

TropCRD (Tropical Crop Resources Database): the multi-tropical crop variation information system

Authors

Jianjia Xiao, Hai Liu, Yangyang Tian, Pengliang An, Biao Liu, ..., Meiling Zou*

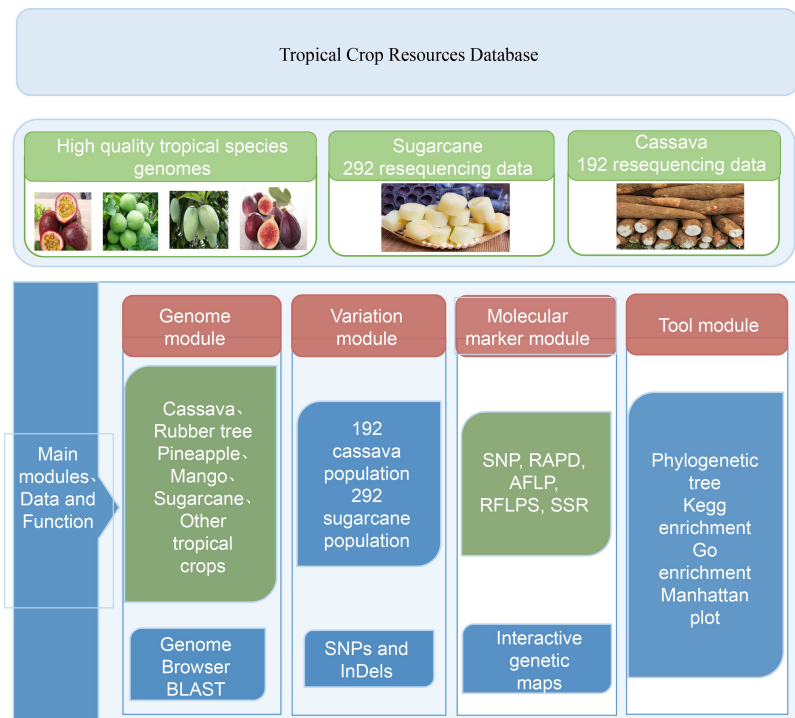
Correspondence

mlzou@hainanu.edu.cn

In Brief

A tropical species database was constructed, which included data storage, species profile, molecular identification, genetic variation and functional analysis.

Graphical abstract



Highlights

- The research materials are tropical crop varieties with high economic value and wide sources.
- Use multiple tools and software to build a website framework.
- It has multiple functions such as sequence alignment, molecular labeling and data visualization.
- Provide a more comprehensive and accurate data base for relevant scientific research, and promote the in-depth development of tropical crop research.

Citation: Xiao J, Liu H, Tian Y, An P, Liu B, et al. 2023. TropCRD (Tropical Crop Resources Database): the multi-tropical crop variation information system. *Tropical Plants* 2:9 <https://doi.org/10.48130/TP-2023-0009>

TropCRD (Tropical Crop Resources Database): the multi-tropical crop variation information system

Jianjia Xiao^{1,2#}, Hai Liu^{3#}, Yangyang Tian^{1,2#}, Pengliang An^{1,2}, Biao Liu^{1,2}, Wanfeng Ma^{1,2}, Ke Deng¹, Xiaorui Dong^{1,2}, Miaohua He^{1,2}, Fen Yu^{1,2} and Meiling Zou^{1,2*}

¹ Sanya Nanfan Research Institute of Hainan University, Hainan Yazhou Bay Seed Laboratory, Sanya 572025, China

² College of Tropical Crops, Hainan University, Haikou 570288, China

³ National Salt and Alkali tolerant Rice Technology Innovation Center, Sanya 572024, China

⁴ Academy of Agriculture and Forestry Sciences, Qinghai University, Xining 810016, Qinghai, China

These authors contributed equally: Jianjia Xiao, Hai Liu, Yangyang Tian

* Corresponding author, E-mail: mlzou@hainanu.edu.cn

Abstract

TropCRD (www.tropical-resources.org.cn) is organized on a crop basis with current *Manihot esculenta* Crantz, *Hevea brasiliensis*, *Saccharum officinarum* L., *Ananas comosus*, *Mangifera indica* L. and other tropical crop species. TropCRD is currently divided into four modules: genome module, variation module, molecular marker module and tools module. These data and tools are available for further research by geneticists and tropical crop breeders, including the application of marker-assisted selection techniques and genomic selective breeding to tropical crop breeding. It provides important support for tropical crop breeding research, and also provides valuable resources for obtaining and analyzing the genome sequence data, genetic variation and functional annotation of tropical crops.

Citation: Xiao J, Liu H, Tian Y, An P, Liu B, et al. 2023. TropCRD (Tropical Crop Resources Database): the multi-tropical crop variation information system. *Tropical Plants* 2:9 <https://doi.org/10.48130/TP-2023-0009>

Introduction

Tropical crops mainly include industrial raw materials such as *Hevea brasiliensis*, *Manihot esculenta* Crantz and *Elaeis guineensis* Jacq., tropical fruits such as *Passiflora caerulea* L., *Litchi chinensis* Sonn., *Mangifera indica* L. and *Saccharum officinarum* L. and spice drinks such as *Coffea arabica* L., *Cinnamomum cassia* (L.) and *Illiciumverum* Hook. f., which are important strategic resources and daily consumer goods. At present, remarkable achievements have been made in the development of the tropical crop industry. Advantageous production areas for major crops have taken shape and relevant industrial systems have been continuously improved, playing an important role in ensuring national defense and economic security, meeting people's living needs, developing non-food biomass energy and increasing farmers' income.

