Molecular Plant
Resource article

CellPress
Partner Journal

TBtools-II: A "one for all, all for one" bioinformatics platform for biological big-data mining

Chengjie Chen[1,2,3,*], Ya Wu[1], Jiawei Li[4], Xiao Wang[5], Zaohai Zeng[1,2,3], Jing Xu[1,2,3], Yuanlong Liu[1,2,3], Junting Feng[6], Hao Chen[7], Yehua He[1] and Rui Xia[1,2,3,*]

[1]State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, College of Horticulture, South China Agricultural University, Guangzhou, Guangdong 510640, China

[2]Key Laboratory of Biology and Germplasm Enhancement of Horticultural Crops in South China, Ministry of Agriculture and Rural Affair, South China Agricultural University, Guangzhou, Guangdong 510640, China

[3]Guangdong Laboratory for Lingnan Modern Agriculture, South China Agricultural University, Guangzhou, Guangdong 510640, China

[4]Guangdong Key Laboratory of Non-human Primate Research, Guangdong-Hongkong-Macau Institute of CNS Regeneration, Jinan University, Guangzhou, Guangdong 510632, China

[5]State Key Laboratory of Crop Stress Adaptation and Improvement, Henan Joint International Laboratory for Crop Multi-Omics Research, School of Life Sciences, Henan University, Kaifeng 475004, China

[6]Key Laboratory of Genetic Improvement of Bananas, Sanya Research Institute, State Key Laboratory of Tropical Crop Breeding, Institute of Tropical Bioscience and Biotechnology, Chinese Academy of Tropical Agricultural Sciences, Haikou, Hainan 571101, China

[7]College of Agronomy, Hunan Agricultural University, Changsha 410128, China

***Correspondence: Chengjie Chen (ccj@scau.edu.cn), Rui Xia (rxia@scau.edu.cn)**

https://doi.org/10.1016/j.molp.2023.09.010

ABSTRACT

Since the official release of the stand-alone bioinformatics toolkit TBtools in 2020, its superior functionality in data analysis has been demonstrated by its widespread adoption by many thousands of users and references in more than 5000 academic articles. Now, TBtools is a commonly used tool in biological laboratories. Over the past 3 years, thanks to invaluable feedback and suggestions from numerous users, we have optimized and expanded the functionality of the toolkit, leading to the development of an upgraded version—TBtools-II. In this upgrade, we have incorporated over 100 new features, such as those for comparative genomics analysis, phylogenetic analysis, and data visualization. Meanwhile, to better meet the increasing needs of personalized data analysis, we have launched the plugin mode, which enables users to develop their own plugins and manage their selection, installation, and removal according to individual needs. To date, the plugin store has amassed over 50 plugins, with more than half of them being independently developed and contributed by TBtools users. These plugins offer a range of data analysis options including co-expression network analysis, single-cell data analysis, and bulked segregant analysis sequencing data analysis. Overall, TBtools is now transforming from a stand-alone software to a comprehensive bioinformatics platform of a vibrant and cooperative community in which users are also developers and contributors. By promoting the theme "one for all, all for one", we believe that TBtools-II will greatly benefit more biological researchers in this big-data era.

Key words: TBtools-II, plugin, biological big data, BSA-seq

Chen C., Wu Y., Li J., Wang X., Zeng Z., Xu J., Liu Y., Feng J., Chen H., He Y., and Xia R. (2023). TBtools-II: A "one for all, all for one" bioinformatics platform for biological big-data mining. Mol. Plant. **16**, 1733–1742.

INTRODUCTION

Bioinformatic data analysis has become an indispensable part of biological research. Different research projects have distinct data analysis needs. To handle massive amounts of biological data, researchers need to master the use of numerous bioinformatics software, especially those only available under a command-line environment, and assemble them into a practical workflow. This presents a great challenge for most wet-lab researchers and biologists. To ease this predicament, we published the first version of TBtools software (Chen et al., 2020), which includes over 130 featured functions. It provides a viable option for researchers and is widely used in the biology community. In recent years, the field of plant genomics and bioinformatics has undergone

Published by the Molecular Plant Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and CEMPS, CAS.

significant advancements. For instance, an increasing number of Telomere-to-Telomere (T2T) and haplotype-resolved genomes have been released, making genome-based research and comparative genomics more feasible for various plant species (Naish et al., 2021; Sun et al., 2022; Shang et al., 2023; Shi et al., 2023). In response, the TBtools software has undergone continuous updates and adaptations to remain at the forefront of these developments.

To date, TBtools has been installed and used on many thousands of computers, and the toolkit has been referenced in more than 5000 academic articles. With the growth of the TBtools user community, we constantly receive feedback and suggestions from our daily interactions with users, most of whom are frontline biological researchers. We have come to realize that a dilemma is arising in biological data analysis. On the one hand, different users have distinct demands for data analysis. They want to do personalized analysis given the specific biological process in which they are interested. Each researcher desires specific data analysis tools or workflows to retrieve the best results. Therefore, we have been asked to incorporate more and more functions into TBtools to meet the increasing demands of this type of personalized data analysis, for instance, analysis of different next-generation sequencing data generated from various sequencing strategies, such as chromatin immunoprecipitation sequencing, DNA affinity purification and sequencing, and bulked segregant analysis sequencing (BSA-seq). On the other hand, the development and addition of new features to TBtools (over 100 since the initial publication) has made the toolkit overly multifaceted and cumbersome, diluting its original focus on functions of broad interest and demand. This in turn has made it difficult for users to quickly locate the functions they need. Furthermore, an increase in the size of the stand-alone software would create complications in terms of distribution, installation, and employment. To alleviate this dilemma, we developed an updated version, TBtools-II, in which we have added a range of new features including a plugin mode. This mode enables users to develop their own plugins and manage their installation, selection, and removal according to their needs. We have extensively tested these new features and plugins. They have proven to be a valuable addition to TBtools and will benefit the growing plant biology user community.

RESULTS

Function enhancements

Thanks to valuable feedback and suggestions from users, we have not only improved the current functionalities of TBtools but have also added a number of innovative features spanning diverse topics, including genomic data analyses, comparative genomics, phylogenetics, omics data analysis, and graphics configuration (Table 1).

