane. To better utilize *E. fulvus* resources and investigate the regulatory mechanisms of traits, we previously performed next-generation transcriptome sequencing on *E. fulvus* (12) and uncovered several drought- and cold-tolerance-related genes, including *DREB* (dehydration responsive element binding protein) (13), *CML* (Calmodulin-like) (14),

Received 28 May 2022; Revised 10 August 2022; Accepted 25 August 2022

© The Author(s) 2022. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License

(<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com



Figure 1. Architecture of EFGD. The EFGD was divided into eight main modules including 'Home', 'Genome', 'Re-sequence', 'Gene', 'Variation', 'Download', 'About' and 'Tools'. The corresponding contents of each module are listed below.

NAC (NAC domain transcription factor) (15) and GRAS (gibberellic acid insensitive, repressor of GAI, and scarecrow) (16). In this study, we produced the whole-genome sequence of *E. fulvus*. The *E. fulvus* genome (0.9 Gb) is much smaller than the *S. spontaneum* L. genome (3.13 Gbp) (17). In addition, 202 samples from Saccharum complex, including 97 *Erianthus*, 56 *Saccharum*, 19 *Miscanthus*, 5 *Narenga* and 1 *Eulalia* were used for whole-genome resequencing and analysis, and a total of 55 564 488 single nucleotide polymorphisms (SNPs) were identified, creating a fine genetic map of *E. fulvus* and providing a large number of genetic markers for the molecular breeding of sugarcane. *Erianthus fulvus* genome and transcriptome information can be obtained at National Center for Biotechnology Information (NCBI) (BioProject ID: PRJNA854329).

In this work, we further developed a comprehensive database (*E. fulvus* genome database, EfGD) for data storage, classification, online analysis and visualization of *E. fulvus* and sugarcane omics data. The EfGD provides scientific researchers with data on the transcriptome and genome sequence for *E. fulvus*, as well as phenotype, resequencing and SNP information for 202 samples from Saccharum complex. In addition, the EfGD also provides search functions, such as BLAST, JBrowse and Publication services, and all data download functions. All data in the EfGD can be downloaded for free.

Materials and methods

Data content

A hybrid sequencing approach of short Illumina reads, long PacBio reads and HiC reads were used to construct the *E. fulvus* ‘ZM13’ genome. The PacBio sequencing resulted in ~132.3 Gb long-reads, Illumina sequencing resulted in ~147.4 Gb short-reads and Hi-C sequencing resulted in 189 Gb (189-fold) high-quality clean reads. In addition, 202 samples, including 121 *Erianthus*, 56 *Saccharum*, 19 *Miscanthus*, 5 *Narenga* and 1 *Eulalia* (Supplementary Table S1), were whole-genome resequenced to generate 2.7 Tb (27.3 billion) paired-end raw reads. Moreover, 14 RNA-seq sequences of leaf tissues (i.e. ZM13, leaf cold 24 h, Er_LT1_1.fq.gz) were also carried out to acquire 23.05 Gb raw RNA-seq data (15). All of these data were integrated into the EfGD. The assembled *E. fulvus* genome was 0.9019 Gb. Using this assembly as the

reference, genome annotation, variant calling and population genetic analysis were conducted. See the following sections for methods.

Genome annotation

The *E. fulvus* genome assembly was annotated for gene content using the NCBI Eukaryotic Genome Annotation Pipeline (18). Tandem Repeats Finder v4.09 (19) was used to search for tandem repeats in repeat annotation. Gene prediction was performed through de novo, homology-based and transcriptome-based approaches. For de novo prediction of protein-coding genes, we used GeneID v1.4.4 (20), GlimmerHMM (21), SNAP (22) and GenScan (23) with default settings. For homology-based gene prediction, protein sequences from *Oryza sativa* (TaxID: 4530), *Arabidopsis thaliana* (TaxID: 3702), *Brachypodium distachyon* (TaxID: 15368), *Sesamum indicum* (TaxID: 1405819), *Zea mays* (TaxID: 4577) and *Sorghum bicolor* (TaxID: 4558) were used for blast analysis. Trinity v2.9.0 (24) assembled transcripts of *E. fulvus* were aligned against the *E. fulvus* genome to generate transcriptome-based gene models using Program to Assemble Spliced Alignments v2.4.1 (25). Protein-coding gene functions were assigned according to the best match using Blastp against Swiss-Prot, Translation of EMBL (TrEMBL) (26) and Kyoto Encyclopedia of Genes and Genomes (KEGG) (27). The InterProScan functional analysis and Gene Ontology (GO) IDs were obtained with InterProScan (28). The GO enrichment was performed with Ontologizer 2.0 (29) with a *P* value cutoff of 0.05.

Digital gene expression profiles

RNA-seq datasets of *E. fulvus* ‘ZM13’ covering low temperature and drought stress treatment of leaves were selected. These raw RNA-seq reads were mapped to the latest reference genome by HISAT2 (30). The matched reads are then presented to StringTie (31) for de novo assembly, during which the expression level of each gene and isoform is estimated.

Variants calling

Whole-genome resequencing data of 202 samples of the Saccharum complex species was generated and mapped to *E. fulvus* reference genome. We performed variant calling for

A

EfGD: *Erianthus fulvus* Genome Database

Erianthus fulvus is a near-source wild species of sugarcane that has the advantages of early maturation, high sugar content, and drought resistance that is of great value to the genetic improvement of sugarcane varieties. In a previous study, we sequenced the whole genome of *E. fulvus* using the second-generation (Illumina) and third-generation (PacBio) sequencing technologies. NextDenovo v2.1 was used for genome assembly analysis and showed that the assembly size was 901.9 Mb. Jvarkit-BioKit was used to mount the Hi-C chromosome on the assembled contigs. All 10 chromosomes (2n=20) were anchored and the mount rate was 93%. We have annotated 27165 protein-coding genes located on 10 chromosomes in the *E. fulvus* genome. On this basis, we performed population genome analysis on 202 materials including 121 *Erianthus*, 56 *Saccharum*, 19 *Miscanthus*, 5 *Narenga*, and 1 *Eulalia*, and annotated 555448 SNPs on 10 chromosomes. These data provide a wealth of genetic information for researchers. For this purpose, we have developed a comprehensive and easy-to-use online database (*Erianthus fulvus* Genome Database, EfGD).