In recent years, the genomes of a number of tropical species have been uncovered, for example, *Passiflora caerulea* L.^[1], *Canna indica* 'Edulis' and *Canna indica* L.^[2], *Litchi chinensis* Sonn.^[3], *Cocos nucifera* L.^[4], etc. Genome and related research results provide valuable resources for the breeding and improvement of many tropical species. At the same time, with the output of a large number of multi-omics data, specific databases for different families have been developed, such as SapBase and ArecaceaeMDB^[5,6]. At present, there is a shortage of databases for tropical crops. TropGeneDB^[7] is currently the only database related to tropical crops; however, it lacks important crops such as *Manihot esculenta* Crantz, *Ananas comosus*, and *Mangifera indica* L. There is also a shortage of analytical and visual tools for breeders to utilize.

Materials and methods

Materials

A total of 192 *Manihot esculenta* Crantz population resequencing data and 292 *Saccharum officinarum* L. population resequencing data were collected, and a number of high quality tropical species genomes and annotated data were collected from public databases as the basis for the establishment of this database.

Methods

The software Linux MongoDB Nginx and PHP development environment HTML5 CSS JavaScript JBrowse 1.16.11 SequenceServer R language R packages such as ggplot2, ggtree and treeio.

The classic Linux, MongoDB, Nginx and PHP development environments are deployed on the Ubuntu 20.04 operating system. The MongoDB database management system manages and stores all variation data (SNPs and InDels) and genotype data. The web front-end uses HTML5, CSS and JavaScript. HTML5 is mainly used for the layout of the page, CSS is used to control the style of the layout of the page, and JavaScript is responsible for implementing some interactive functions. Visualize the genome using JBrowse 1.16.11. SequenceServer is used for BLAST comparison. The rest of the data presentation and visualization is done in R language. It has been tested in browsers such as Firefox, Google Chrome and Internet Explorer.

Results

Current status of genetic, molecular and phenotypic studies of tropical plants

We have created TropCRD to provide data and technical support for tropical crop breeding. TropCRD was created to store genetic, molecular and phenotypic data on tropical crop species, with the variation information, molecular markers, quantitative trait loci, genetic maps, genetic diversity and phenotypic diversity studies data. At present, the database has included a total of five tropical species genomes, 12,850 QTL markers, 484 germplasm resources of tropical species, 75,396 SNPs and InDels, and 220,090 genes annotated. The genetic linkage map of the three species was constructed (Table 1).

The TropCRD is based on tropical crops and committed to building a platform serving tropical crop breeding, displaying various types of data from multiple dimensions, providing a more comprehensive display platform for data display of omics research, and providing a variety of new ideas and strategies for related in-depth research (Fig. 1).

Database software

TropCRD uses the classic Linux, MongoDB, Nginx and PHP development environment. It is deployed on the Ubuntu 20.04 operating system. All variation data (SNPs and InDels) and genotype data in the database are managed and stored by the MongoDB database management system. Its web front end

Table 1. Overview of genetic, molecular and phenotypic data in TropCRD.

| Species | QTLs | Germplasms | SNPs | Genes | Genetic maps |
|---------------------------------|-------|------------|--------|---------|--------------|
| <i>Manihot esculenta</i> Crantz | 4,501 | 192 | 30,427 | 33,030 | 1 |
| <i>Saccharum officinarum</i> L. | — | 292 | 44,969 | 112,787 | — |
| <i>Ananas comosus</i> | — | — | — | 31,585 | — |
| <i>Mangifera indica</i> L. | 1,409 | — | — | — | 1 |
| <i>Hevea brasiliensis</i> | 6,940 | — | — | 42,688 | 1 |

uses HTML5, CSS and JavaScript. HTML5 is mainly used for page layout, CSS is used to control the style of page layout, and JavaScript is responsible for the implementation of some interactive functions. We used JBrowse 1.16.11 to visualize the genome. SequenceServer is used for BLAST comparison. The Echarts visualization library was used to present the genetic map, while the rest of the data presentation and visualization were handled by R language, which has been tested in browsers including Firefox, Google Chrome and Internet Explorer.

Genome module

The Genome module provides users with the genomes of tropical species such as *Manihot esculenta* Crantz, *Hevea brasiliensis*, *Saccharum officinarum* L., *Ananas comosus*, *Mangifera indica* L., *Jatropha curcas* L. and *Pennisetum sinese* Roxb. The module adopts JBrowse software. Users can view genome sequence, gene, transcript sequence information and view annotation files, information about variation data, and complete visualizations in the Genome browser. For example, the user selects chromosome 3: 17,775,000 ~ 17,785,000 bp to browse the genome region (Fig. 2) and clicks 'Manes.OG101200' to display the type, location, length, and specific sequence information of the gene (Fig. 3). Support sequence alignment, can select multiple databases for comparison and analysis at the same time, compare the homology information between the output sequence, help judge the source of the input sequence or the evolutionary relationship with the known sequence (Fig. 4).

Molecular marker module

The molecular marker module included 37,852 markers of *Manihot esculenta* Crantz, *Hevea brasiliensis*, *Ananas comosus* and *Mangifera indica* L., including SNP, SSR, EST SSR, SCAR, CAPS, ISSR, RAPD, AFLP and RFLPS. Based on published articles^[8–17], interactive genetic maps are constructed and markers are arranged on each chromosome. Users can select the genetic map region they are interested in according to their own needs, obtain the marker ID, genetic map location,