BLAST Zone: A new function for advanced comparative genomic analysis

Although there are several BLAST-based tools available for comparing two sequences or sequence files, including those present in the first release of TBtools, there is still room for further improvement. These tools often require resetting or rebuilding of the BLAST database and do not support concurrent querying against multiple BLAST databases at once. We have resolved these issues by developing the "BLAST Zone" function, which enables users to store and manage BLAST databases using a tree-like directory system. Users are able to select multiple BLAST databases at once and conduct sequence blasting across a few sequence libraries from various species. Additionally, a companion function, phylogenetic tree construction, has also been integrated, enabling the creation and visualization of an evolutionary tree for the sequences obtained from BLAST searches.

In addition, we often encounter requests to search for the best homologous gene from a species of interest for a given gene with a well-studied function. BLAST merely provides a list of homologs sorted by bit scores/e-values, without considering the phylogenetic relationship. Although a few other tools like OrthoMCL (Li et al., 2003) and Orthofinder (Emms and Kelly, 2019) have been designed for similar purposes, they only support Unix-based operating systems and often require protein searches at a genome-wide level, which usually takes a remarkable amount of processing time. A new function named "Find the Best Homology" has been introduced in TBtools-II, enabling the identification of the homologous genes in a specified species for a target sequence within seconds. This function calls BLASTP (Camacho et al., 2009) for bidirectional protein sequence alignment and employs MUSCLE (Edgar, 2004), trimAl (Capella-Gutierrez et al., 2009), and IQ-TREE2 (Minh et al., 2020) for the construction of a maximum-likelihood tree. This function is extremely useful for researchers working on non-model plants to locate orthologs for genes of interest reported in model species.

Streamlined evolutionary and selection analysis

Phylogenetic tree construction is a common task in molecular biology. Normally, it includes several steps, including sequence acquisition, multiple sequence alignment, alignment trimming, substitution model selection, and ultimately tree reconstruction, which can be quite laborious. TBtools-II includes an easy-to-use function "One Step ML Tree", which seamlessly integrates MUSCLE (Edgar, 2004), trimAl (Capella-Gutierrez et al., 2009), and IQ-TREE2 (Minh et al., 2020) in a sequential workflow. Additionally, it comes with an illustration module for visualizing the resulting phylogenetic tree. In addition, evolutionary selection pressure for gene pairs can also be evaluated in TBtools-II. In the function named "Advanced Ka/Ks Calculator", we implement Needleman–Wunsch alignment (Needleman and Wunsch, 1970) and the Ka/Ks calculation algorithm (Nei and Gojobori, 1986) with pure Java code for both concurrent and high-throughput selection pressure analysis.

Optimized data visualization functionality

In addition to newly developed analysis functions, TBtools-II has undergone extensive optimization in terms of data visualization. Typically, users prefer to use pre-existing templates for graph preparation, if available, and save intermediate files for future use. TBtools has tailored solutions for these data visualization needs. In particular, for heatmaps and gene location maps, which often require modification or updating of datasets, TBtools-II supports the export and import of plot parameter files. This feature allows users to reuse previously defined parameters to prepare charts with a style similar to that used previously. For larger-scale visualization projects, such as Circos plots, TBtools has an instant archive mechanism for intermediate files to

Class	Functions
Genomic data analyses	GXF Rename/Fix/Stat/Patch/Select
	Fasta Window Stat
	Gene Density Profile
	SSR Miner
	Quick Protein Anno
	Protein Parameter Calc
Phylogenetics	One Step ML Tree
	Needleman–Wunsch Alignment
	Advanced Ka/Ks Calculator
	Protein Pairwise SimilarityMatrix
Comparative genomics	BLAST Zone
	Find the Best Homology
	PAF Viz
	One Step MCScanX-Super Fast
	Quick Genome Dot Plot
Omics data analysis	TAU Calc
	Expression Correlation Calc
	Volcano Plot
	DEGs Dist Plot
	SAM/BAM/CRAM BIN Cov
	VCF BIN Cov
	Admixture Q Matrix Viz
Graphics configuration	Heatmap Config
	Gene Location Config
	Advanced Circos Project

Table 1. List of several enhanced functions in TBtools-II.

facilitate their saving, sharing, and reutilization on different computers. For instance, in the "Advanced Circos" function, all intermediate files are storable, distributable, and reusable (Chen et al., 2022).

Overview of the plugin mode in TBtools-II

Vision of "one for all, all for one"

TBtools has been widely embraced since its initial public release because of its exceptional ease of use (with a user-friendly interface), comprehensive functionality, and powerful interactive capabilities. TBtools now boasts a community of many thousands of researchers who rely on the toolkit to fulfill most of their daily data analysis needs, such as sequence management, routine next-generation sequencing data mining, and graph preparation. TBtools has become an indispensable bioinformatics tool in biological labs, especially for wet-lab-based biologists.

However, no matter how powerful modern data analysis software are, it is impossible for any one program to meet all needs of all users. TBtools is no exception. While the standardized upstream bioinformatics is well established, downstream analysis varies greatly depending on individual researcher needs. Personalized data analyses require customized tools with adjustable parameters. We have recently launched a new plugin mode in TBtools-II (Figure 1A). Specifically, we have developed specific functions as dispensable plugins to cater to non-widespread demand. This feature enables users to acquire and remove plugins as needed. In addition, high-quality software cannot be developed without an active user community. Users are the best resources for software development. Thus, we encourage TBtools users with strong bioinformatics skills to contribute their scripts, pipelines, or tools as independent TBtools plugins, which can be shared and distributed throughout the TBtools community.

The theme of TBtools-II is "one for all, all for one" (我为人人, 人人为我) (Figure 1A). We believe that as a powerful and comprehensive biological toolkit, TBtools-II is more than just a software—it is a growing open platform that can provide a one-stop solution to a diverse range of biological data analysis needs. We want to emphasize that all TBtools users are not just utilizing the tool but also actively contributing to its development. We appeal to our users: let us work together to make TBtools-II the best it can be for everyone.