News & events >>

- 07-15 2020 Enhanced sugar accumulation and regulated plant hormone signaling genes contribute to cold tolerance in hypopolysaccharide *Saccharum spontaneum*
- 08-05 2019 Comparative Analysis of two Sugarcane Ancestors *Sa cocharum officinarum* and *S. spontaneum* based on Complete Chloroplast Genome Sequences and Photosynthetic Ability in Cold Stress
- 09-09 2010 Characterization of the Chromosomal Transmission of Intergenic Hybrids of *Saccharum* spp. and *Erianthus fulvus* by Genomic *in situ* Hybridization

External links >>

EnsemblPlants, NCBI National Center for Biotechnology Information, UniProt, SMART, The MEME Suite, Pfam, SGD

B

About EfGD

Erianthus fulvus is a near-source wild species of sugarcane that has the advantages of early maturation, high sugar content, and drought resistance that is of great value to the genetic improvement of sugarcane varieties. In a previous study, we sequenced the whole genome of *E. fulvus* using the second-generation (Illumina) and third-generation (PacBio) sequencing technologies. NextDenovo v2.1 was used for genome assembly analysis and showed that the assembly size was 901.9 Mb. Jvarkit-BioKit was used to mount the Hi-C chromosome on the assembled contigs. All 10 chromosomes (2n=20) were anchored and the mount rate was 93%. We have annotated 27165 protein-coding genes located on 10 chromosomes in the *E. fulvus* genome. On this basis, we performed population genome analysis on 202 materials including 121 *Erianthus*, 56 *Saccharum*, 19 *Miscanthus*, 5 *Narenga* and 1 *Eulalia*, and annotated 555448 SNPs on 10 chromosomes. These data provide a wealth of genetic information for researchers. For this purpose, we have developed a comprehensive and easy-to-use online database (*Erianthus fulvus* Genome Database, EfGD). EfGD is committed to providing a comprehensive omics database for researchers. At present, EfGD has realized the integration of important datasets including the sequencing, assembly, functional gene annotation, gene expression, phenotype, resequencing, and variation annotation of the *E. fulvus* genome. Also, EfGD provides data analysis and visualization tools for Blast and Brw. The integration of these data and tools makes EfGD a valuable database and all data in EfGD can be freely downloaded.

Architecture

Funding

The sub-project of the National Key Research and Development Program of China (2019YFD1000503); National Natural Science Foundation of China (31960051, 31560113); Key Project of Applied Basic Research Program of Yunnan Province (2015FA004); ESI Discipline Promotion Program of Yunnan Agricultural University (2019YNAUESHMS01)

Contact us

Group: Study on genetic diversity of sugarcane germplasm resources
Lab: The Key Laboratory for Crop Production and Intelligent Agriculture of Yunnan Province
Institute: Sugarcane Research Institute of Yunnan Agricultural University
Address: Sugarcane Research Institute of Yunnan Agricultural University, Kunming, 650201, China
E-mail: lsh10@ynia.com

Figure 2. ‘Home’ and ‘About’ modules. (A) ‘Home’ module, basic introduction to EfGD and external quick links. (B) ‘About’ module, detailed introduction to EfGD.

202 genomes using the ‘HaplotypeCaller’ module of GATK v3.3-0-g37228af (32) with the ‘-ERC GVCF’ setting, and all the gVCF output files were combined into a Variant Call Format (VCF) file using ‘CombineGVCFs’. Joint genotyping was conducted for each of the chromosomes using ‘GenotypeGVCFs’ with the ‘-L’ setting (33). To reduce the error rate of calling variants, SNPs were filtered by hard-filtering of ‘VariantFiltration’. SNPs were filtered with the—filter

Expression ‘QUAL < 30.0 || QD < 2.0 || MQ < 40.0 || FS > 60.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0 || SOR > 5.0’ (34). Only insertions and deletions shorter than or equal to 40 bp were considered. Indels and SNPs with none biallelic, >50% missing calls and minor allele frequency (MAF) < 0.005 were removed, which yielded the basic set. Finally, a total of 55 564 488 SNPs from 202 samples including 121 Erianthus, 56 Saccharum, 19 Miscanthus, 5 Narenga and 1 Eulalia were