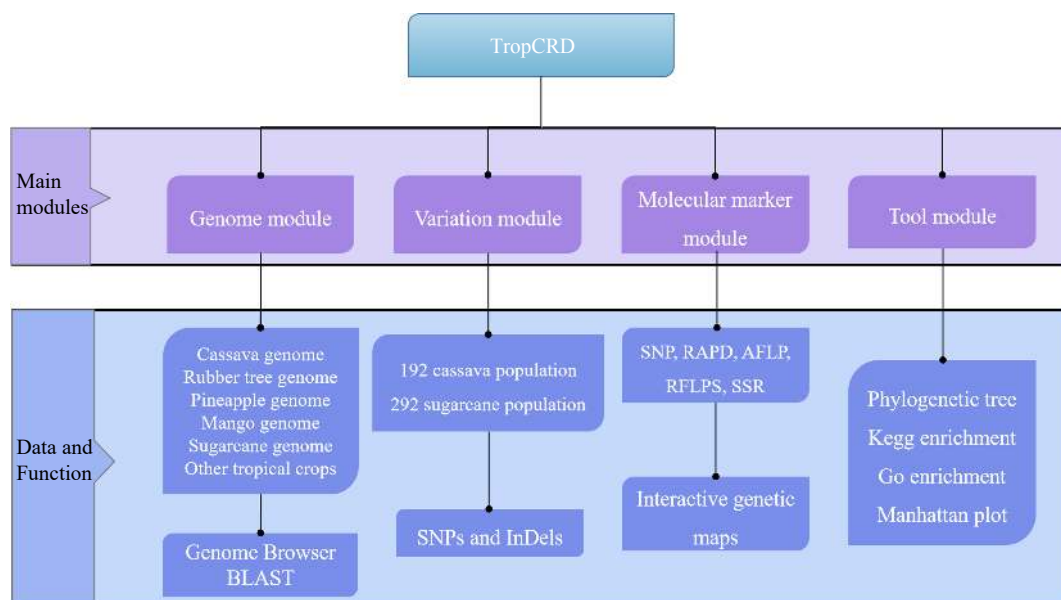


Fig. 1 Main data modules of tropical crop database.

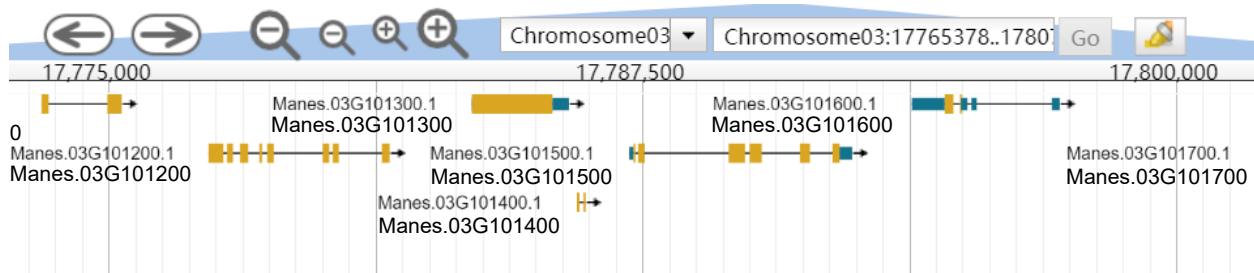


Fig. 2 The genome region of chromosome 3: 17,775,000 ~ 17,800,000 bp.

gene Manes.03G101200

Primary Data

| | |
|----------|--------------------------------------------|
| Name | Manes.03G101200 |
| Type | gene |
| Position | Chromosome03:17777349..17781579 (+ strand) |
| Length | 4,231 bp |

Attributes

| | |
|--------------------|---------------------------|
| Ancestoridentifier | cassava4.1_033544m.g.v4.1 |
| Id | Manes.03G101200.v6.1 |
| Seq_id | Chromosome03 |
| Source | phytozomev10 |

Region sequence

```
>Chromosome03 Chromosome03:17777349..17781579 (+ strand)
class=gene length=4231
ATGGCGTTGAAGCGGAGTCACACGACGGAGCTAGGCTGCATAGCTTGCGAGGAAGTGAAGCGA
CCTCGGTGCGGAAAGGAAGGGTGGCTGGTCGAAAACCCCTAATATTCTCTGCGCTCTTGACA
CCCATCAATTGCCCTCGCAAATCGTTCTTTCATCTTAGTTACGGGGTGGGATGATGGGCCT
CGCCGAGGATCCGACCCGATTTGTCTCCCATAGAGGCCGAGTTCATAACCTCTCTGGAATG
GCTGGTCTTTGACGAGATTCGAGTGATTGCGGTGGGCACCTCTTGTGGGTATTCTTGTTTT
ATTCTTTGGATGGCAATGGTTTCATAGACAGGTTAGTGAGTTGTTGGTTTGCCTTCCATG
ATTTAATTAATTTGAATAAGTATTTTCAGAAAAAATATCCCGTTATTCTGCTGCAGATG
GTGTATCCTGGGCGGATTCGAAAATAAGAGTCCGTGGAAGTAAAGAAAGATTTGACCCATCA
AACATCATCTTCAGAGGAGATTTCTATTGTTATTCGTGGTGTAATTGCTGTTTTGATGGCT
```

Fig. 3 Sequence information of the gene 'Manes.03G101200'.

genome location and other information of all markers in the region (Fig. 5).

Variation module

At present, in the tropical fruit genome database variation module, resequencing variation data of multiple species populations are included, and users can choose the species they are interested in. This module has realized the retrieval function of *Manihot esculenta* Crantz and *Saccharum officinarum* L., including 192 *Manihot esculenta* Crantz materials and 292 *Saccharum officinarum* L. samples with resequencing data. A total of 75,396 SNP and InDel sites were obtained through the processing of resequencing data. Among them, *Saccharum officinarum* L. had 44,969 sites and *Manihot esculenta* Crantz had 30,427 sites. Users can retrieve SNPs sites by gene ID, input gene ID, select upstream and downstream ranges and variation types, and variation information will be displayed in the form of charts (Fig. 6).

Tools module

The tools module includes phylogenetic tree drawing, GO function enrichment analysis, Kegg enrichment analysis, and Manhattan map. The tools are all drawn in R language on the server background. R packages such as ggplot2, ggtree, treeio were used to complete data processing and image rendering in the background, providing users with the most reliable recognition results. Users click the corresponding function in 'Tools', select the species, enter the ID of the gene to be mapped in the blank below the species, and click the 'Search' button to obtain the functional relationship diagram of related genes and the phylogenetic relationship between species genes.