From a toolkit to a platform

With the plugin mode, users can download and install specific plugins from the TBtools plugin store according to their research needs (Figure 1B). If a plugin is no longer needed, users can also delete it easily. To date, the TBtools plugin store has more than 50 featured plugins (Supplemental Table 1), and this has significantly expanded the functionality of the toolkit. There are plugins available for simple text processing, such as "Fastq to Fasta", and for quick batch gene function annotation, such as "Quick Protein Anno", as well as for primer batch design and specificity evaluation, such as "Batch Primer Design" and "Primer Check". Other plugins, such as "DESeq2 GUI Wrapper", are designed for gene differential expression analysis, alongside BSA-seq and chromatin immunoprecipitation sequencing analysis-related workflows. Notably, these plugins can be used in Unix-based operating systems as well as in Microsoft Windows directly without the need to install Windows Subsystem for Linux 2 or virtual machines. In addition, the plugin store also has several functions for data visualization, such as "Bar Plot", "Bubble Plot" and "Sankey Plot".

Overall, the plugin mode is elevating TBtools from a basic toolkit to a comprehensive bioinformatics platform.

From user to developer

The TBtools user community continues to grow, and we often receive inquiries from skillful users asking if TBtools can support the functions in which they are interested. Some of these users are well-trained bioinformaticians with experience in developing established bioinformatics tools or pipelines. Their primary goal is to spread their tools to a wider audience, alleviate their daily data analysis workload, streamline their analysis procedures, or assist the instruction and transmission of skills within their laboratories. To meet these needs, we have developed the "CLI Program Wrapper Creator" plugin, which enables users to package almost any command-line software, pipeline, or script (in program languages such as Java, Python, Perl, and R) into a TBtools-based plugin with a user-friendly graphical interface. They simply need to list the necessary parameters in a form (Figure 1C). This plugin enables easy and real-time update interface design and saving of intermediate files. Users can also easily export the plugins they create for rapid sharing and distribution. Given the rich resource of bioinformatics tools written in R, another plugin named "R Plugin Demo" is specifically designed

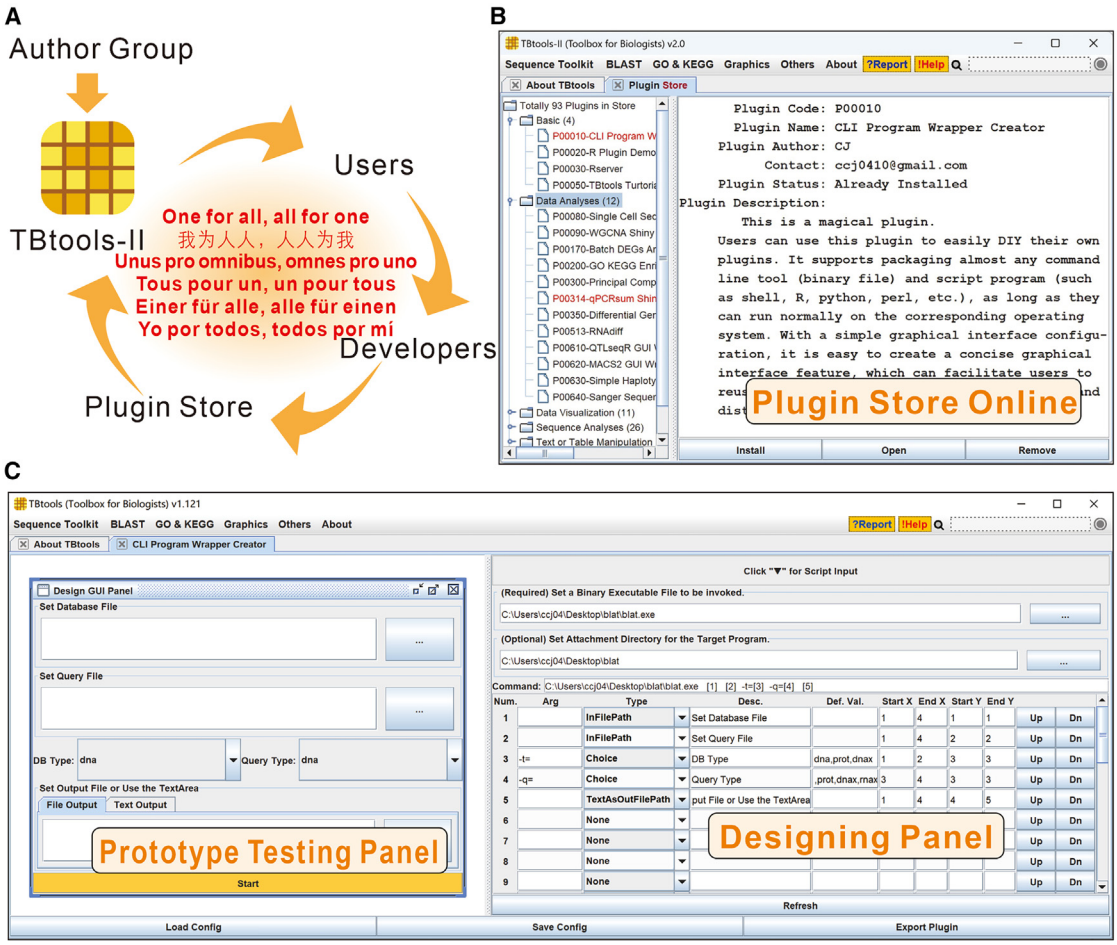


Figure 1. The plugin mode launched in TBtools-II.

(A) By launching the innovative plugin mode in TBtools-II, we aim to foster an active and cooperative community of users, developers, and authors who share in the theme of "one for all, all for one". Translations of "one for all, all for one" in Chinese, Latin, French, German, and Spanish are included.

(B) Plugins in the TBtools-II plugin store are organized in a tree-like structure, with plugin identifiers and names on the left and a detailed description of the main function on the right. Users can easily install, open, and remove plugins according to personal needs.

