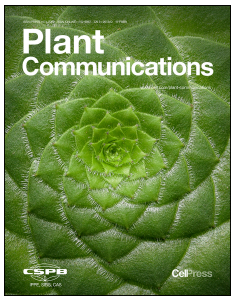


Utility of Triti-Map for bulk-segregated mapping of causal genes and regulatory elements in Triticeae

Fei Zhao, Shilong Tian, Zijuan Li, Luhuan Ye, Yili Zhuang, Meiyue Wang, Yilin Xie, Qiuhong Wu, Shenghao Zou, Wan Teng, Yiping Tong, Dingzhong Tang, Zhiyong Liu, Yijing Zhang

PII: S2590-3462(22)00051-7

DOI: https://doi.org/10.1016/j.xplc.2022.100304

Reference: XPLC 100304

To appear in: *PLANT COMMUNICATIONS*

Received Date: 5 July 2021

Revised Date: 13 January 2022

Accepted Date: 13 February 2022

Please cite this article as: Zhao, F., Tian, S., Li, Z., Ye, L., Zhuang, Y., Wang, M., Xie, Y., Wu, Q., Zou, S., Teng, W., Tong, Y., Tang, D., Liu, Z., Zhang, Y., Utility of Triti-Map for bulk-segregated mapping of causal genes and regulatory elements in Triticeae, *PLANT COMMUNICATIONS* (2022), doi: https://doi.org/10.1016/j.xplc.2022.100304.

Utility of Triti-Map for bulk-segregated mapping of causal genes and regulatory elements in Triticeae

Fei Zhao[1,2], Shilong Tian[1,2], Zijuan Li[1,2], Luhuan Ye[1,2], Yili Zhuang[1,2], Meiyue Wang[1,2], Yilin Xie[1,2], Qiuhong Wu[3], Shenghao Zou[4,5], Wan Teng[2,3], Yiping Tong[2,3], Dingzhong Tang[3,4,5], Zhiyong Liu[2,3†], Yijing Zhang[6†]

[1]National Key Laboratory of Plant Molecular Genetics, CAS Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China
[2]University of the Chinese Academy of Sciences, Beijing 100049, China
[3]State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, The Innovative Academy of Seed Design, Chinese Academy of Sciences, Beijing 100101, China
[4]State Key Laboratory of Ecological Control of Fujian-Taiwan Crop Pests, Fujian Agriculture and Forestry University, Fuzhou, 350002 China
[5]Key Laboratory of Ministry of Education for Genetics, Breeding and Multiple Utilization of Crops, Plant Immunity Center, Fujian Agriculture and Forestry University, Fuzhou, 350002 China
[6]State Key Laboratory of Genetic Engineering, Collaborative Innovation Center of Genetics and Development, Department of Biochemistry, Institute of Plant Biology, School of Life Sciences, Fudan University, Shanghai 200438, China

[†]**Authors for correspondence**

E-mail: zyliu@genetics.ac.cn
zhangyijing@fudan.edu.cn

Phone: +86-10-64806422
+86-21-5492-4206

Short title: Triti-Map for Triticeae gene mapping

Highlights:

Triti-Map provides computational modules facilitating Triticeae gene-mapping. The

Mapping Module and Assembly Module help locate candidate genes and intergenic regulatory elements and the Web-based Annotation Module provides comprehensive downstream annotation and analyses.

Keywords: Agronomic Gene Mapping, Triticeae, Wheat, Bulk Segregated ChIP-seq, Triti-Map

ABSTRACT

Triticeae species, including wheat, barley, and rye, are critical for global food security. Mapping agronomically important genes is crucial for elucidating molecular mechanisms and improving crops. However, Triticeae includes many wild relatives with desirable agronomic traits, and frequent introgressions occurred during Triticeae evolution and domestication. Thus, Triticeae genomes are generally large and complex, making the localization of genes or functional elements controlling agronomic traits challenging. Here, we developed Triti-Map, which contains a suite of user-friendly computational packages specifically designed and optimized to overcome the obstacles of gene-mapping in Triticeae, as well as a web interface integrating multi-omics data from Triticeae for efficient mining of genes or functional elements controlling particular traits. The Triti-Map pipeline accepts both DNA and RNA bulk-segregated sequencing data as well as traditional QTL data as inputs for locating genes and elucidating their functions. We illustrate the usage of Triti-Map with a combination of bulk segregated ChIP-seq data to detect a wheat disease resistance gene and its promoter sequence absent in the reference genome and clarify its evolutionary process. We hope that Triti-Map will facilitate gene isolation and accelerate Triticeae breeding.

INTRODUCTION

Triticeae species, including wheat and barley, are among the major food crops. Transposable element bursts before and after the divergence of Triticeae species contributed to their large genome sizes (Wicker et al., 2018). For example, the worldwide staple common wheat has a genome size of 16 Gb (International Wheat Genome Sequencing Consortium(IWGSC)et al., 2018). Additionally, Triticeae species have complex genomes, with many wild relatives with desirable agronomic traits deployed for crop improvement. Distant hybridizations and introgressions occurred frequently during the evolution and domestication of Triticeae, as well as modern breeding processes (Feldman and Levy, 2012). The considerable size and complexity of the genomes of these species are substantial obstacles for researchers trying to localize genes or functional elements controlling specific agronomic characteristics.

