

ScDB: A Resource for Sugarcane Genomics and Molecular Biology

Boddu Sangavi*, Bhagyasri Majhi, Soumyanetra Saha and Vattsal Rana

Department of Genetics and Plant Breeding, N. M. College of Agriculture, Navsari Agricultural University, Navsari, Gujarat-396450

Corresponding Author:

Introduction

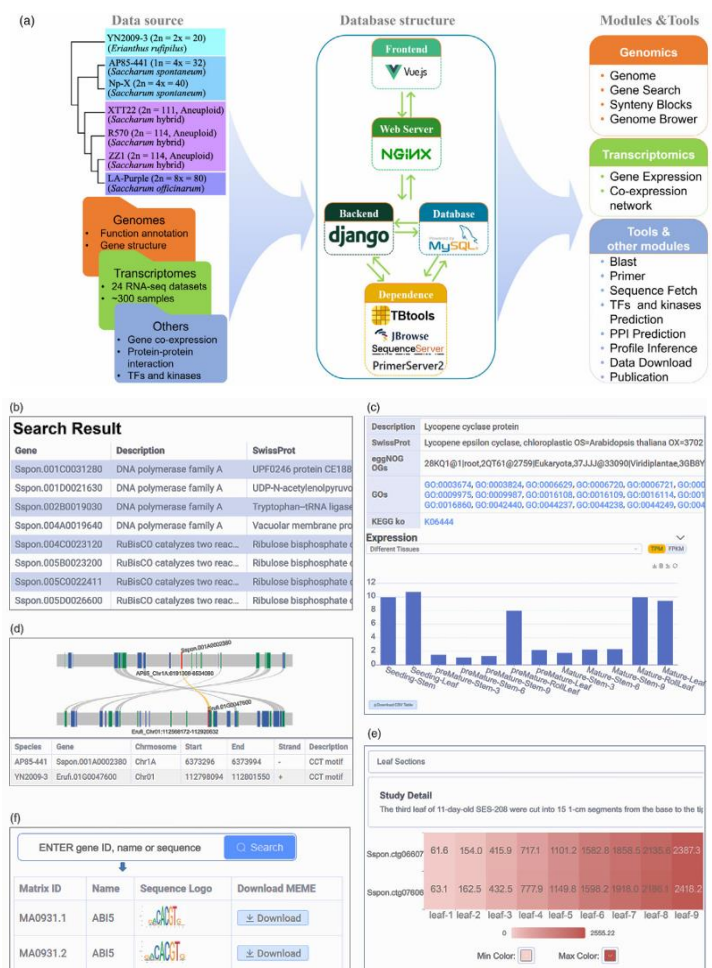
The most significant sugar crop in the world, sugarcane is used as the main raw material to make both sugar and biofuels. A modern cultivar of sugarcane that was created by intentional interspecific hybridization of *Saccharum spontaneum* and *Saccharum officinarum*. The development of superior sugarcane varieties depends on the use of wild resources, and the omics and genomic investigations of these resources offer important new understandings of their molecular mechanisms. But historically, researchers have faced difficulties due of the sugarcane genome's intricacy. In earlier research, we spearheaded the genome assembly of a haploid *S. spontaneum* AP85-441 (Zhang *et al.*, 2018) and developed the method for addressing a complicated autopolyploid at the allele level. Next, we mapped the chromosomal evolution in *S. spontaneum* and tracked the origins of *Saccharum* Np-X (Zhang *et al.*, 2022).

Furthermore, the website was able to put together a complete, gap-free diploid genome of *Erianthusrufipilus* YN2009-3, which provided insight into the evolutionary genomic footprints of the highly polyploid *Saccharum* (Wang *et al.*, 2023). In the meantime, we are pleased to introduce the *Saccharum* hybrid XTT22 genome, which is regarded as the most important development in sugarcane research. Additionally, other groups have conducted genome studies on sugarcane in a similar manner. A. D'Hont's team and Muqing Zhang's team released the genomes of modern sugarcane R570 and ZZ1 this year, respectively (Bao *et al.*, 2024; Healey *et al.*, 2024).

Overview of Sugarcane Database

Building on this framework, this is the first easily navigable multi-omics database for six 3386 *Saccharum* species (AP85-441, Np-X, LA-Purple, XTT22, R570, ZZ1) and an *Erianthusrufipilus* (YN2009-3): ScDB (Saccharum genomic database, <https://sugarcane.gxu.edu.cn/scdb>). There are now 38.91 Gb of genomic assembly sequences in ScDB, which include 1,366,608 genes. ScDB also has 24 transcriptome projects with over 2.5 TB of data and more than 300 sugarcane samples. Additionally, to

make using ScDB easier, 12 online features that users use frequently have been created. These include "Gene Search," "Orthologous Gene Search," "Synteny Block," "Genome Browser," "Gene Expression," "Co-expression Network," "Blast," "Primer," "Sequence Fetch," "Transcription Factors," "Protein Interaction Network," and "Profile Inference" (Figure 1a).