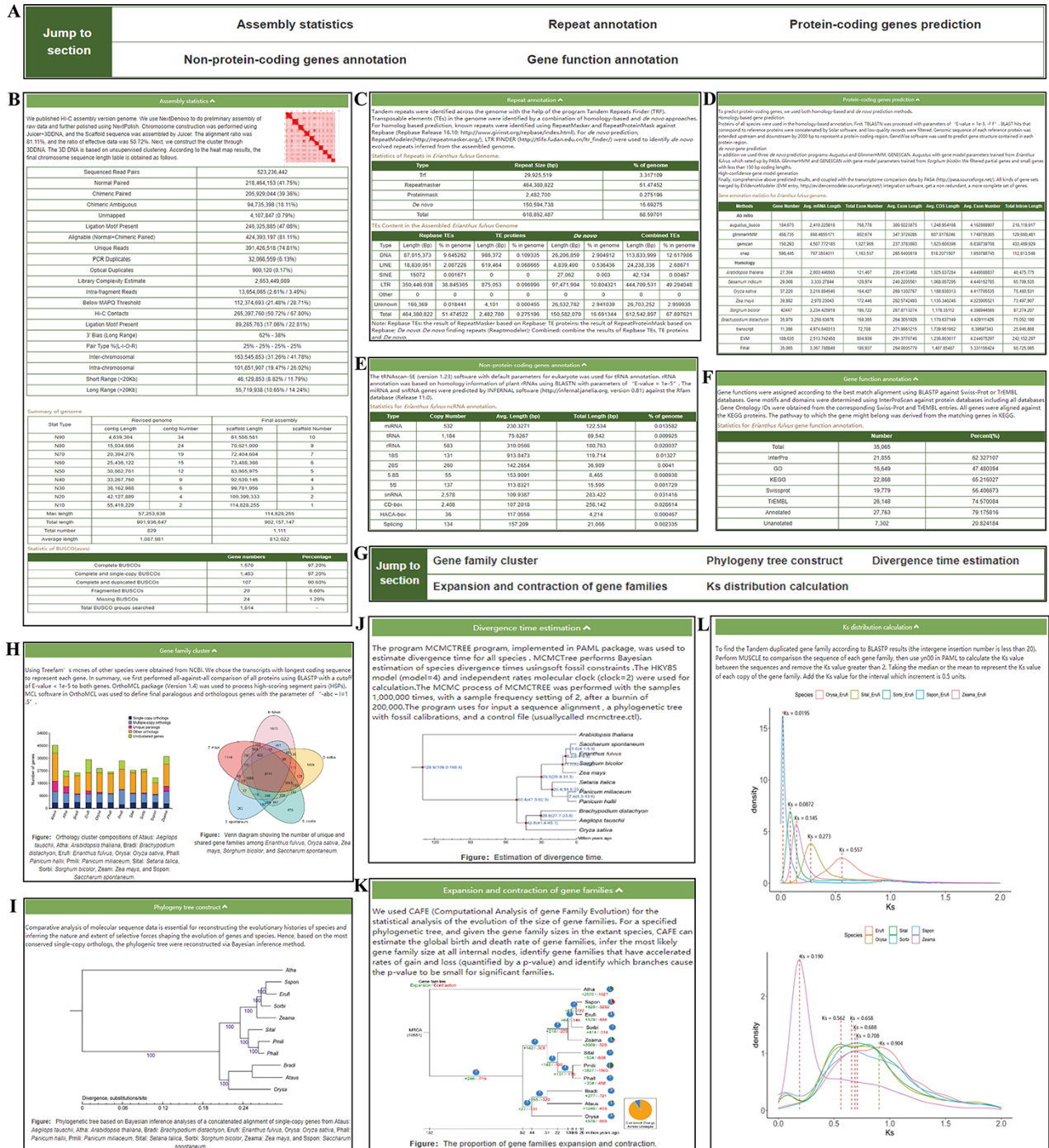


Figure 3. ‘Genome’ module. The ‘Genome’ module has subpages: ‘Statistics’ (A–F) and ‘Analysis’ pages (G–L). (A) The navigation bar of the ‘Statistics’ page provides quick access to each section. (B) Assembly statistics. (C) Repeat annotation. (D) Protein-coding genes prediction. (E) Nonprotein-coding genes annotation. (F) Gene functional annotation. (G) The navigation bar of the ‘Analysis’ page. (H) Gene family cluster. (I) Phylogenetic tree. (J) Divergence time. (K) Expansion and contraction of gene families. (L) Ks distribution.

identified. For each population, variants with an MAF less than 0.05 were filtered. Based on the annotation file, all the qualified variants were annotated using ANNOVAR v2015-12-14 (35). An in-house script was used to count the number of each annotation item.

Population structure and phylogeny

We pruned the autosomal SNPs to remove linkage disequilibrium (LD) by PLINK (Version 1.90b3.38) (36) with

‘-indep-pairwise 50 10 0.2’. Principal component analysis (PCA) was performed using Genome-wide Complex Trait Analysis (version: 1.25.3) (37). We conducted ADMIXTURE (version 1.3.0) (38) analyses with K (individual ancestry component) from 2 to 10 to infer population structure. For each K, 10 replicates were run with random seeds between 1 and 100. The seed with the lowest cross-validation error was used in the database. We used the whole-genome SNPs to construct the maximum likelihood (ML) phylogenetic tree with