Recently developed mapping methods [e.g., Bulked Segregant Analysis (BSA)] have significantly decreased the cost and labor associated with genetic research (Zou et al., 2016). Combining BSA with RNA-seq and exome-seq methods has further lowered the total cost of map-based cloning (Miller et al., 2013; Hill et al., 2013). Several modified strategies based on BSA have enhanced genetic analysis via improved sequence assembly or optimized calculations (Abe et al., 2012; Takagi et al., 2013a; Fekih et al., 2013). The accuracy and resolution of map-based cloning depend on the accuracy and distribution of genetic markers heterogeneous among populations. Table 1 lists the advantages and disadvantages of different sequencing strategies used to obtain molecular markers in terms of overcoming these obstacles. These segregant strategies have facilitated the detection of essential loci controlling target traits.

However, for Triticeae species with extensive and frequent introgression, gene mapping is confronted with the following difficulties:

First, high computational cost due to the large genome. Specific packages and optimized parameters are needed to overcome the obstacles presented by the large genomes.

Second, the causal gene may not be present in the reference genome within the candidate region. Two strategies may be employed. The first is to collect syntenic regions and genes from other Triticeae genomes. The second is to enrich functional regions by RNA-seq or ChIP-seq followed by sequencing and *de novo* assembly. Recent report demonstrated that a large fraction of genes and regulatory elements could be captured by ChIP-seq, without relying on reference genome sequences (Qi et al., 2018).

Third, the casual loci may be present in the regulatory region. For example, a recent report in barley revealed that deletion of one 'TA' short tandem repeat in the promoter region was sufficient to confer the six-rowed trait (Wang et al.,2021). The large intergenic regions in Triticeae genomes have a substantial abundance of regulatory elements. To identify the regulatory region and the gene region simultaneously, a new strategy involving the application of chromatin immunoprecipitation (ChIP)-seq technology to capture the core genome was recently proposed (Qi et al., 2018).

Forth, due to the generally significant linkage disequilibrium of Triticeae, gene mapping strategies generally result in large candidate regions, which needs functional annotation to narrow down candidate gene or elements based on multi-omics information. The tissue specificity and specific response to certain treatment may give important clues about the causal gene or element, which require well processed and organized multi-omics data.

In this study, we introduce Triti-Map, a tool for efficient gene-mapping based on bulk segregated DNA or RNA sequencing data and QTL data in Triticeae. Triti-Map contains a series of computational packages specifically optimized for Triticeae species together with a web-interface integrating multi-omics data from Triticeae to maximize the mining of public data and sequencing results to identify candidate genes. We illustrate how Triti-Map may be used to locate candidate genes controlling disease resistance based on the bulked segregant ChIP-seq method, making it an easy-to-use resource for efficient gene mapping in Triticeae species.

RESULTS

Computational modules of Triti-Map for locating candidate intervals and detecting specific candidate sequences

The major obstacles for Triticeae gene-mapping include the large genomes and a high frequency of introgressed genes that may not be present in reference genomes. Here we present Triti-Map, which contains a series of computational packages and a web-based interface specifically designed and optimized for Triticeae species to narrow down candidate intervals and detect specific candidate sequences (Figure 1). Figure 1 illustrates the workflow of the Triti-Map computational modules. Triti-Map uses a pair of bulk sequencing datasets (another pair of parental datasets is optional) and a user-set parameter list to reveal candidate regions and sequences. Triti-Map supports the processing of segregated DNA and RNA sequencing results and traditional QTL data. The input data types are automatically detected and subjected to appropriate analyses.

The package uses the following analysis pipeline. Briefly, the pre-processed reads are analyzed by two modules. The Interval Mapping Module maps the reads to a reference genome, after which a traditional method is used for BSA-based interval

detection. The Assembly Module assembles the sequenced reads to identify sequences absent in the reference genome. The assembled sequences specific to the bulk exhibiting the target trait are kept for subsequent analyses. These two computational modules (i.e., the Interval Mapping Module and the *De novo* Assembly Module) integrate computational steps using Snakemake (Koster and Rahmann, 2012). Users only need to set the basic configuration parameters to complete the analysis and obtain candidate genomic regions and phenotype-associated sequences lacking in the reference genome.

The Web-based Annotation Module of Triti-Map for locating functional genes and regulatory elements

After detecting a candidate functional interval, candidate genes or regulatory elements need to be identified and localized. This is a challenging task considering the substantial linkage disequilibrium of wheat populations and the frequency of introgressions (Cheng et al., 2019; Zhou et al., 2020; He et al., 2019). Thus, we developed a web-based platform integrating multi-omics data, sequence data, and functional information to maximize the mining of public data and sequencing results to identify candidate genes (Figure 2A). First, to address the possibility a causal gene may not be present in the reference genome, a comprehensive collection of colinear regions across Triticeae species, including *H. vulgare*, *Ae. tauschii*, *T. urartu*, *T. dicoccoides*, *T. turgidum*, and *T. aestivum*, is retrieved for any input candidate interval. Lineage-specific and shared genes as well as functional information are listed. Second, for the *de novo* assembled sequences specific to the bulk with the target trait (Figure 2B), functional annotation and phylogenetic analysis of sequences via comparison with all publicly available sequences using the EBI RESTful API is conducted to help narrow down the candidates. Third, for genes within a candidate region, a phylogenetic tree representing the gene evolutionary process is constructed, enabling users to deduce the association between the presence of a gene and a particular trait. Fourth, for SNPs contributing to interval analyses, the potential functions and feature distributions are presented, which may help to identify functional elements.