(C) The "CLI Program Wrapper Creator" plugin was developed for packaging scripts, command-line programs, and pipelines into a plugin compatible with TBtools-II. User-defined parameters and changes in the designing interface on the right trigger instant refreshes in the prototype testing interface on the left.

to convert existing R scripts into graphical plugins in TBtools. For these plugins adopted from published tools or pipelines, acknowledgments and references will be credited to their original developers with the citation information of the original work provided when users use the plugin.

To date, more than half of the plugins available in the TBtools plugin store have been developed and released to the public voluntarily by advanced TBtools users. As such, more and more users are becoming developers of the toolkit, resulting in a thriving and self-motivated user community.

Featured R plugins

To highlight the powerful capabilities of the TBtools plugin mode, three user-developed plugins, which have already been employed and extensively used in the community, were selected for demonstration.

1) Gene differential expression analysis is one of the most common tasks in data analysis. The "DESeq2 GUI Wrapper" plugin (Figure 2A) was designed specifically for this purpose. It encompasses an R script that makes use of the widely used DESeq2 package (Love et al., 2014) for data analysis and the ggplot2 package (Wickham, 2016) for data visualization. Using this plugin, users can now easily perform gene differential expression analysis among multiple datasets with simple button clicks.
2) Recently, gene co-expression analysis has become more and more popular. The plugin "WGCNA Shiny" (Figure 2B) was developed for this. This plugin employs the Shiny module in R to provide a user-friendly, interactive interface for the WGCNA (weighted gene co-expression network analysis) package (Langfelder and Horvath, 2008). By utilizing this plugin, users are able to perform almost all gene co-expression analyses with ease. It is worth noting that the author of this plugin drew inspiration from the iterative WGCNA package (Greenfest-Allen et al., 2017) and incorporated a similar iterative co-expression module analysis into the plugin. This feature often leads to better aggregation of gene modules.

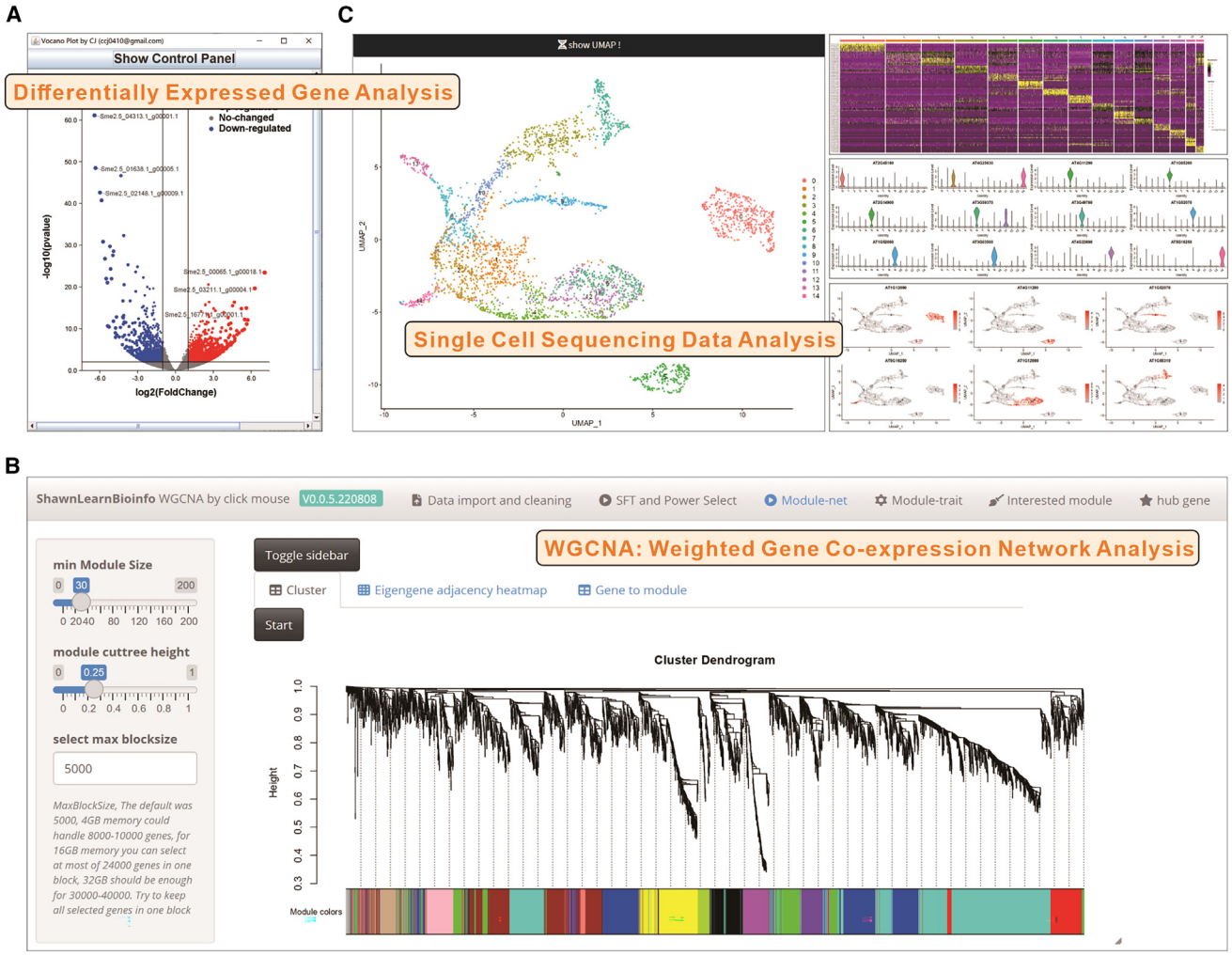


Figure 2. Three R plugins of TBtools-II contributed by users.
(A) A volcano plot shows expression changes of differentially expressed genes analyzed by the "DESeq GUI Wrapper Panel" plugin.
(B) A screenshot displaying the partitioning of co-expressed gene modules generated by gene co-expression network analysis using the "WGCNA ShinyApp" plugin.
(C) Screenshot of segregated cell types and marker genes identified from single-cell transcriptome analysis using the "Seurrat ShinyApp" plugin.