ScDB is made up of a core database, a backend application server, a frontend web interface, and a number of analytical and visualization tools. The six primary sections of the database are "Home," "Genomics," "Transcriptomics," "Tools," "Download," and "Publication." The homepage includes connections to a number of tools, descriptions of *Saccharum* species and *Erianthusrufipilus*, an overview of ScDB, and an advanced search engine. Users can search by gene ID, gene name, GO number, and KEGG

number using the advanced search function (Figure 1b).

Figure 1: Overview of ScDB and its functions in multi-omics analysis. (a) The phylogenetic relationships and data sources of all existing species in ScDB, the construction process and the modules and tools included. (b) Advanced search of the home page. (c) Part of the gene details page: gene function annotations and expression data for different studies. (d) The gene synteny blocks are obtained using the synteny blocks search function. The results are presented as a synteny diagram and table. (e) Gene expression heatmaps allow users to select different studies and Expression Units (TPM or FPKM) and customize colour schemes. (f) The Profile Inference tool enables users to match known motifs by gene ID, gene name and amino acid sequence, and meme files are provided for download. This pictorial representation is taken from Chen, *Set al.*, 2024

The "Genome" module contains functions for "Genome," "Gene Search," "Synteny Blocks," and "Genome Browser." The "Genome" feature displays sequenced *Saccharum* species and *Erianthusrufipilus* along with information about their geographic distribution and evolutionary relationships. Users can view detailed genomic information and images for each variety, as well as structural annotations for each chromosome. The "Gene Search" feature allows users to look up multiple genes using either gene IDs or specific chromosome regions. The "Search by Range" option incorporates a chromosome selection tool, which makes it simpler for users who are not as familiar with the genome to navigate.

In the "Gene Expression" section, users can access expression data for a variety of genes, choose their preferred studies, choose the expression units (either Transcripts Per Million or Fragments Per Kilobase Million), and alter the heatmap's color scheme to suit their tastes. The "Transcriptomics module" provides search and visualization capabilities for gene expression (Figure 1e) and co-expression gene networks.

Functions for "Blast," "Primer," "Sequence Fetch," "Transcription Factors," "Protein Interaction Network," and "Profile Inference" are all included in the "Tools" module. The "Blast" tool uses several data sets to do homology searches. The primer design tool is called "Primer." Chromosome sequences can be

extracted from a designated region using "Sequence Fetch." We identified the transcription factor and kinase families of *Saccharum* species and *Erianthusrufipilus* using the iTAK (Zheng *et al.*, 2016) software in the "Transcription Factors" section. Users can click on the name of any transcription factor or kinase family to view a list of all the genes that are members of that family and search for the gene family that the gene belongs to.

Users can use gene IDs to find protein interaction networks for certain genes in "Protein Interaction Network." The findings are shown in a table that can be saved as an SVG graphic or as an interactive network diagram that can be saved as a CSV file. By matching gene ID, gene name, and protein sequence in "Profile Inference," users can search the Jaspas database for motifs. They can also download meme format files that can be used to predict binding with upstream sequences retrieved from the gene information page. The "Download" module offers the ability to download chromosomal data and annotations.

Conclusion

In conclusion, ScDB includes transcriptome data, genomic assemblies, and annotations for six *Saccharum* species as well as *Erianthusrufipilus*. ScDB also offers a number of practical modules for search, analysis, and visualization to improve the usability and effectiveness of data collection and analysis. To make sure that ScDB is a strong and long-lasting platform for collecting and analyzing sugarcane data, it will be expanded in the future with further sugarcane genome data, other omics data levels (proteomics, epigenetics, ncRNA, etc.), and additional data analysis tools.

References

- Bao, Y., Zhang, Q., Huang, J., Zhang, S., Yao, W., Yu, Z., Deng, Z. et al. (2024) A chromosomal-scale genome assembly of modern cultivated hybrid sugarcane provides insights into origination and evolution. *Nat. Commun.* 15, 3041.
- Chen, S., Feng, X., Zhang, Z., Hua, X., Zhang, Q., Chen, C., et al. (2024) ScDB: A comprehensive database dedicated to *Saccharum*, facilitating functional genomics and molecular biology studies in sugarcane. *Plant Biotechnology Journal*, pbi.14457.

Healey, A.L., Garsmeur, O., Lovell, J.T., Shengquiang, S., Sreedasyam, A., Jenkins, J., Plott, C.B. et al. (2024) The complex polyploid genome architecture of sugarcane. *Nature* 628, 804–810.

Wang, T., Wang, B., Hua, X., Tang, H., Zhang, Z., Gao, R., Qi, Y. et al. (2023) A complete gap-free diploid genome in *Saccharum* complex and the genomic footprints of evolution in the highly polyploid *Saccharum* genus. *Nature Plants* 9, 554–571.

Zhang, J., Zhang, X., Tang, H., Zhang, Q., Hua, X., Ma, X., Zhu, F. et al. (2018) Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. *Nat. Genet.* 50, 1565–1573.

Zhang, Q., Qi, Y., Pan, H., Tang, H., Wang, G., Hua, X., Wang, Y. et al. (2022) Genomic insights into the recent chromosome reduction of autopolyploid sugarcane *Saccharum spontaneum*. *Nat. Genet.* 54, 885–896.