Because the enrichment of specific epigenetic markers reflects the presence of active regulatory elements in non-coding regions, we included a search engine and a

genome browser for detecting and visualizing epigenetic modifications within candidate regions or regions surrounding candidate genes and SNPs (Figure 2B).

Optimization to overcome specific challenges in Triticeae gene mapping

The pipeline was optimized in the following ways to address specific challenges in Triticeae gene mapping (Figure 3).

First, the software and parameters were optimized to decrease the analysis time. Because of the large genomes and long chromosomes of Triticeae species, the commonly used tools for analyzing genomic loci intervals (e.g., bedtools (Quinlan and Hall, 2010)) are very slow. We used GIGGLE (Layer et al., 2018), which quickly compares a large number of wheat genomic intervals based on a temporal indexing scheme using a B+ tree to create a single index of the genome intervals, thereby significantly shortening the time required for analysis. To detect variants, we split the genome according to chromosomes and used GATK HaplotypeCaller (Van der Auwera et al., 2013) for parallel analysis. For DNA-seq-type sequence alignment, BWA-mem2 (Vasimuddin et al., 2019) is the default alignment program, which is faster than bwa (Li and Durbin, 2009) because of enhanced cache reuse, simplified algorithms, and the use of SIMD wherever applicable. The alignment results produced by this program are identical to those of bwa.

Second, given the frequent introgressions and distant hybridizations between Triticeae species, a candidate gene may be absent in the mapped interval of the reference genome. Triti-Map can retrieve new bulk-specific sequences via *de novo* assembly and comparison. Furthermore, it collects Triticeae genomic regions colinear with candidate regions and provides detailed functional annotations to increase the chances of identifying a candidate gene for a particular trait.

Third, regarding sequence comparisons and functional annotations, because of the rapid increase in the available genome sequence and multi-omics data, rather than adopting a local database, the pipeline uses EMBL-EBI RESTful APIs (Madeira et al., 2019) to obtain up-to-date sequence information.

Multiple strategies were employed to simplify the use of Triti-Map. The software and environment required for analyses can be quickly deployed through Conda. All analysis modules were developed and integrated based on the Snakemake workflow

management system. Moreover, the configuration is simple, flexible, and easy to use. Interval mapping (Figure 3, left) and *de novo* sequence assembly (Figure 3, right) can be conducted separately or together with other modules (Figure 3). The Triti-Map usage document (https://github.com/fei0810/Triti-Map/wiki) provides complete instructions and a case study.

Triti-Map and bulk-segregated ChIP-seq applied to identify a disease resistance gene and its promoter region not present in the reference genome

We next illustrate how Triti-Map with the combination of bulk-segregated ChIP-seq help detect the disease resistance gene, which is not present in the reference genome as recently reported (Wu et al., 2021). Two bulked segregant pools were collected from the F_2 progeny of a cross between Xuezao (susceptible to powdery mildew) and 3D249 (resistant to powdery mildew). For each pooled sample, a ChIP-seq analysis was performed for three histone marks (H3K4me3, H3K27me3, and H3K36me3) closely associated with gene activities. The sequencing data were analyzed using the Interval Mapping Module of the Triti-Map package, resulting in the identification of a 6 Mb region on chromosome 7A (chr7A: 724,111,912–730,119,678) highly associated with powdery mildew resistance (Figure 4A, Supplemental Figure S1 and S2). High-quality SNPs within this region were used as input data for the Triti-Map Web-based Annotation Module. Nonsynonymous mutations were detected in two disease resistance-related genes (*TraesCS7A02G551900* and *TraesCS7A02G555200*). Additionally, the integrated epigenetic and motif information revealed the presence of the DNase I hypersensitive site (DHS) as well as H3K36me3, H3K4me3, and H3K9ac in the *TraesCS7A02G551900* promoter, suggesting the promoter is highly accessible to transcription factors (e.g., ABF1) (Figure 4B, Supplemental Table S2). However, the results of an experimental validation indicated that these two genes did not fully segregate with the disease resistance trait.

We hypothesized that the candidate gene is present in other species or wheat populations but not in the reference genome. We used the Assembly Module to detect disease resistance-specific sequences. Among the 10,429 resistant pool-specific scaffolds, 1,704 were partially mapped to Triticeae genomic regions colinear with the candidate region using the collinearity function of the Web-based Annotation Module (Figure 4C). A subsequent functional annotation using EMBL-EBI hmmscan API

detected nine sequences encoding protein domains common among *R* genes. Additionally, four sequences encoding the NB-ARC domain and five sequences encoding the LRR domain were detected (Supplemental Table S3). The EMBL-EBI blast API of the Assembly Module detected sequences encoding one NB-ARC domain and one LRR domain that were highly similar (>99.7% sequence identity) to *Pm60*, which was originally identified as the gene responsible for powdery mildew resistance in the diploid species *T. urartu* (Zou et al., 2018) (Figure 4D, Supplemental Table S4). These two sequences accounted for 69% of the total length of *Pm60*. The *Pm60* sequence was further extended using the sequencing data, resulting in 235 bp extension at the 5' end and 203 bp at the 3' end (10% of the total length of *Pm60*) (Qi et al., 2018). The 5' end was enriched for both active and repressive marks, indicating that this gene is at a bivalent state (Qi et al., 2018).