3) Single-cell transcriptomics is a rapidly developing field in plant biology, and the analysis of such data can be complex and challenging. The commonly used R language program package "Seurat" (Hao et al., 2021) was transformed into a TBtools plugin named "Seurat Shiny", providing users with a user-friendly, easy-to-use function for single-cell transcriptome data analysis (Figure 2C).

Plugins for post-genomic-era data analyses

As plant biology research advances into the post-genomic era, vast amounts of sequence resources, such as whole-genome re-sequencing data, are becoming readily available. Making proficient use of these public resources can greatly accelerate scientific research and crop breeding. Therefore, we have developed a series of TBtools plugins, which empower TBtools users to conduct more data analysis tasks on their local computers.

Rapid development of molecular markers

Molecular marker-assisted breeding has been widely applied in the breeding of several major crops—including rice, maize, and wheat—and has yielded highly promising results. However, the development of practical and effective molecular markers has been a persistent challenge, often requiring a significant investment of both time and resources. Fortunately, with the growing availability of genome sequencing and high-throughput sequencing data, we now have an unprecedented opportunity to harness these resources and accelerate progress in breeding programs.

In TBtools-II, we have developed the "SSR Miner" (SSR stands for simple sequence repeat) function for the rapid and efficient identification of SSR loci at the whole-genome level (Figure 3A). To compare two genome sequences of two species or two haploids, users can also apply the "Genome VarScan" plugin to quickly identify structure variation regions (Figure 3B). The "Batch Target Region Primer Design" plugin can be used for batch design of PCR primers specific to the obtained SSR loci and structure variation regions (Figure 3C), and the "Primer Check" plugin allows users to simulate PCR experiments and generate *in silico* gel images, making it straightforward to visually assess the specificity or polymorphism of the primers (Figure 3D).

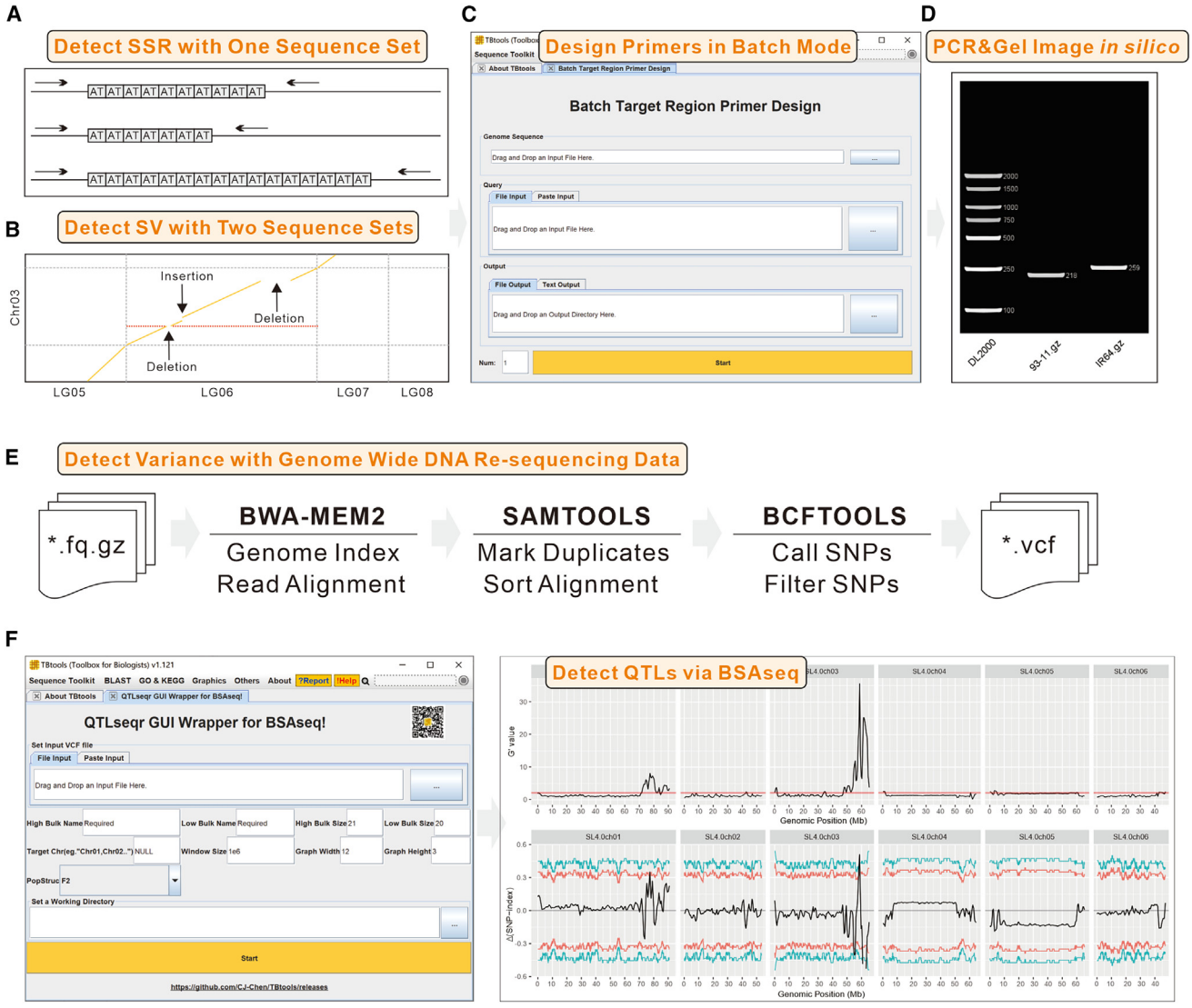


Figure 3. Advanced plugins for post-genomics data analysis in TBtools-II.

(A) The "SSR Miner" plugin offers a robust and reliable method for the identification of SSR loci from a single sequence dataset.

(B) The "Genome VarScan" plugin supports the detection of structure variation between two sequence datasets.

(C) The "Batch Target Region Primer Design" plugin enables primer design in batches.