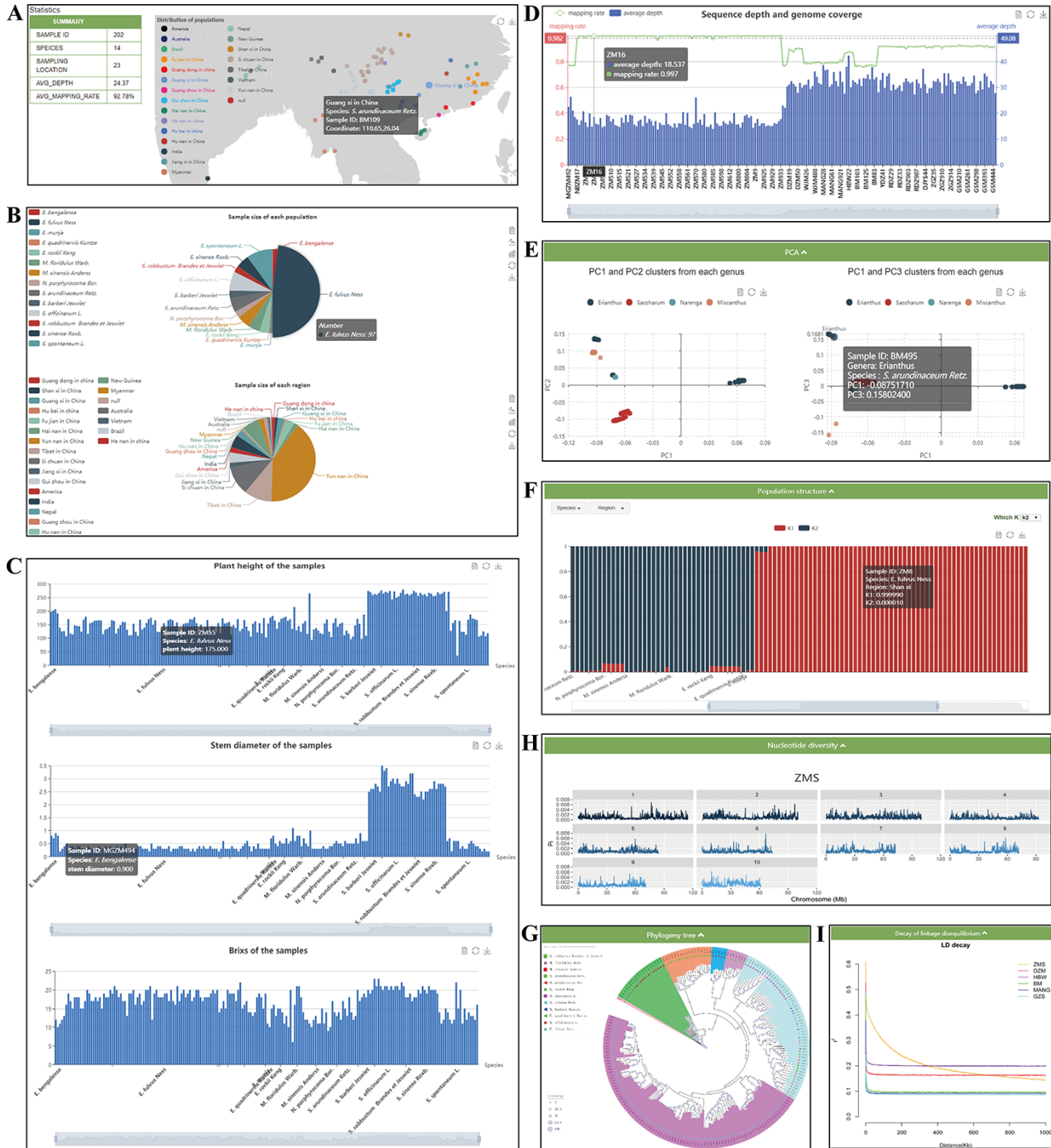


Figure 4. ‘Re-sequence’ module. The ‘Re-sequence’ module has subpages: ‘Statistics’ (A–D) and ‘Analysis’ pages (E–I). (A) Geographic distribution of 202 individuals. (B) Sample size of each species and region. (C) Statistics of plant height, stem diameter and brix for each sample. (D) Resequencing quality statistics. (E) PCA clusters from 202 individuals. (F) Bar plot of admixture analysis for all samples with K from 2 to 10. (G) Phylogenetic tree. (H) Nucleotide diversity statistics. (I) Decay of LD.

100 bootstraps using MrBayes (version: 3.1.2) (39). The tool iTOL (<http://itol.embl.de>) was used to color the phylogenetic tree.

Implementation

The EfGD is based on the Apache webserver (<http://www.apache.org>), adopts ThinkPHP5.1 (<http://www.thinkphp.cn>) based Fastadmin template and applies programming language including PHP-HTML5 and JavaScript (JS). MySQL (<https://www.mysql.com>) is used for data sorting, storage and management, and the AJAX asynchronous loading scheme is used for quick data loading and function implementation. To provide an interactive use experience, Echarts (<https://echarts.apache.org/zh/index.html>), JBrowse (40) and viroBlast (41) have been applied.

Results

To organize different types of data and analyze the results, the EfGD was divided into eight main modules (Figure 1): the ‘Gene’ and ‘Variation’ modules are the two most important modules providing detailed information on genes and variation data; the ‘Genome’ and ‘Re-sequence’ modules provide an overall view of the genetic background of *E. fulvus* and phylogenetic relationship of Saccharum complex populations; the ‘Home’ and ‘About’ modules introduce the database, project and our team and the ‘Download’ module enables the download of several data types. The ‘Tools’ module provides BLAST, JBrowse and Publication services. For each module, a data description and basic functions can be seen below.

Home and about

The ‘Home’ module is a landing page of the EfGD that provides an overview of the database. In addition, ‘Home’ provides external quick links to databases and news related to the Saccharum complex (Figure 2A). The ‘About’ module

provides a general introduction to the EfGD and detailed contact information (Figure 2B).

Genome

The ‘Genome’ module has two sections: ‘Statistics’ and ‘Analysis’ (Figure 3). Five parts are included in the ‘Statistics’ section: assembly statistics, repeat annotation, protein-coding genes, nonprotein-coding genes and functional annotation. This information will help users understand the *E. fulvus* genome more clearly. Some basic results are also included in the ‘Analysis’ section, such as gene family cluster, phylogeny tree, divergence time, expansion and contraction of gene families and Ks distribution.

Re-sequence

The ‘Re-sequence’ module contains two sections: ‘Statistics’ and ‘Analysis’ (Figure 4). ‘Statistics’ presents summaries for samples embedded in the database. This page shows real-time statistics of all samples embedded in the database, and the sample information is counted from different angles and a variety of statistical graphs. The geographic distribution and summary information are shown in Figure 4A. In addition, pie charts of the sample size of each species and region are provided (Figure 4B). The phenotypic statistics of plant height, stem diameter and brix of the 202 samples are provided by the histogram (Figure 4C). Resequencing quality statistics for each sample are shown at the bottom (Figure 4D). All the figures on this page are dynamic and interactive, and users can view the current group information more clearly.