A candidate gene may be introduced from distant relatives or lost during evolution. To assess these two possibilities, we performed collinearity and evolutionary analyses. Colinear regions corresponding to the candidate region of chromosome 7A were detected in both chromosome 7B and chromosome 7D, both of which contain an *R* gene with high sequence homology to *Pm60*. No highly homologous gene was detected in the tetraploid A and hexaploid A subgenomes, implying that *Pm60* originated in the common ancestor of diploid wheat, but was lost in the A subgenome progenitor before tetraploidization. The corresponding genes in subgenomes B and D underwent divergent evolution and no longer contributed to powdery mildew resistance (Figure 4D). Considered together, these results indicate that bulk-segregated ChIP-seq and Triti-Map enable the rapid identification of a candidate causal gene for a specific phenotype that is not present in the reference genome, facilitating functional and evolutionary analyses.

DISCUSSION

In this study, we developed Triti-Map, consisting of a suite of scripts and web-based platform, specifically to address the challenges of Triticeae gene mapping (i.e., large genomes and frequent introgressions) and could be applied if the candidate gene isn't present in available reference genome sequences.

First, this pipeline increased the likelihood of identifying agronomically genes lacking in the reference genome by integrating information regarding *de novo*

assembled sequences, colinear regions in Triticeae species corresponding to candidate loci, and homologous sequences in public databases. Second, since the ChIP-based strategy could help capture the core genomic regions, including genes and regulatory elements (Qi et al., 2018), bulk-segregated ChIP-seq facilitated detection of non-reference genes and elements, as well as decreasing time and labor required for the analysis and enrichment of the core genome including both gene body and regulatory regions. Third, in addition to ChIP-seq data, the Triti-Map pipeline also accepts other types of DNA and RNA sequencing data as well as traditional QTL data as the input for locating genes and elucidating their functions. The workflow of this pipeline based on Snakemake is easy to maintain and expand. The ability to add more interval mapping and analysis modules later contributes to the broad utility and flexibility of Triti-Map. Fourth, a comprehensive collection of multi-omics data and a systematic curation of functional and evolutionary information facilitate the functional characterization of sequences. This is useful for precisely and reliably localizing candidate genes and for hypothesis-driven research into specific mechanisms.

There are several publicly available web-based resources for wheat genomic data mining, including resources for visualizing and analyzing BSA results (Zhang et al., 2021) and for searching for collinearity and homology among Triticeae species (Chen et al., 2020). Moreover, the database of wheat genomic variations as well as a wheat multi-omics database is available (Blake et al., 2019; Wang W et al., 2020; Chen Y et al., 2020; Zhang L et al., 2021; Ma et al.,2021). Table 2 compared Triti-Map with five wheat genome variation databases and wheat multi-omics databases published. Further mining of the data obtained from bulk-segregated ChIP-seq and Triti-Map using these resources will provide insights regarding downstream mechanisms.

We anticipate that Triti-Map will improve gene isolation and cloning as well as breeding in Triticeae by decreasing the time and labor required to identify agronomically important genes.

METHODS

Data pre-processing

The following data processing pipeline is included in Triti-Map. Triti-Map accepts ChIP-seq and WGS (DNA-seq data) or RNA-seq data as the input. The Fastp (0.20.1) program (Chen et al., 2018) is used to remove adapter sequences and low-quality sequencing bases from the raw data. The pre-processed DNA-seq reads are mapped to the reference genome using the default parameters of BWA-mem2 (Vasimuddin et al., 2019). The RNA-seq reads are mapped using the two-pass mode of STAR (Dobin et al., 2013). Briefly, the reads are first mapped to the reference genome, after which the junction information obtained from the mapping results is applied to reconstruct the genome index and perform a second round of mapping.

Variant detection and candidate genomic interval identification

To detect variants from the DNA-seq data, GATK (Van der Auwera et al., 2013) MarkDuplicates is used to remove the duplicated reads resulting from PCR amplification errors during the library preparation step. The RNA-seq data are processed using the pipeline recommended by GATK (https://gatk.broadinstitute.org/). Uniquely mapped reads are extracted, and the variant sites are detected using GATK HaplotypeCaller. To detect mutations, the genome is split according to chromosomes for parallel computing. After all processes are completed, the data for the variants on each chromosome are merged using GATK MergeVcfs to generate a VCF file containing the information for all mutation sites.

The variant sites are further filtered according to the quality of the variant sites and genotype information using GATK SelectVariants and GATK VariantFiltration. The following combined criteria are used: QD > 2, FS < 60, MQ > 40, SOR < 10, ReadPosRankSum > −8.0, MQRankSum > −12.5, and QUAL > 30. To ensure SNPs are accurately identified, SNPs within a 35 bp sequence are also filtered. Genotype filtering is performed depending on the availability of the parental data and information regarding dominant or recessive alleles. Moreover, mutation sites that satisfy the following conditions are also eliminated: homozygous and identical mutations in two mixed pools; homozygous and identical mutations in the parents; heterozygous mutations in recessive parents or mixed pools; mutations in recessive mixed pools that are homozygous and different from those in the recessive parent. The filtered VCF file is converted to a matrix format using a self-written bash script for the subsequent interval mapping.

The current Triti-Map pipeline supports the use of the allele frequency difference (ΔSNP-index) method (Takagi et al., 2013b) for mapping candidate intervals. The QTLseqr (Mansfeld and Grumet, 2018) package is used to calculate the ΔSNP-index and identify the trait-associated interval. The results are filtered and visualized using a self-written R script.