(D) The "e-Gel Image" plugin generates *in silico* gel image of predicted PCR products.

(E) A brief workflow for the rapid calling of SNPs from resequencing data in TBtools-II.

(F) The interface of the "QTLseqr GUI Wrapper" plugin is shown along with exemplary results generated from BSA-seq data analysis.

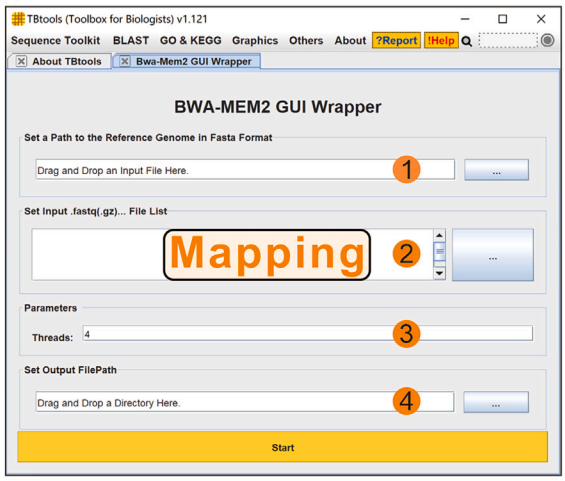
Similarly, with the advancement of high-throughput sequencing technologies, nearly all laboratories can afford the cost of whole-genome sequencing. Investigating the genomic differences of self-sequenced materials against publicly available resources is becoming a routine task. In TBtools, three plugins, i.e., BWA-MEM2 GUI Wrapper (Li, 2013; Vasimuddin et al., 2019), SAMtools GUI Wrapper (Li et al., 2009), and BCFtools GUI Wrapper (Danecek et al., 2021)/Freebayes GUI Wrapper (Garrison and Marth, 2012), can be used in sequence to quickly and accurately identify genomic variations (Figure 3E). This set of plugins is especially beneficial for small research groups that focus on molecular marker-assisted breeding. Using the "QTLseqr GUI Wrapper" plugin (Mansfeld and Grumet, 2018), users can also easily conduct BSA-seq data analysis on a personal computer (Figure 3F), as shown using test data below.

Demonstration of BSA-seq data analysis

To demonstrate the diverse and practical functionalities of TBtools plugins, we utilized a publicly available sequencing dataset from tomato (Soyk et al., 2019) as an illustrative example to introduce the BSA-seq data analysis workflow in a stepwise manner.

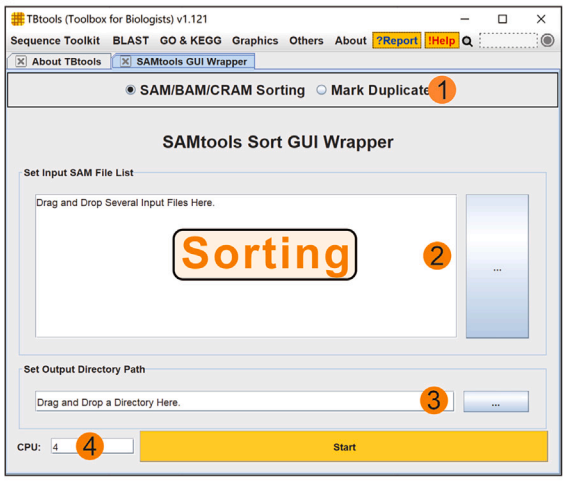
Firstly, users can use the "BWA-MEM2 GUI Wrapper" plugin to perform read alignment (Figure 4A). The tomato reference genome sequence can be downloaded from the SGN database (Fernandez-Pozo et al., 2014), while the sequencing data can be obtained from the NCBI SRA database. After selecting the

A



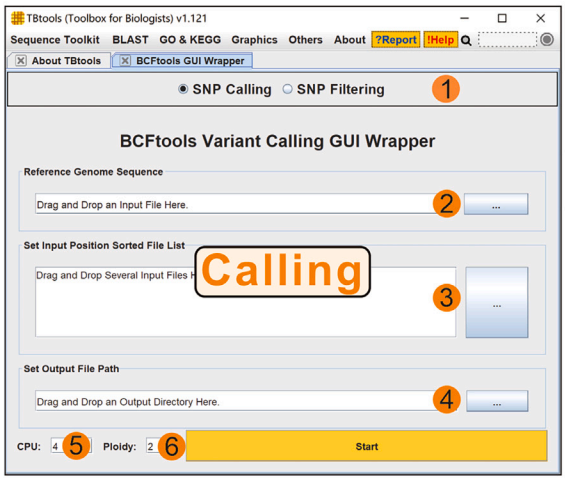
1	Genome Sequence:	tomato.genome.fa
2	Bulk Sequencing Data:	branched.R1.fq branched.R2.fq suppressed.R1.fq suppressed.R2.fq
3	No. Threads:	4
4	Output Directory:	C:\Users\ccj04\Desktop\BSAseq
	Resultant Files:	branched.bam suppressed.bam

B



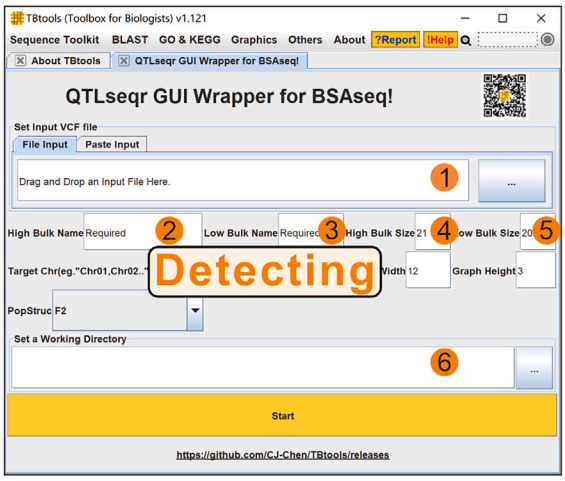
1	Task:	Sort Alignment	Mark Duplicates
2	Alignment Files:	branched.bam suppressed.bam	branched.sorted.bam suppressed.sorted.bam
3	Output Directory:	C:\Users\ccj04\Desktop\BSAseq	
4	No. Threads:	4	
	Resultant Files:	*.sorted.bam	*.sorted.markdup.bam