In the ‘Analysis’ section, we provide five dynamic and interactive charts, so that users can view the information from 202 samples more clearly. The PCA part presents results for two datasets: all 202 individuals belonging to *Erianthus* and outgroup taxa in PC1 with PC2, and the other is PC1 with PC3 (Figure 4E). The population structure part presents the

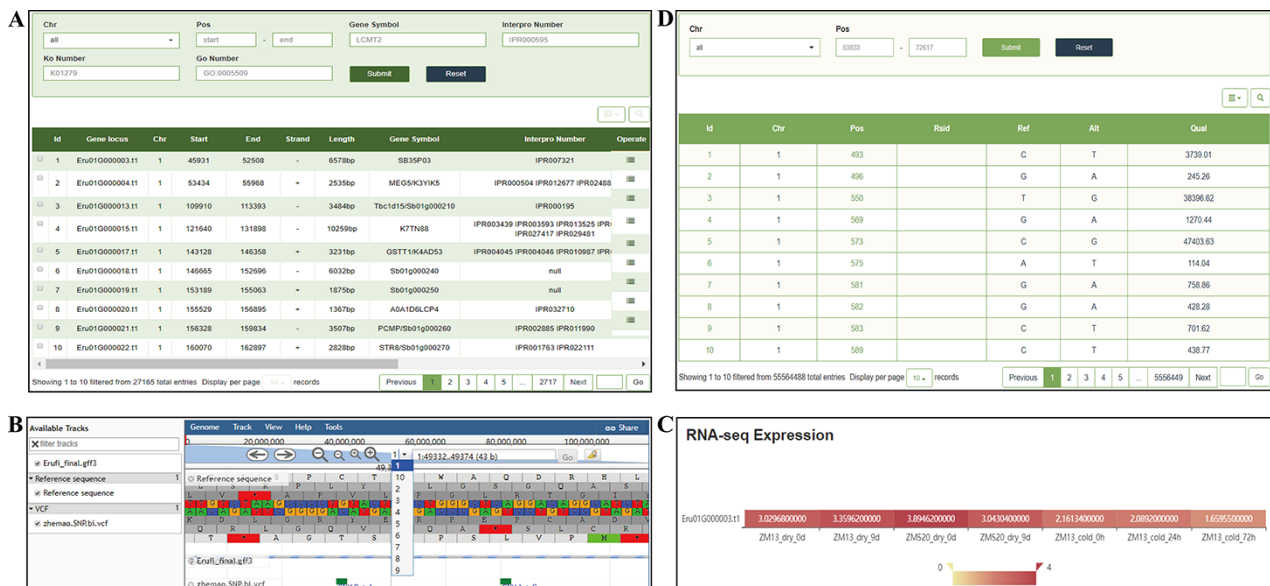


Figure 5. ‘Gene’ and ‘Variation’ modules. (A) ‘Gene’ search module. (B) Click on Gene ID to display detailed information in the JBrowse. (C) Click on Gene ID to display the expression level. (D) ‘Variation’ module.

proportions of the proposed ancestry components in each sample with bar plots. The length of each colored bar indicates the proportion of representative ancestry in each individual. The number of proposed ancestries is defined by the 'which K' option includes ADMIXTURE clustering results for populations with K from 2 to 10 (Figure 4F). In the

phylogeny tree section (Figure 4G), the ML tree of 202 samples was constructed with 100 bootstraps. In the nucleotide diversity section (Figure 4H), boxplots were used to show the genetic diversity indexes from six Saccharinae populations. The decay of the LD part presents the results for six populations (Figure 4I). In addition, the small components