De novo assembly and bulk-specific sequence screening and annotation

The pre-processed DNA-seq data are assembled using ABySS (version 2.0.2) (Jackman et al., 2017), with k = 90. The transcripts detected by RNA-seq are assembled using the default parameters of rnaSPAdes (version 3.15.0) (Bushmanova et al., 2019).

Assembled sequences longer than 500 bp are retained and mapped to the reference genome using the default parameters of bwa-mem2 (DNA-seq) and minimap2-axsplice--split-prefix (RNA-seq) (Li, 2018). The assembled sequences of the two mixed pools not aligned or partially aligned (i.e., reads containing soft-clipped fragments or with an alternative alignment with cigar strings SA:Z and XA:Z) to the reference genome are selected for the subsequent analysis. The screening process involves the Seqkit (Shen et al., 2016) and BEDOPS (version 2.4.36) (Neph et al., 2012) programs, as well as a self-written bash script.

Target trait-related bulk-specific sequences are obtained via a reciprocal BLAST analysis of sequences from two bulks. To further narrow down the candidate sequences, all regions of Triticeae genomes colinear with the candidate intervals identified using the Interval Mapping Module are identified. The bulk-specific sequences highly similar to these regions are considered as candidate sequences (BLAST default parameters). The candidate sequences are functionally annotated using EBI HMMER3 hmmscan API. Their homologous sequences and information regarding their functions in plants are retrieved from the Ensembl plant database using EBI Blast API. Combining the above information, Triti-Map will generate a table containing functional annotations and homologous sequence information for the candidate new sequences and positional information for comparison with the reference genome.

Triticeae species data collection and processing

Genome sequences and annotation details are obtained from the Ensembl database (Yates et al., 2020) for the following six Triticeae species: *Hordeum vulgare*, *Aegilops tauschii*, *Triticum urartu*, *Triticum dicoccoides*, *Triticum turgidum*, and *Triticum aestivum*. Information is also retrieved for 10 hexaploid wheat genomes(Walkowiak et al., 2020). Epigenomic data from Triticeae species are compiled, including different histone modification ChIP-seq and DNase I hypersensitive site (DHS) data (Figure 2A, Supplemental Table S1). Additionally, MACS (Zhang et al., 2008) is used to identify read-enriched regions, whereas MotifScan (Sun et al., 2018) is used to locate transcription factor-binding motifs. OrthoFinder(Emms and Kelly, 2019) is applied to identify orthologous genes across Triticeae species. JCVI MCScan (Tang et al., 2008) is used to identify gene pairs in colinear regions. EggNOG-mapper (Huerta-Cepas et al., 2017) is used to functionally annotate genes in different Triticeae species, whereas eGPS (Yu et al., 2019) is used to perform a population genetics analysis with a high-density genetic variation map (VMap 1.0) of wheat (Zhou et al., 2020).

Web-based platform construction

A web-based platform was developed using R Shiny, and the front frame was produced using bs4Dash. In this platform, ANNOVAR (Wang et al., 2010) is used to annotate the uploaded mutation information, whereas GIGGLE (Layer et al., 2018) is used to annotate epigenetic modifications and motifs. The annotations and the results of other analyses are displayed in a formatted table using the R reactable package. The distribution of variant site positions on genes is displayed using the R trackviewer (Ou and Zhu, 2019) package. The distribution of the variant site features, the distribution of chromosomal density, and the results of the collinearity analysis are visualized using Echarts4r and plotlyR. Moreover, EBI API (Madeira et al., 2019) is used for the functional annotation of sequences and the determination of sequence similarity. The evolutionary tree for homologous genes from different subgenomes is constructed using ggtree (Yu et al., 2017). The genome browser containing data for the apparent modifications in each species is developed based on the Jbrowse (Buels et al., 2016) configuration.

Availability of the Triti-Map package and web-based interface

The Triti-Map package was developed using Snakemake (Koster and Rahmann, 2012) and Conda. The package and manual are available online (https://github.com/fei0810/Triti-Map). Triti-Map can be installed from Bioconda (Grüning et al., 2018) and Docker. The web-based Triti-Map interface is also accessible (http://bioinfo.cemps.ac.cn/tritimap).

Sample and sequencing data processing for a case study involving the detection of a disease resistance gene

The powdery mildew-resistant common wheat cultivar 3D249 is an F_7 wheat-WEW introgression line developed by Professor Tsomin Yang of China Agricultural University, Beijing, China (pedigree: Jingshuang 27//Yanda 1817/WE18/3/Wenmai 4). Common wheat cultivar Xuezao is highly sensitive to *Blumeria graminis* f. sp. *graminis* (*Bgt*)#E09. Two-week-old seedlings of 30 F_3 generation homozygous resistant and susceptible materials derived from a Xuezao × 3D249 hybridization were pooled to construct resistant and susceptible DNA bulks for a ChIP-seq analysis. The ChIP experiments involved antibodies specific for H3K27me3 (Upstate, USA, Cat. 07-449), H3K4me3 (Abcam, Cat. Ab8580), and H3K36me3 (Abcam, Cat. Ab9050). The HiSeq 2500 system was used for sequencing (150 bp paired-end reads) (Beijing Nuohe Company).**DATA ACCESS**

The package and manual are available online (https://github.com/fei0810/Triti-Map). Triti-Map can be installed from Bioconda and Docker. The web-based Triti-Map interface is also accessible (http://bioinfo.cemps.ac.cn/tritimap).