C



1	Task:	Call SNPs	Filter SNPs
2	Genome Sequence:	tomato.genome.fa	
3	Input Files:	*.sorted.markdup.bam	raw.bcf
4	Output File Path:	C:\Users\ccj04\Desktop\BSAseq	
5	Ploidy:	2	
6	Number of Threads:	4	
	Resultant Files:	raw.bcf	filtered.vcf

D



1	Input Files:	filtered.vcf
2	High Bulk Name:	branched
3	Low Bulk Name:	suppressed
4	High Bulk Size:	21
5	Low Bulk Size:	20
6	Output File Path:	C:\Users\ccj04\Desktop\BSAseq
	Resultant Files:	TBtools_QTLseqR_BSA_QLT.csv
		Gprime.pdf deltaSNP.pdf
		nSNPs.pdf negLog10Pval.pdf

(legend on next page)

genome sequence and sequencing data, users can set the number of threads, such as four, and specify a working directory, then simply click "Start" and wait for the program to finish. Subsequently, two corresponding alignment result files in BAM format will be generated.

Generally, mutation detection software requires the read alignment results to be sorted by genomic position with duplicates removed. This can be achieved by using the "SAMtools GUI Wrapper" plugin (Figure 4B). Two modes, "SAM/BAM/CRAM Sorting" and "Mark Duplicates" can be used sequentially. The input file is the BAM file obtained from the previous step. The user should specify a working directory and the number of available threads and then click "Start". This process will finally generate a read alignment file with the suffix ".sorted.markdup.bam".

Afterward, the user can choose to use either the "BCFtools Variant Calling GUI Wrapper" or the "Freebayes GUI Wrapper" plugin for whole-genome mutation detection and filtering (Figure 4C). Usage of the two plugins is similar, with the user setting the reference genome sequence, inputting the sorted and deduplicated read alignment result file, specifying the output file path, setting the number of threads and ploidy, and then clicking "Start" to complete the analysis. To filter mutation sites, the user can simply adjust the mode of the BCFtools GUI Wrapper plugin to "SNP Filtering" and then set corresponding files and parameters as prompted by the interface.

Finally, the user inputs the obtained high-quality mutation site file into the "QTLseqr GUI Wrapper" plugin (Figure 4D), sets the sample IDs and the numbers of individuals in the two pool samples, provides the output directory, and adjusts other parameters as needed. After clicking "Start", the user will obtain the corresponding quantitative trait locus site information table and resultant figures (Figure 3F).

DISCUSSION

Since its official public release in 2020, TBtools has been used by hundreds of thousands of researchers and has gained far more attention than expected. Over the past 3 years, we have strived to deliver a great user experience while refining the data analysis efficiency of TBtools. The core functionality has been optimized and upgraded by the incorporation of more than 100 new functions, resulting in this upgraded version—TBtools-II. As an integral part of this upgrade, we introduced the plugin mode to better meet personalized data analysis needs. Although there are methods available for quickly packaging command-line tools, such as PyQT, wxPython, and Perl/Tk, they often require users to be proficient with a programming language. TBtools-II simplifies this process with its plugin "CLI Program Wrapper Creator", making it easy for users to develop plugins in a standardized manner. Now, an increasing number of TBtools users are taking part in the development of TBtools plugins, creating a supportive and collaborative ecosystem for all TBtools users. This "one for all, all for one" concept is transforming TBtools into a dynamic user-oriented bioinformatics platform.

Nowadays, bioinformatics analysis faces challenges from the growing size of sequencing data and the amount of data that require processing in a single analysis. Large amounts of data often require large computation resources, which is usually beyond the capacity of local personal computers. In order to meet the demands for heavy computing power, large web-based analysis platforms with substantial computation capacity have been developed. The best example in the field of biology is the Galaxy platform (Giardine et al., 2005), which has been popularly used, especially for data-intensive biomedical research. But as we know, nothing interesting is ever completely one-sided. Web-based platforms have their own shortcomings. First, web-based applications might suffer from unstable networks or low internet speed, which affects the distant transfer of big data. Besides, there are always concerns regarding data safety during data transfer, such as data leaks, illegal usage, and data integrity. Moreover, online platforms often bear a heavier maintenance burden because of the need to keep up with frequent upgrades in network infrastructure or changes in internet usage policies, thus requiring more labor and continuous financial support. In contrast, high-density computing resources are not usually required for most common biological data analysis tasks in wet labs. Local implementation is more convenient and safer, eliminating the need for remote data transfer. As a stand-alone software primarily used on personal computers, TBtools-II excels in this respect, providing a reliable and stable solution for daily data manipulation and analysis of small- or moderately sized datasets. All its functions and plugins can be operated offline. Even without long-term maintenance in the future, users can continue to employ and extend this software as needed. In summary, we believe that TBtools-II is not a replacement for any existing tool or platform but rather serves as an additional supplement to meet the diverse demands of researchers for convenient and efficient biological data analysis.

DATA AND CODE AVAILABILITY

TBtools-II is freely available to non-commercial users at https://github.com/CJ-Chen/TBtools/releases. Demo data can be freely downloaded from https://tbtools.cowtransfer.com/s/631b6a609d354e.

All tests were conducted using six threads, with a peak memory footprint at around 12 GB on a personal computer with a single Intel Core i5 processor (12600KF), 16 GB RAM, and 500 GB storage space and jobs completed in less than 5 h.

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Molecular Plant Online*.

Figure 4. Demonstration of the streamlined workflow of BSA-seq analysis using incorporated plugins in TBtools-II.
(A) The "BWA-MEM2 GUI Wrapper" plugin is used for read alignment.
(B) The "SAMtools GUI Wrapper" plugin is designed for sorting alignment results and removing duplicates.
(C) The "VCFtools GUI Wrapper" plugin is developed for the identification of variants.
(D) The "QTL interval detection" plugin is used for the detection of QTLs.
The numbers in each screenshot correspond to the numbered operation steps indicated.