As shown in Fig. 7, Input Gene IDs into input gene IDs, 'Manes.01G000100', 'Manes.01G000200', 'Manes.01G000300', 'Manes.01G000400', 'Manes.01G000500', 'Manes.01G000600', 'Manes.01G000700', 'Manes.01G000800', 'Manes.01G000900', 'Manes.01G0001000', phylogenetic tree, GO functional enrich-

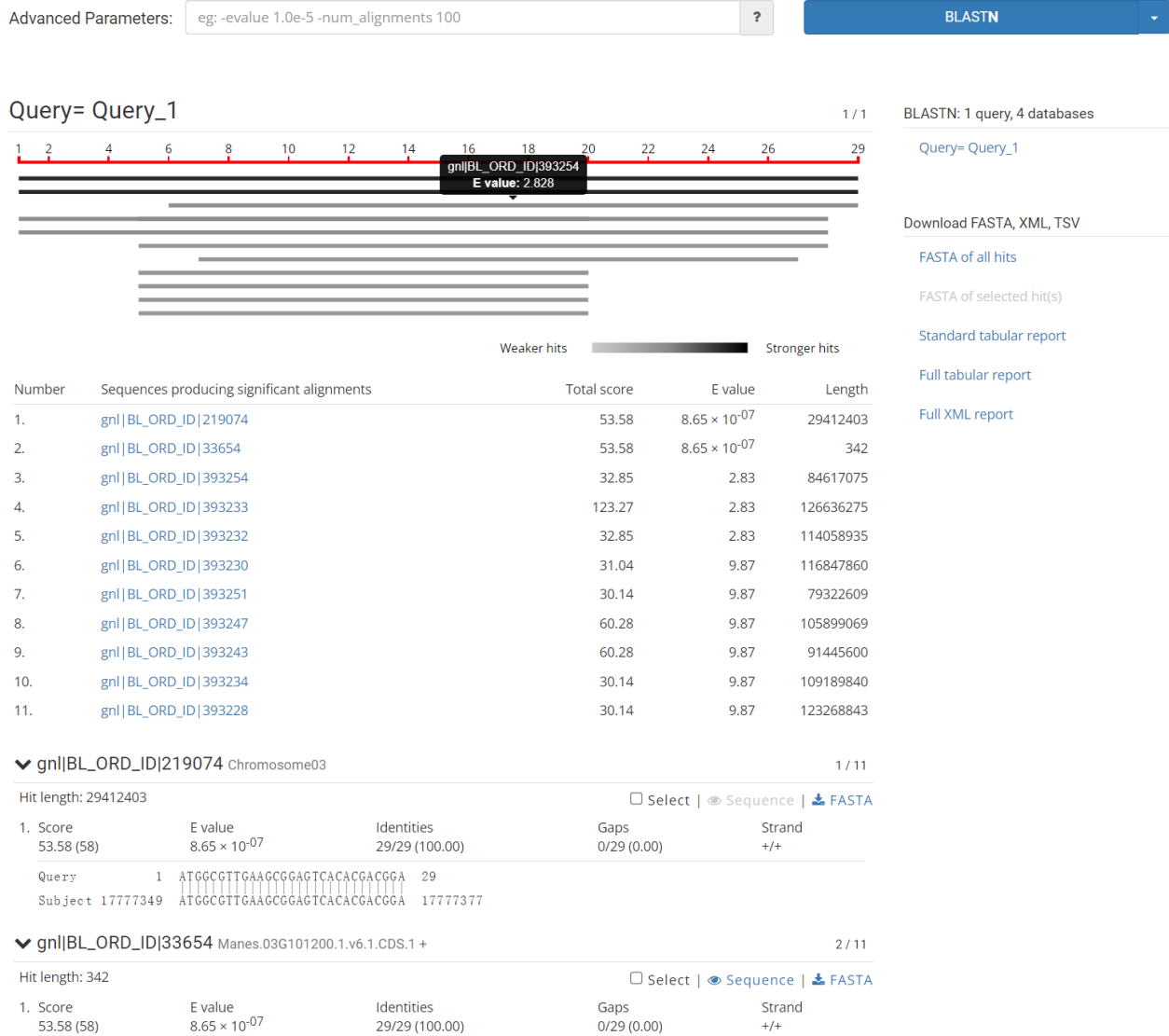


Fig. 4 BLAST sequence alignment interface.

ment bubble map and Kegg metabolic pathway enrichment bubble map were drawn (Fig. 7).

It can be seen that in the phylogenetic tree, Manes.01G0001000 and Manes.01G000900, Manes.01G000300 and Manes.01G000600, Manes.01G000700 and Manes.01G000800 have close homology (Fig. 8).

The top ten enrichment functions in the GO functional enrichment bubble chart are commitment complex, spliceosomal complex assembly, nuclear-transcribed mRNA catabolic process, deadenylation-dependent decay, nuclear-transcribed mRNA catabolic process, exonucleolytic, RNA splicing, via transesterification reactions, RNA splicing, via transesterification reactions with bulged adenosine as nucleophile, regulation of alternative mRNA splicing, via spliceosome, spliceosomal snRNP assembly, mRNA splicing, via spliceosome and P-body.

There are three pathways enriched in the bubble diagram of Kegg metabolic pathway, namely Spliceosome [BR:ko03041], Spliceosome, Systemic lupus erythematosus and Mitophagy - yeast (Fig. 9).

The Manhattan map is drawn using ShinyAIM^[18] software,

which can download the character data provided by ourselves or upload the file to generate an interactive Manhattan map. Users can click 'GWAS' in 'Tools' to enter the Manhattan map drawing interface, download the Manhattan map summary table of GWAS analysis results of tropical fruit traits provided in the database, or upload their own data in the specified format to complete the Manhattan map drawing. Click 'Browse' to upload the data to be analyzed. When the upload is completed, 'the upload complete' will be displayed. Users can check whether the table header is included or not.