The ChIP-seq data have been deposited in the Sequence Read Archive (SRA) and assigned the identifier accession PRJNA725543 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA725543).

DISCLOSURE DECLARATION

The authors declare that they have no conflict of interest.

ACKNOWLEDGEMENT

We thank Huang Tao for his help in maintaining the high-performance computing server. We thank Zhougeng Xu for insightful communications. We are grateful to Martin Mascher for facilitating access to wheat 10+ genomes data. We thank Robbie Lewis, MSc, from Liwen Bianji (Edanz) (www.liwenbianji.cn/) for editing a draft of this manuscript.

AUTHOR CONTRIBUTIONS

Z.Y.L. and Y.J.Z. conceived the project. F.Z. designed the software; F.Z and S.L.T. performed the software coding. F.Z., M.Y.W., S.L.T., Y.L.X. and M.M. analyzed the data. Z.J.L., L.H.Y., Y.L.Z., and Q.H.W. performed the experiments. Z.F. and Y.J.Z. prepared the figures and wrote the manuscript, other co-authors critically reviewed and modified the manuscript.

FUNDING

This study was supported by the National Science Fund for Excellent Young Scholars (32022012).

CONFLICT OF INTEREST

The authors declare no conflict of interest

REFERENCES

Abe, A. et al. (2012). Genome sequencing reveals agronomically important loci in rice using MutMap. Nat Biotechnol **30**: 174–178.

Avni, R., Zhao, R., Pearce, S., Jun, Y., Uauy, C., Tabbita, F., Fahima, T., Slade, A., Dubcovsky, J., & Distelfeld, A. (2014). Functional characterization of GPC-1 genes in hexaploid wheat. Planta 239(2): 313–324.

Avni, R. et al. (2017). Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. Science **357**: 93–97.

Blake, V. C., Woodhouse, M. R., Lazo, G. R., Odell, S. G., Wight, C. P., Tinker, N. A., Wang, Y., Gu, Y. Q., Birkett, C. L., Jannink, J. L., Matthews, D. E., Hane, D. L., Michel, S. L., Yao, E., & Sen, T. Z. (2019). GrainGenes: centralized small grain resources and digital platform for geneticists and breeders. Database : the journal of biological databases and curation 2019: baz065.

Buels, R., Yao, E., Diesh, C.M., Hayes, R.D., Munoz-Torres, M., Helt, G., Goodstein, D.M., Elsik, C.G., Lewis, S.E., Stein, L., and Holmes, I.H. (2016). JBrowse: a dynamic web platform for genome visualization and analysis. Genome Biol **17**: 1–12.

Bushmanova, E., Antipov, D., Lapidus, A., and Prjibelski, A.D. (2019). rnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. GigaScience **8**.

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics **34**: i884–i890.

Chen, Y., Song, W., Xie, X., Wang, Z., Guan, P., Peng, H., Jiao, Y., Ni, Z., Sun, Q., and Guo, W. (2020). A Collinearity-Incorporating Homology Inference Strategy for Connecting Emerging Assemblies in the Triticeae Species as a Pilot Practice in the Plant Pangenomic Era. Mol Plant **13**: 1694–1708.

Cheng, H. et al. (2019). Frequent intra- and inter-species introgression shapes the landscape of genetic variation in bread wheat. Genome Biology **20**: 136.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics **29**: 15–21.

Dong, C. et al. (2020). Combining a New Exome Capture Panel With an Effective varBScore Algorithm Accelerates BSA-Based Gene Cloning in Wheat. Front Plant Sci **11**.

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

Fekih, R. et al. (2013). MutMap+: genetic mapping and mutant identification without crossing in rice. PLoS One **8**: e68529.

Feldman, M. and Levy, A.A. (2012). Genome evolution due to allopolyploidization in wheat. Genetics **192**: 763–774.

Golicz, A.A. et al. (2016). The pangenome of an agronomically important crop plant Brassica oleracea. Nature Communications **7**: 13390.

Grüning, B., Dale, R., Sjödin, A., Chapman, B.A., Rowe, J., Tomkins-Tinch, C.H., Valieris, R., Köster, J., and Bioconda Team (2018). Bioconda: sustainable and comprehensive software distribution for the life sciences. Nat Methods **15**: 475–476.

He, F. et al. (2019). Exome sequencing highlights the role of wild-relative introgression in shaping the adaptive landscape of the wheat genome. Nat Genet **51**: 896–904.

Hill, J.T., Demarest, B.L., Bisgrove, B.W., Gorsi, B., Su, Y.-C., and Yost, H.J. (2013). MMAPPR: mutation mapping analysis pipeline for pooled RNA-seq. Genome Res. **23**: 687–697.

Hindorff, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S., and Manolio, T.A. (2009). Potential etiologic and functional implications of

genome-wide association loci for human diseases and traits. Proceedings of the National Academy of Sciences **106**: 9362–9367.

Huerta-Cepas, J., Forslund, K., Coelho, L.P., Szklarczyk, D., Jensen, L.J., von Mering, C., and Bork, P. (2017). Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. Mol Biol Evol **34**: 2115–2122.

International Wheat Genome Sequencing Consortium (IWGSC) et al. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science **361**.

Jackman, S.D., Vandervalk, B.P., Mohamadi, H., Chu, J., Yeo, S., Hammond, S.A., Jahesh, G., Khan, H., Coombe, L., Warren, R.L., and Birol, I. (2017). ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. Genome Res. **27**: 768–777.