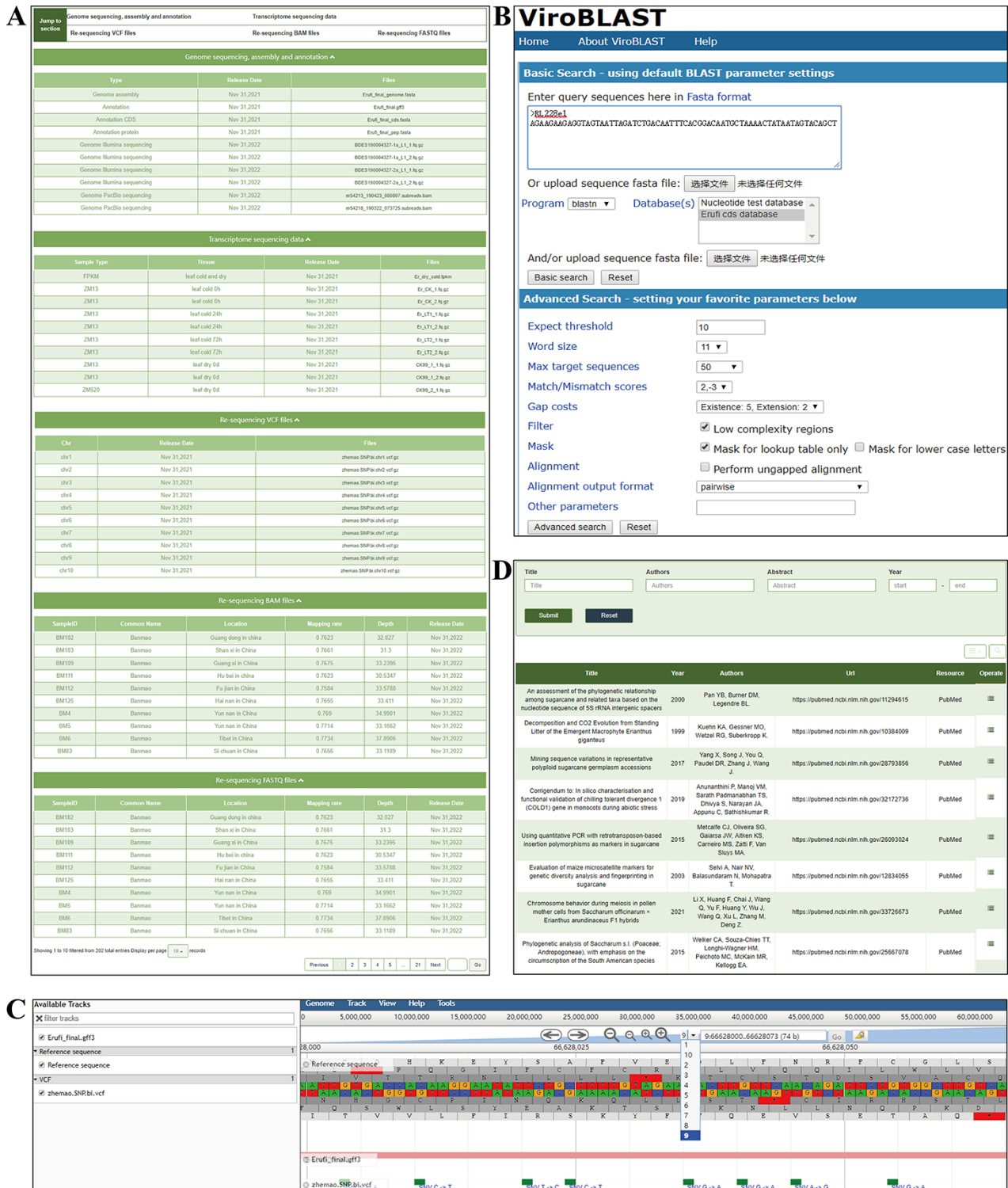


Figure 6. 'Download' and 'Tools' modules. (A) 'Download' module, providing genome, transcriptome, variation and resequencing files to download. (B) BLAST Tool. (C) JBrowse tool. (D) Publication tool.

equipped for each graph allow filtering by the legend and the structural data can be presented by clicking the button in the upper right corner.

Gene

The ‘Gene’ page provides gene data from the latest assembly and annotation of the *E. fulvus* genome. Users can enter the InterPro domain, GO and KEGG orthology identifiers, key words for gene symbols or gene positions in the search box for genes or gene families and the detailed corresponding information, such as the structural and functional annotation of genes, can be acquired. Each gene entry contains 10 pieces of information, such as gene locus, chromosome location and length, and the users can select the columns that show the information they want to focus on through the upper right option of the display list (Figure 5A). In addition, we deployed a mini JBrowse window part under the table, which allows viewing the location and variation of the target gene, and the detailed sequence information by clicking (Figure 5B). In the RNA-seq expression section, users can select genes according to specific needs and observe the gene expression level from the heatmap of the ticked-related genes in the gene expression module (Figure 5C).

Variation

In the current version, the ‘Variation’ module contains 55 564 488 nonredundant SNPs from 202 resequencing materials. A search tool allows users to search information of SNP according to ‘Chr’ and ‘Pos’. The search results provide chromosome number, specific sites, Reference SNP ID (RsID) and quality information (Figure 5D).

Download

The ‘Download’ module contains information and links to almost all publicly available *E. fulvus* genome and transcriptome datasets, all of which are published for the first time in this database (Figure 6A). The data of the genome, transcriptome sequencing, variation files and all individuals’ whole-genome resequencing files can be downloaded from this module. If users have any questions about data uploading or downloading, they can contact us using the published contact information.

Tools

Three tools, BLAST, JBrowse and Publication, are integrated into the ‘Tools’ module, which will help users perform web-based BLAST with our assembly, focus on the target genomic region and find literature and/or books related to the Saccharum complex. The query sequence in ‘BLAST’ sections can be entered by either copying or uploading the sequence file. After setting parameters and clicking on the search bottom, the alignment result will appear on a new page, with the identities and percentage between query and subject and the alignment score. The final BLAST result can be downloaded in FASTA format (Figure 6B). Sets of genomic, gene and density information of SNP/300 kb allele frequencies can be explored using the JBrowse. Users can choose tracks of interest in the ‘Select Tracks’ menu and a region of interest. As a combination of database and interactive web pages, *E. fulvus* genomes, gene annotation and variants information have been imported into JBrowse. In addition, by clicking a

specific data entry, the detailed feature page of data entries can be opened (Figure 6C). The ‘Publication’ section contains a total of 910 studies or books. By searching key words in title, abstract, author and/or year of publication, a list is provided and clicking on ‘Operate’ opens a new page with detailed information, such as the abstract, publication type, etc. (Figure 6D). In future updates, other interesting data will be added to the genome browser.

Conclusion and plan

The EfGD is committed to providing a comprehensive omics database for researchers. At present, the EfGD has integrated important data, including the sequencing, assembly, functional gene annotation, gene expression, phenotype, resequencing and variation annotation of the *E. fulvus* genome. In addition, the EfGD provides tools BLAST and JBrowse for data analysis and visualization. The integration of these data and tools makes the EfGD a valuable database.

In the future, we will update the EfGD regularly, develop more convenient online data analysis tools and integrate multi-omics, variety resources, phenotypic and trait regulation to build a comprehensive research and analysis database of *E. fulvus*. At the same time, we hope that many researchers will take advantage of these resources and provide comments and suggestions so that the EfGD can become a one-stop *E. fulvus* research platform with multifaceted omics data and analysis tools.

Supplementary data

Supplementary data are available at Database Online.

Acknowledgements

We thank many users for their valuable suggestions regarding the EfGD, thank Professor Tong Li, Director of The Key Laboratory for Crop Production and Intelligent Agriculture of Yunnan Province, for his support and thank the Network Management Center of Yunnan Agricultural University for network optimization.

Funding

Major Science and Technology Projects in Yunnan Province (202202AE090021); Biological Resources Digital Development and Application Project (202002AA100007); Special project of The Key Laboratory for Crop Production and Intelligent Agriculture of Yunnan Province (202105AG070007); Subproject of the National Key Research and Development Program of China (2018YFD1000503); National Natural Science Foundation of China (31960451 and 31560417); Key Project of Applied Basic Research Program of Yunnan Province (2015FA024); ESI Discipline Promotion Program of Yunnan Agricultural University (2019YNAUESIMS01).