After uploading, you can select the traits for analysis and set an appropriate $-\log_{10}$ p-value threshold for filtering based on your requirements. The generated Manhattan map can be interactive. The mouse on each point will display the SNP information of that point and you can freely drag the selection area, zoom in and out to move and save the picture. As shown in the figure, the $-\log_{10}$ p-value threshold was set to 5, and a total of 8 significantly related SNP sites were screened out, among which there were two significant sites on chromosome 4 (Fig. 10).

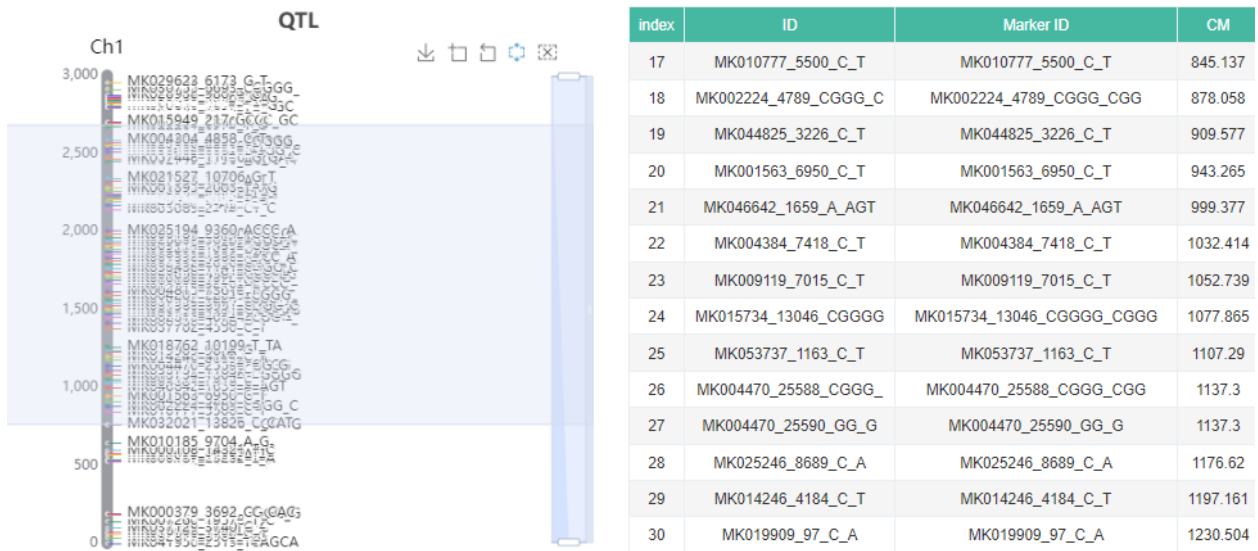


Fig. 5 QTL information of *Manihot esculenta Crantz* can be obtained by interactive genetic map.

Manihot esculenta Crantz gene regional variation information

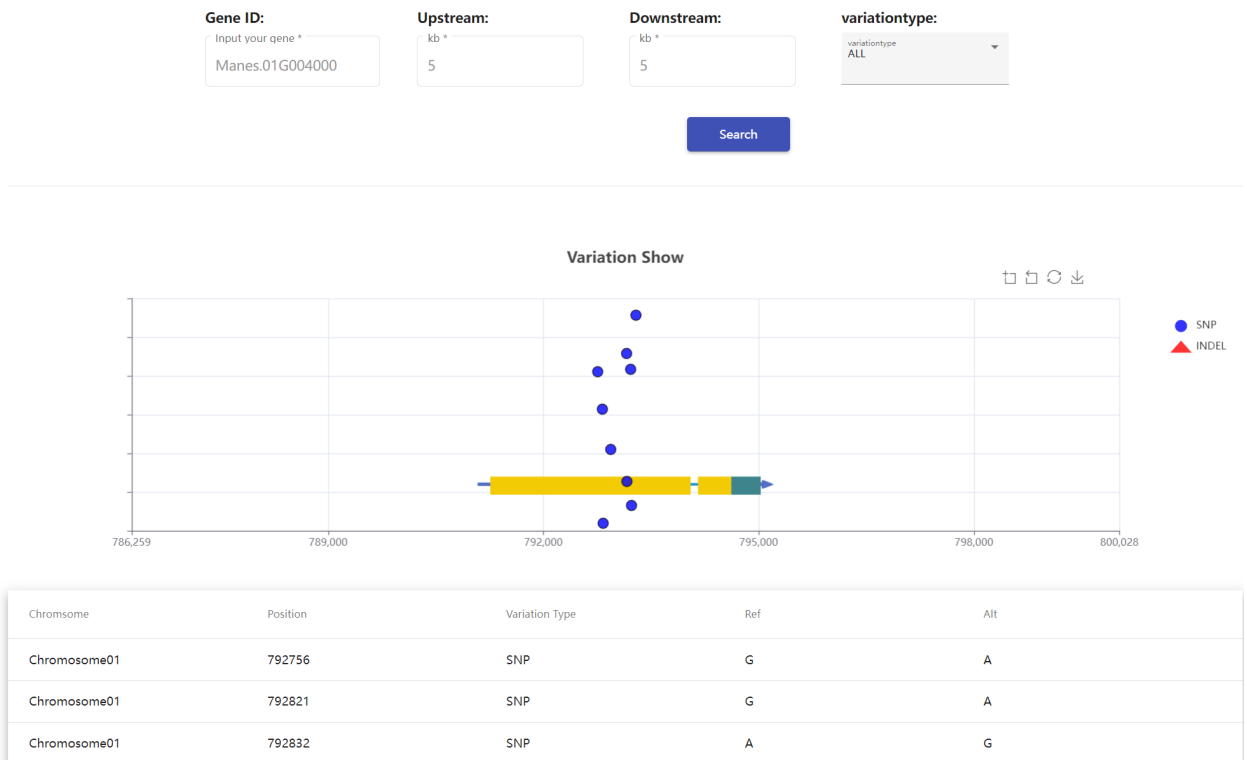


Fig. 6 Retrieval of SNP sites upstream and downstream of the gene 'Manes.01G004000'.