FUNDING

This work is supported by the Key Area Research and Development Program of Guangdong Province (2022B0202070003, and 2021B0707010004). This work is also supported by the National Science Foundation of China (#32072547, and #32102320); the National Key Research and Development Program (2021YFF1000101, and 2019YFD1000500); the Special Support Program of Guangdong Province (2019TX05N193); the Scientific Research Foundation of the Hunan Provincial Education Department (20A261),); and the open competition program of top ten critical priorities of Agricultural Science and Technology Innovation for the 14th Five-Year Plan of Guangdong Province (2022SDZG05). C.C. is supported by the Guangzhou Municipal Science and Technology Plan Project (2023A04J0113). J.F. is supported by the Hainan Provincial Natural Science Foundation of China (323QN279).

AUTHOR CONTRIBUTIONS

C.C. and R.X. conceived the project; C.C. and R.X. and designed the functions of the toolkit. C.C. performed all the Java coding. J.L. and X.W. developed the "Seurat ShinyApp" and "WGCNA shinyApp" plugins, respectively. Y.W., Z.Z., J.X., Y.L., J.F., H.C., and Y.H. tested the functions and helped with the preparation of the tutorial manual. C.C. and R.X. prepared the figures and wrote the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

We thank all labmates in the Xia lab, the He lab, and the Xu lab for their generous help. We express our gratitude to all the TBtools plugin developers. We extend our appreciation to Guiyang Watchbio Co., Ltd., Yang Shao, Qi Zhao, and Jianming Zeng for their constructive feedback in building the TBtools ecosystem. We are also grateful for the kind advice from many thousands of TBtools users, especially the >40 advanced users. No conflict of interest is declared.

Received: May 21, 2023
Revised: August 22, 2023
Accepted: September 17, 2023
Published: September 22, 2023

REFERENCES

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinform **10**:421.

Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics **25**:1972–1973.

Chen, C., Chen, H., Zhang, Y., Thomas, H.R., Frank, M.H., He, Y., and Xia, R. (2020). TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. Mol. Plant **13**:1194–1202.

Chen, C., Wu, Y., and Xia, R. (2022). A painless way to customize Circos plot: From data preparation to visualization using TBtools. iMeta **1**:e35.

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., and Li, H. (2021). Twelve years of SAMtools and BCFtools. GigaSci **10**. giab008–4.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32**:1792–1797.

Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. **20**:238.

Fernandez-Pozo, N., Menda, N., Edwards, J.D., Saha, S., Tecle, I.Y., Strickler, S.R., Bombarely, A., Fisher-York, T., Pujar, A., Foerster, H., et al. (2014). The Sol Genomics Network (SGN)—from genotype to phenotype to breeding. Nucleic Acids Res. **43**:D1036–D1041.

Garrison, E.P., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. Preprint at arXiv. https://doi.org/10.48550/arXiv.1207.3907.

Giardine, B., Riemer, C., Hardison, R.C., Burhans, R., Elnitski, L., Shah, P., Zhang, Y., Blankenberg, D., Albert, I., Taylor, J., et al. (2005). Galaxy: A platform for interactive large-scale genome analysis. Genome Res. **15**:1451–1455.

Greenfest-Allen, E., Cartailler, J.-P., Magnuson, M.A., and Stoeckert, C.J. (2017). iterativeWGCNA: iterative refinement to improve module detection from WGCNA co-expression networks. Preprint at bioRxiv. https://doi.org/10.1101/234062.

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell **184**:3573–3587.e29.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinform **9**:559.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at arXiv. https://doi.org/10.48550/arXiv.1303.3997.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics **25**:2078–2079.

Li, L., Stoeckert, C.J., and Roos, D.S. (2003). OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. Genome Res. **13**:2178–2189.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. **15**:550.

Mansfeld, B.N., and Grumet, R. (2018). QTLseqr: An R Package for Bulk Segregant Analysis with Next-Generation Sequencing. Plant Genome **11**, 180006.

Naish, M., Alonge, M., Wlodzimierz, P., Tock, A.J., Abramson, B.W., Schmücker, A., Mandáková, T., Jamge, B., Lambing, C., Kuo, P., et al. (2021). The genetic and epigenetic landscape of the *Arabidopsis* centromeres. Science **374**:eabi7489.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol. Biol. Evol. **37**:1530–1534.

Needleman, S.B., and Wunsch, C.D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. **48**:443–453.

Nei, M., and Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3**:418–426.

Shang, L., He, W., Wang, T., Yang, Y., Xu, Q., Zhao, X., Yang, L., Zhang, H., Li, X., Lv, Y., et al. (2023). A complete assembly of the rice Nipponbare reference genome. Mol. Plant **16**:1232–1236, published August 2023. https://doi.org/10.1016/j.molp.2023.08.003.

Shi, X., Cao, S., Wang, X., Huang, S., Wang, Y., Liu, Z., Liu, W., Leng, X., Peng, Y., Wang, N., et al. (2023). The complete reference genome for grapevine (*Vitis vinifera* L.) genetics and breeding. Hortic. Res. **10**:uhad061.

Soyk, S., Lemmon, Z.H., Sedlazeck, F.J., Jiménez-Gómez, J.M., Alonge, M., Hutton, S.F., Van Eck, J., Schatz, M.C., and Lippman, Z.B. (2019). Duplication of a domestication locus neutralized a cryptic variant that caused a breeding barrier in tomato. Nat. Plants **5**:471–479.

Sun, H., Jiao, W.-B., Krause, K., Campoy, J.A., Goel, M., Folz-Donahue, K., Kukat, C., Huettel, B., and Schneeberger, K. (2022).

Chromosome-scale and haplotype-resolved genome assembly of a tetraploid potato cultivar. Nat. Genet. **54**:342–348.

Vasimuddin, M., Misra, S., Li, H., and Aluru, S. (2019). Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. In 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS), pp. 314–324.

Wickham, H. (2016). Package 'ggplot2'. Create elegant data visualizations using the grammar of graphics. Version **2**:1–89.