Conflict of interest

There is no conflict of interest.

References

1. Hamada, Y.M. (2020) *Agribusiness as the Future of Agriculture: The Sugarcane Industry under Climate Change in the*

- Southeast Mediterranean*. Apple Academic Press, New York. [10.1201/9780429321702](https://doi.org/10.1201/9780429321702).
2. Yu, H. and Zhao, N.X. (2004) Geographical distribution of Saccharinae (Gramineae). *J. Trop. Subtrop. Bot.*, **12**, 29–35. [10.3969/j.issn.1005-3395.2004.01.005](https://doi.org/10.3969/j.issn.1005-3395.2004.01.005).
 3. Wang, Z.T., Pan, Y.B., Luo, J. *et al.* (2020) SSR-based genetic identity of sugarcane clones and its potential application in breeding and variety extension. *Sugar Tech.*, **22**, 1–12. [10.1007/s12355-019-00788-9](https://doi.org/10.1007/s12355-019-00788-9).
 4. Ravinder, K., Appunu, C., Durai, A.A. *et al.* (2015) Genetic confirmation and field performance comparison for yield and quality among advanced generations of *Erianthus arundinaceus*, *E. bengalense* and *Saccharum spontaneum* cyto-nuclear genome introgressed sugarcane intergeneric hybrids. *Sugar Tech.*, **17**, 379–385. [10.1007/s12355-014-0333-2](https://doi.org/10.1007/s12355-014-0333-2).
 5. Feng, X.P., He, Y.Q., Fang, J. *et al.* (2015) Comparison of the growth and biomass production of *Miscanthus sinensis*, *Miscanthus floridulus* and *Saccharum arundinaceum*. *Span. J. Agric. Res.*, **23**, e0703–e0703. [10.5424/sjar/2015133-7262](https://doi.org/10.5424/sjar/2015133-7262).
 6. He, S.C., Yang, Q.H., Xiao, F.H. *et al.* (1994) Investigations and collections of wild germplasm plants related to sugarcane in China. *Sugarcane*, **1**, 11–17.
 7. Ying, X.M. (2013) Analysis of genetic diversity for germplasm resources of *Erianthus fulvus*. *Chin. Acad. Agric. Sci.*, **12**, 6–10.
 8. Tian, C.Y., Wang, X.H., Li, F.S. *et al.* (2015) Morphological diversity of wild sugarcane species *Erianthus fulvus*. *Chin. Agric. Sci. Bulletin*, **31**, 97–102. [10.11924/j.issn.1000-6850.casb14120014](https://doi.org/10.11924/j.issn.1000-6850.casb14120014).
 9. Wang, X.H., Yang, Q.H., Li, F.S. *et al.* (2010) Characterization of the chromosomal transmission of intergeneric hybrids of *Saccharum* spp. and *Erianthus fulvus* by genomic in situ hybridization. *Crop Sci.*, **50**, 1642–1648. [10.2135/cropsci2010.01.0004](https://doi.org/10.2135/cropsci2010.01.0004).
 10. Zhao, J.L., Wang, X.H., Li, F.S. *et al.* (2014) Genetic diversity of wild sugarcane species *Erianthus fulvus* based on SSR. *Mol. Plant Breed.*, **12**, 323–331. [10.13271/j.mpb.012.000323](https://doi.org/10.13271/j.mpb.012.000323).
 11. Li, F.S., Lin, W.F. and He, S.C. (2004) Identification of intergeneric F1 hybrids between *S. officinarum* and *E. fulvus*. *Chin. J. Tropical Crops*, **25**, 102–105. [10.3969/j.issn.1000-2561.2004.04.020](https://doi.org/10.3969/j.issn.1000-2561.2004.04.020).
 12. Meng, Y., Chen, S.Y., Xu, R. *et al.* (2018) Digital gene expression profiles of wild species of *Erianthus fulvus* response to low temperature stress. *Mol. Plant Breed.*, **16**, 3877–3886. [10.13271/j.mpb.016.003877](https://doi.org/10.13271/j.mpb.016.003877).
 13. Qian, Z.F., Meng, Y., Xu, R. *et al.* (2021) Cloning and expression analysis of ErDREB1A gene in the wild species of *Erianthus fulvus*. *Genomics Appl. Biol.*, **40**, 827–834. [10.13417/j.gab.040.000827](https://doi.org/10.13417/j.gab.040.000827).
 14. Qian, Z.F., Gu, S.J., Zhao, X.T. *et al.* (2022) Transcriptome-wide identification and cold stress expression analysis of CML gene family in *Erianthus fulvus*. *J. Agric. Biotechnol.*, 1–11. [10.3969/j.issn.1674-7968.2022.05.006](https://doi.org/10.3969/j.issn.1674-7968.2022.05.006).
 15. Ma, H., Cao, Z.Q., Xu, R. *et al.* (2021) Cloning and expression analysis of EfNAC44 gene in *Erianthus fulvus*. *Mol. Plant Breed.*, **19**, 2550–2556. [10.13271/j.mpb.019.002550](https://doi.org/10.13271/j.mpb.019.002550).
 16. Shen, X.Y., Gu, S.J., Xie, L.Y. *et al.* (2020) Cloning and expression analysis of EfGRAS gene of relative sugarcane wild species *Erianthus fluvus*. *Chin. J. Tropical Crops*, **41**, 2113–2119. [10.3969/j.issn.1000-2561.2020.10.017](https://doi.org/10.3969/j.issn.1000-2561.2020.10.017).
 17. Zhang, J.S., Zhang, X.T., Tang, H.B. *et al.* (2018) Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. *Nat. Genet.*, **50**, 1565–1573. [10.1038/s41588-018-0237-2](https://doi.org/10.1038/s41588-018-0237-2).
 18. Pruitt, K.D., Brown, G.R., Hiatt, S.M. *et al.* (2013) RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res.*, **42**, D756–63. [10.1093/nar/gkt1114](https://doi.org/10.1093/nar/gkt1114).
 19. Gary, B. (1999) Tandem Repeats Finder: a program to analyze DNA sequences. *Nucleic Acids Res.*, **27**, 573–580. [10.1093/nar/27.2.573](https://doi.org/10.1093/nar/27.2.573).