Data availability statement

The data used in this study are available from the corresponding author upon request

Discussion

Currently, in the absence of a comprehensive breeding program, it takes a long time to breed varieties and new varieties with marketable traits. In the early stages of breeding,

much time, space and resources are invested in selection and genetic progress after the initial crossing with the parent genotype. Many important agronomic traits of tropical crops are quantitative traits, which are controlled by a large number of microgenes. TropCRD promotes the breeding of tropical crops by constructing the genetic map of tropical crops and seeking QTL loci related to traits. On the other hand, genetic variation has been widely used in human diseases, identification of genetic loci related to important agronomic or economic traits

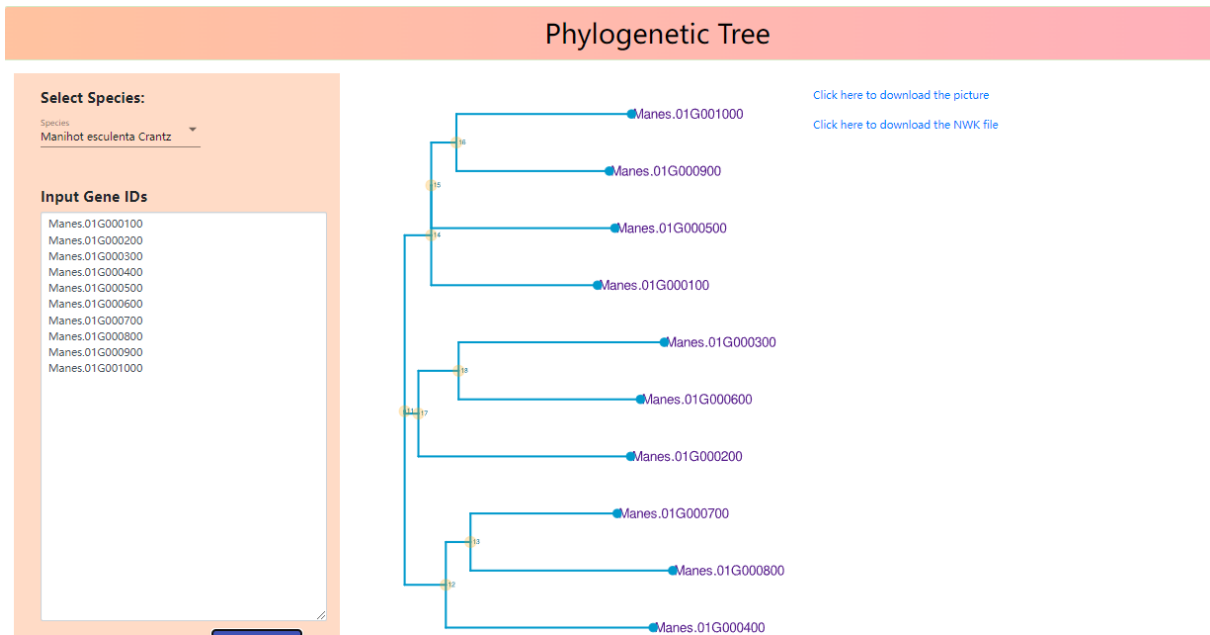


Fig. 7 Mapping phylogenetic trees based on genes of *Manihot esculenta* Crantz.

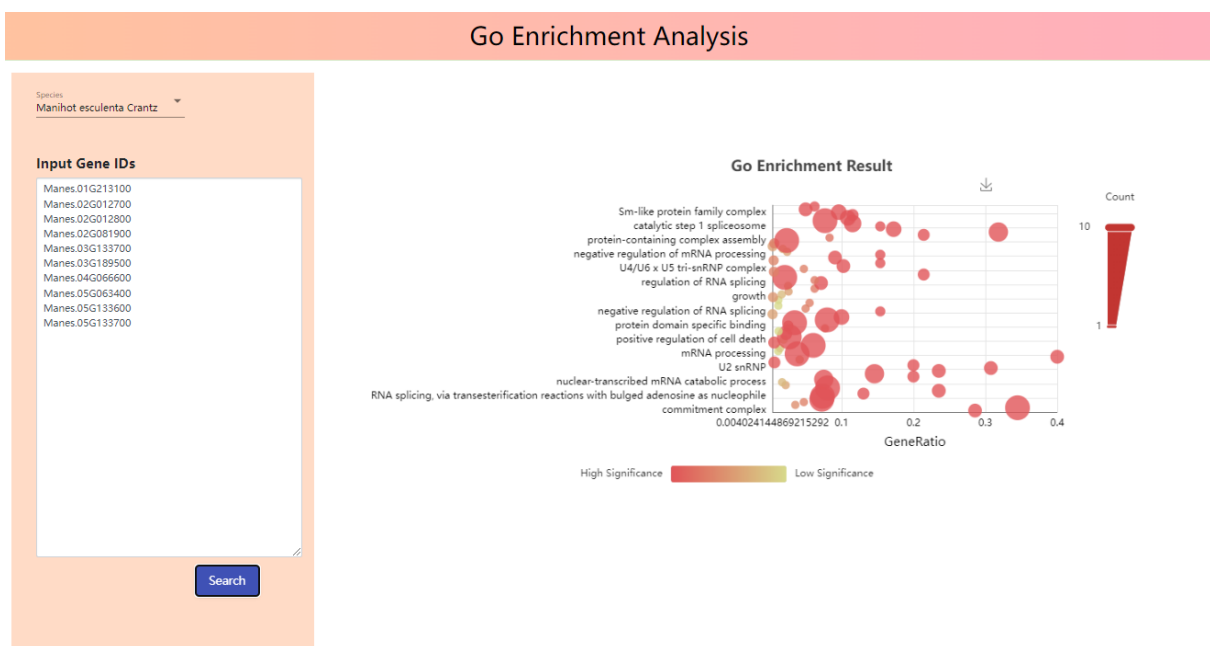


Fig. 8 GO enriched bubble map based on genes of *Manihot esculenta* Crantz.

in animals and plants, cloning of important functional genes, marker-assisted selection breeding, etc. TropCRD can be used in crop genetic breeding through the retrieval of genetic variation, including single nucleotide polymorphism (SNP) and small fragment insertion/deletion (InDel). TropCRD is important for the accurate identification of germplasm resources and the discovery of excellent alleles^[19].