Jaganathan, D., Bohra, A., Thudi, M., and Varshney, R.K. (2020). Fine mapping and gene cloning in the post-NGS era: advances and prospects. Theor Appl Genet **133**: 1791–1810.

Koster, J. and Rahmann, S. (2012). Snakemake--a scalable bioinformatics workflow engine. Bioinformatics **28**: 2520–2522.

Layer, R.M., Pedersen, B.S., DiSera, T., Marth, G.T., Gertz, J., and Quinlan, A.R. (2018). GIGGLE: a search engine for large-scale integrated genome analysis. Nat Methods **15**: 123–126.

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics **34**: 3094–3100.

Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25**: 1754–1760.

Li, L., Li, D., Liu, S., Ma, X., Dietrich, C.R., Hu, H.-C., Zhang, G., Liu, Z., Zheng, J., Wang, G., and Schnable, P.S. (2013). The Maize glossy13 Gene, Cloned via BSR-Seq and Seq-Walking Encodes a Putative ABC Transporter Required for the Normal Accumulation of Epicuticular Waxes. PLOS ONE **8**: e82333.

Ling, H.-Q. et al. (2018). Genome sequence of the progenitor of wheat A subgenome Triticum urartu. Nature **557**: 424–428.

Liu, S., Yeh, C.-T., Tang, H.M., Nettleton, D., and Schnable, P.S. (2012). Gene Mapping via Bulked Segregant RNA-Seq (BSR-Seq). PLoS ONE **7**: e36406.

Luo, M.-C. et al. (2017). Genome sequence of the progenitor of the wheat D genome Aegilops tauschii. Nature **551**: 498–502.

Ma, S. W. et al. (2021). WheatOmics: a platform combining multiple omics data to accelerate functional genomics studies in wheat. Molecular Plant.

Maccaferri, M. et al. (2019). Durum wheat genome highlights past domestication signatures and future improvement targets. Nat. Genet. **51**: 885–895.

Madeira, F., Park, Y.M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R.N., Potter, S.C., Finn, R.D., and Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res **47**: W636–W641.

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

Mascher, M. et al. (2017). A chromosome conformation capture ordered sequence of the barley genome. Nature **544**: 427–433.

Miller, A.C., Obholzer, N.D., Shah, A.N., Megason, S.G., and Moens, C.B. (2013). RNA-seq–based mapping and candidate identification of mutations from forward genetic screens. Genome Res. **23**: 679–686.

Mo, Y., Howell, T., Vasquez-Gross, H., de Haro, L.A., Dubcovsky, J., and Pearce, S. (2018). Mapping causal mutations by exome sequencing in a wheat TILLING population: a tall mutant case study. Mol Genet Genomics **293**: 463–477.

Montenegro, J., Golicz, A.A., Bayer, P., Hurgobin, B., Lee, H., Chan, C.K., Visendi, P., Lai, K., Doležel, J., Batley, J., and Edwards, D. (2017). The pangenome of hexaploid bread wheat. The Plant journal : for cell and molecular biology.

Neph, S. et al. (2012). BEDOPS: high-performance genomic feature operations. Bioinformatics **28**: 1919–1920.

Ou, J. and Zhu, L.J. (2019). trackViewer: a Bioconductor package for interactive and integrative visualization of multi-omics data. Nat Methods **16**: 453–454.

Qi, M. et al. (2018). CGT-seq: epigenome-guided de novo assembly of the core genome for divergent populations with large genome. Nucleic Acids Res. **46**: e107.

Quinlan, A.R. and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics **26**: 841–842.

Ramirez-Gonzalez, R.H., Segovia, V., Bird, N., Fenwick, P., Holdgate, S., Berry, S., Jack, P., Caccamo, M., and Uauy, C. (2015). RNA-Seq bulked segregant analysis enables the identification of high-resolution genetic markers for breeding in hexaploid wheat. Plant Biotechnology Journal **13**: 613–624.

Ryan, S., Willer, J., Marjoram, L., Bagwell, J., Mankiewicz, J., Leshchiner, I., Goessling, W., Bagnat, M., and Katsanis, N. (2013). Rapid identification of kidney cyst mutations by whole exome sequencing in zebrafish. Development **140**: 4445–4451.

Shcherban, A. B., & Salina, E. A. (2017). Evolution of VRN-1 homoeologous loci in allopolyploids of Triticum and their diploid precursors. BMC plant biology, 17(Suppl 1): 188.

Shen, W., Le, S., Li, Y., and Hu, F. (2016). SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. PLoS One **11**: e0163962.

Su, Z., Hao, C., Wang, L., Dong, Y., and Zhang, X. (2011). Identification and development of a functional marker of TaGW2 associated with grain weight in bread wheat (Triticum aestivum L.). Theor Appl Genet **122**: 211–223.

Sun, H., Wang, J., Gong, Z., Yao, J., Wang, Y., Xu, J., Yuan, G., Zhang, Y., and Shao, Z. (2018). Quantitative integration of epigenomic variation and transcription factor binding using MAmotif toolkit identifies an important role of IRF2 as transcription activator at gene promoters. Cell Discovery **4**: 1–4.

Takagi, H. et al. (2013a). MutMap-Gap: whole-genome resequencing of mutant F2 progeny bulk combined with de novo assembly of gap regions identifies the rice blast resistance gene Pii. New Phytol. **200**: 276–283.