 20. Parra, G., Blanco, E. and Guigó, R. (2000) GeneID in drosophila. *Genome Res.*, **10**, 511–515. [10.1101/gr.10.4.511](https://doi.org/10.1101/gr.10.4.511).
 21. Majoros, W.H., Pertea, M. and Salzberg, S.L. (2004) TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics*, **20**, 2878–2879. [10.1093/bioinformatics/bth315](https://doi.org/10.1093/bioinformatics/bth315).
 22. Korf, I. (2004) Gene finding in novel genomes. *BMC Bioinform.*, **5**, 59. [10.1186/1471-2105-5-59](https://doi.org/10.1186/1471-2105-5-59).
 23. Burge, C. and Karlin, S. (1997) Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.*, **268**, 78–94. [10.1006/jmbi.1997.0951](https://doi.org/10.1006/jmbi.1997.0951).
 24. Grabherr, M.G., Haas, B.J., Yassour, M. *et al.* (2011) Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.*, **29**, 644–652. [10.1038/nbt.1883](https://doi.org/10.1038/nbt.1883).
 25. Haas, B.J., Salzberg, S.L., Wei, Z. *et al.* (2008) Automated eukaryotic gene structure annotation using EvidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.*, **9**, R7. [10.1186/gb-2008-9-1-r7](https://doi.org/10.1186/gb-2008-9-1-r7).
 26. Boeckmann, B. (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.*, **31**, 365–370. [10.1093/nar/gkg095](https://doi.org/10.1093/nar/gkg095).
 27. Minoru, K., Susumu, G., Shuichi, K. *et al.* (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res.*, **32**, D277–280. [10.1093/nar/gkh063](https://doi.org/10.1093/nar/gkh063).
 28. Zdobnov, E.M. and Rolf, A. (2001) InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics*, **17**, 847–848. [10.1093/bioinformatics/17.9.847](https://doi.org/10.1093/bioinformatics/17.9.847).
 29. Bauer, S., Grossmann, S., Vingron, M. *et al.* (2008) Ontologizer 2.0—a multifunctional tool for GO term enrichment analysis and data exploration. *Bioinformatics*, **24**, 1650–1651. [10.1093/bioinformatics/btn250](https://doi.org/10.1093/bioinformatics/btn250).
 30. Kim, D., Paggi, J.M., Park, C. *et al.* (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.*, **37**, 709–915. [10.1038/s41587-019-0201-4](https://doi.org/10.1038/s41587-019-0201-4).
 31. Pertea, M., Pertea, G.M., Antonescu, C.M. *et al.* (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.*, **33**, 290–295. [10.1038/nbt.3122](https://doi.org/10.1038/nbt.3122).
 32. Brouard, J.S., Schenkel, F., Marete, A. *et al.* (2019) The GATK joint genotyping workflow is appropriate for calling variants in RNA-seq experiments. *J. Anim. Sci. Biotechnol.*, **10**, 1–6. [10.1186/s40104-019-0359-0](https://doi.org/10.1186/s40104-019-0359-0).
 33. Areeya, D., Licht, T., Penpicha, T. *et al.* (2020) An optimized genomic VCF workflow for precise identification of *Mycobacterium tuberculosis* cluster from cross-platform whole genome sequencing data. *Infect. Genet. Evol.*, **79**, 104152. [10.1016/j.meegid.2019.104152](https://doi.org/10.1016/j.meegid.2019.104152).
 34. Ma, J., Aagaard, K., Gevers, D. *et al.* (2013) 585: distinct vaginal microbiome profiles are associated with host mitochondrial DNA (mtDNA) haplogroup and SNP variants. *Am. J. Obstet. Gynecol.*, **208**, S250–S250. [10.1016/j.ajog.2012.10.751](https://doi.org/10.1016/j.ajog.2012.10.751).
 35. Wang, K., Li, M. and Hakonarson, H. (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.*, **38**, e164–e164. [10.1093/nar/gkaq603](https://doi.org/10.1093/nar/gkaq603).
 36. Zheng, X., Gogarten, S.M., Lawrence, M. *et al.* (2017) SeqArray—a storage-efficient high-performance data format for WGS variant calls. *Bioinformatics*, **33**, 2251–2257. [10.1093/bioinformatics/btx145](https://doi.org/10.1093/bioinformatics/btx145).
 37. Yang, J., Lee, S.H., Goddard, M.E. *et al.* (2011) GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.*, **88**, 76–82. [10.1016/j.ajhg.2010.11.011](https://doi.org/10.1016/j.ajhg.2010.11.011).
 38. Alexander, D.H., Novembre, J. and Lange, K. (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.*, **19**, 1655–1664. [10.1101/gr.094052.109](https://doi.org/10.1101/gr.094052.109).
 39. Ronquist, F. and Huelsenbeck, J.P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574. [10.1093/bioinformatics/btg180](https://doi.org/10.1093/bioinformatics/btg180).
 40. Skinner, M.E., Uzilov, A.V., Stein, L.D. *et al.* (2009) JBrowse: a next-generation genome browser. *Genome Res.*, **19**, 1630–1638. [10.1101/gr.094607.109](https://doi.org/10.1101/gr.094607.109).
 41. Deng, W.J., Nickle, D.C., Gerald, L.H. *et al.* (2007) ViroBLAST: a stand-alone BLAST web server for flexible queries of multiple databases and user's datasets. *Bioinformatics*, **23**, 2334–2336. [10.1093/bioinformatics/btm331](https://doi.org/10.1093/bioinformatics/btm331).