The Tropical Crop Resource Database is the first database built based on tropical crops. Different from other databases, it takes crops growing in tropical regions as research objects and covers the genomic data of crops in tropical regions around the world. It provides valuable resources for researchers to obtain

and analyze genome sequence data, variation data and functional annotations of tropical crops. There are bioinformatics analysis tools, which provide molecular basis for the study of tropical crops. Secondly, the tropical crop resource database is the first database that combines Hyper-seq sequencing technology and builds tropical crop BLAST database based on Hyper-seq data.

In addition, the database still has a lot of room for improvement and expansion. The first is the integration of different types of data. At present, only genome sequence data, variation data and functional annotations are included in the database. However, other types of data, such as transcriptome

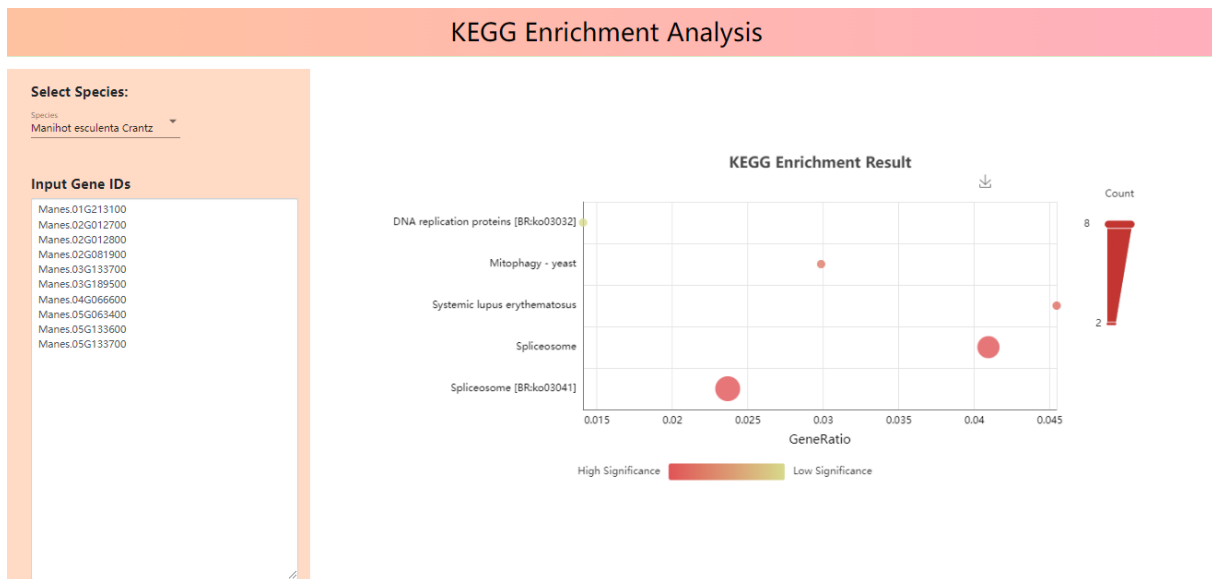


Fig. 9 Kegg enriched bubble map based on genes of *Manihot esculenta Crantz*.

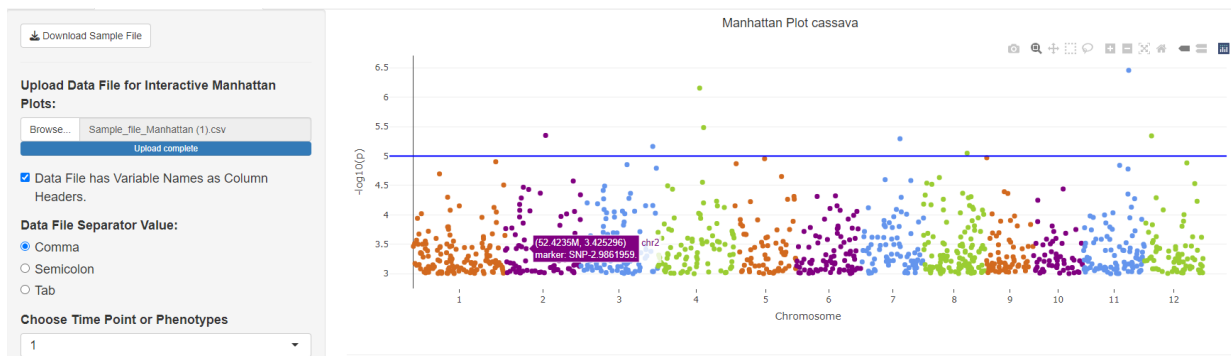


Fig. 10 Manhattan Plot according to the GWAS results.

data, epigenetic data, and metabolome data, can also provide valuable insights into the molecular mechanisms behind these important fruit traits. So integrating these different types of data into a tropical fruit DNA database could greatly improve their usefulness. Second is the standardization of data formats and annotations. There is a lack of standardization of the formats and annotations currently used in different tropical crop databases. This can make it difficult for researchers to compare data across different databases and can lead to inconsistencies in data analysis. Therefore, the establishment of standardized format and annotation of data can improve the interoperability of data and enhance the repeatability of research results. Finally, expanding the database to include more species could greatly improve their usefulness. The current database focuses on a few key species, such as *Carica papaya* L., *Musa nana* Lour., *Ananas comosus* and *Mangifera indica* L.. However, there are many other tropical crops that have not yet been fully characterized at the genomic level. Thus, expanding the database to include more species could provide researchers with new insights into the genetic basis of these important traits in fruits.

With the continuous development of sequencing technology, data in the field of tropical crops is also increasing. Therefore, multi-omics data are systematically integrated and

analyzed by using a database as the carrier, and different data formats are visualized by using bioinformatics tools, which is convenient for botanists and breeders to mine the genes of related agronomic traits and promote the development and utilization of excellent variety traits and speed up breed selection. It not only provides an important platform for the future genetic breeding of tropical crops, but also provides an important reference for other crops to integrate and use multi-omics data to promote the development of the breeding industry. In the future, the tropical crop database will be updated with additional data on tropical species, as well as more multidimensional omics data and further data analysis tools to provide important support for tropical crop breeding research.

Acknowledgments

This research was supported by Sanya Yazhou Bay Science and Technology City (SCKJ-JYRC-2022-65), Science and Technology special fund of Hainan Province (ZDYF2022XDNY149) and National Key R&D Program of China (2019YFD1000500).

Conflict of interest

The authors declare that they have no conflict of interest.

Dates

Received 29 May 2023; Accepted 5 June 2023; Published online 30 June 2023

References

- Xia Z, Huang D, Zhang S, Wang W, Ma F, et al. 2021. Chromosome-scale genome assembly provides insights into the evolution and flavor synthesis of passion fruit (*Passiflora edulis* Sims). *Horticulture Research* 8:14
- Fu Y, Jiang S, Zou M, Xiao J, Yang L, et al. 2022. High-quality reference genome sequences of two Cannaceae species provide insights into the evolution of Cannaceae. *Frontiers in Plant Science* 13:955904
- Hu G, Feng J, Xiang X, Wang J, Salojärvi J, et al. 2022. Two divergent haplotypes from a highly heterozygous lychee genome suggest independent domestication events for early and late-maturing cultivars. *Nature genetics* 54:73–83
- Wang S, Xiao Y, Zhou Z, Yuan J, Guo H, et al. 2021. High-quality reference genome sequences of two coconut cultivars provide insights into evolution of monocot chromosomes and differentiation of fiber content and plant height. *Genome Biology* 22:304
- Li J, Chen C, Zeng Z, Wu F, Feng J, et al. 2022. SapBase (Sapinaceae Genomic DataBase): a central portal for functional and comparative genomics of Sapindaceae species. *bioRxiv* Preprint
- Yang Z, Liu Z, Xu H, Li Y, Huang S, et al. 2023. A comprehensive multi-omics database for Arecaceae breeding and functional genomics studies. *Plant Biotechnology Journal* 21:11–13
- Hamelin C, Sempere G, Jouffe V, Ruiz M. 2013. TropGeneDB, the multi-tropical crop information system updated and extended. *Nucleic Acids Research* 41:D1172–D1175
- Zou M, Lu C, Zhang S, Chen Q, Sun X, et al. 2017. Epigenetic map and genetic map basis of complex traits in cassava population. *Scientific Reports* 7:41232
- Xia Z, Liu K, Zhang S, Yu W, Zou M, et al. 2018. An ultra-high density map allowed for mapping QTL and candidate genes controlling dry latex yield in rubber tree. *Industrial Crops & Products* 120:351–56
- de Sousa N, Carlier J, Santo T, Leitão J. 2013. An integrated genetic map of pineapple (*Ananas comosus* (L.) Merr.). *Scientia Horticulturae* 157:113–18
- Kuhn DN, Bally ISE, Dillon NL, Innes D, Groh AM, et al. 2017. Genetic map of mango: A tool for mango breeding. *Frontiers in Plant Science* 8:577
- Lespinasse D, Rodier-Goud M, Grivet L, Leconte A, Legnate H, et al. 2000. A saturated genetic linkage map of rubber tree (*Hevea* spp.) based on RFLP, AFLP, microsatellite, and isozyme markers. *TAG Theoretical and Applied Genetics* 100:127–38
- Tran DM, Clément-Demange A, Déon M, Garcia D, le Guen V, et al. 2016. Genetic determinism of sensitivity to *Corynespora cassiicola* exudates in rubber tree (*Hevea brasiliensis*). *PLoS One* 11:e0162807
- Rabbi IY, Kulembeka HP, Masumba E, Marri PR, Ferguson M, et al. 2012. An EST-derived SNP and SSR genetic linkage map of cassava (*Manihot esculenta* Crantz). *Theoretical and Applied Genetics* 125:329–42
- International Cassava Genetic Map Consortium (ICGMC), et al. 2015. High-resolution linkage map and chromosome-scale genome assembly for cassava (*Manihot esculenta* Crantz) from 10 populations. *G3 Genes|Genomes|Genetics* 5:133–44
- Garcia-Oliveira AL, Kimata B, Kasele S, Kapinga F, Masumba E, et al. 2020. Genetic analysis and QTL mapping for multiple biotic stress resistance in cassava. *PLoS One* 15:e0236674
- Ewa F, Asiwe JNA, Okogbenin E, Ogbonna AC, Egesi C. 2021. KASPar SNP genetic map of cassava for QTL discovery of productivity traits in moderate drought stress environment in Africa. *Scientific Reports* 11:11268
- Hussain, W, Campbell, M, Walia, H, Morota, G, et al. 2018. ShinyAIM: Shiny-based application of interactive Manhattan plots for longitudinal genome-wide association studies. *Plant Direct* 2:e00091
- Yang Z, Liang C, Wei L, Wang S, Yin F, et al. 2022. BnVIR: bridging the genotype-phenotype gap to accelerate mining of candidate variations underlying agronomic traits in *Brassica napus*. *Molecular Plant* 15:779–82



Copyright: © 2023 by the author(s). Published by Maximum Academic Press on behalf of Hainan University. This article is an open access article distributed under Creative Commons Attribution License (CC BY 4.0), visit <https://creativecommons.org/licenses/by/4.0/>.