Takagi, H. et al. (2013b). QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. Plant J. **74**: 174–183.

Tang, H., Bowers, J.E., Wang, X., Ming, R., Alam, M., and Paterson, A.H. (2008). Synteny and Collinearity in Plant Genomes. Science **320**: 486–488.

Van der Auwera, G.A. et al. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics **43**: 11.10.1-11.10.33.

Vasimuddin, Md., 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) (IEEE: Rio de Janeiro, Brazil), pp. 314–324.

Walkowiak, S. et al. (2020). Multiple wheat genomes reveal global variation in modern breeding. Nature **588**: 277–283.

Wang, J. H. et al. (2021). Genetic and molecular characterization of determinant of six-rowed spike of barley carrying vrs1.a4. Theor Appl Genet 134(10):3225-3236.

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Research **38**: e164–e164.

Wang, W., Wang, Z., Li, X., Ni, Z., Hu, Z., Xin, M., Peng, H., Yao, Y., Sun, Q., and Guo, W. (2020). SnpHub: an easy-to-set-up web server framework for exploring large-scale genomic variation data in the post-genomic era with applications in wheat. Gigascience **9**.

Wang, Y. et al. (2017). Mapping stripe rust resistance gene YrZH22 in Chinese wheat cultivar Zhoumai 22 by bulked segregant RNA-Seq (BSR-Seq) and comparative genomics analyses. Theor Appl Genet **130**: 2191–2201.

Wicker, T., Gundlach, H., Spannagl, M., Uauy, C., Borrill, P., Ramírez-González, R.H., De Oliveira, R., Mayer, K.F.X., Paux, E., Choulet, F., and International Wheat Genome Sequencing Consortium (2018). Impact of transposable elements on genome structure and evolution in bread wheat. Genome Biology **19**: 103.

Wu, Q. et al. (2021). Bulked segregant CGT-Seq-facilitated map-based cloning of a powdery mildew resistance gene originating from wild emmer wheat (Triticum dicoccoides). Plant Biotechnol J.

Yao, W., Li, G., Zhao, H., Wang, G., Lian, X., and Xie, W. (2015). Exploring the rice dispensable genome using a metagenome-like assembly strategy. Genome Biology **16**: 187.

Yates, A.D. et al. (2020). Ensembl 2020. Nucleic Acids Res **48**: D682–D688.

Yu, D. et al. (2019). eGPS 1.0: comprehensive software for multi-omic and evolutionary analyses. National Science Review **6**: 867–869.

Yu, G., Smith, D.K., Zhu, H., Guan, Y., and Lam, T.T.-Y. (2017). ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods in Ecology and Evolution **8**: 28–36.

Zhang, L. et al. (2021). WheatGmap: a comprehensive platform for wheat gene mapping and genomic studies. Mol Plant **14**: 187–190.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol **9**: R137.

Zhao, G. et al. (2017). The Aegilops tauschii genome reveals multiple impacts of transposons. Nature Plants **3**: 946–955.

Zhou, Y. et al. (2020). Triticum population sequencing provides insights into wheat adaptation. Nat Genet **52**: 1412–1422.

Zou, C., Wang, P., and Xu, Y. (2016). Bulked sample analysis in genetics, genomics and crop improvement. Plant Biotechnol J **14**: 1941–1955.

Zou, S., Wang, H., Li, Y., Kong, Z., and Tang, D. (2018). The NB-LRR gene Pm60 confers powdery mildew resistance in wheat. New Phytologist **218**: 298–309.

TABLE AND FIGURES LEGENDS

Table 1: Pros and cons of different sequencing strategies for identifying molecular markers.

Data type	Sequencing cost	Library construction	Genomic coverage	SNP identification accuracy	Hardware requirements	References
WGS-based	High	Simple	Whole-genome	No obvious bias	High	(Abe et al., 2012; Fekih et al., 2013; Takagi et al., 2013a, 2013b)
RNA-seq based	Low	Simple	Expressed gene	Affected by gene expression level and alternative splicing	Low	(Liu et al., 2012; Li et al., 2013; Hill et al., 2013; Zhou et al., 2020)
Exome Capture	Low	Complicated	Exons designed in probe	Affected by reference genome, gene annotation and probe design	Low	(Ryan et al., 2013; Mo et al., 2018; Dong et al., 2020)
ChIP-seq based	Low	Median	Core-genome including gene and regulatory elements	No obvious bias	Median	(Qi et al., 2018; Wu et al., 2021)

Table 2: Comparison of features between Triti-Map and published databases

	GrainGenes Blake et al., 2019	Wheat-SnpHub-Portal Wang W et al., 2020	GeneTribe Chen Y et al., 2020	WheatGmap Zhang L et al., 2021	WheatOmics Ma et al.,2021	TritiMap
Description	An improved resource for the small-grains community	SnpHub, an easy-to-set-up web server framework for exploring large-scale genomic variation data in the post-genomic era with applications in wheat	Triticeae GeneTribe, a collinearity-incorporating homology inference strategy for connecting emerging assemblies in the Triticeae Tribe as a pilot practice in the plant pangenomic era	WheatGmap, which integrates multiple BSA mapping models and large amounts of public data to accelerate gene cloning and functional research and facilitate resource sharing	WheatOmics, a platform combining multiple omics data to accelerate functional genomics studies in wheat Mol Plant	Triti-Map is composed of both gene mapping scripts and downstream analysis tools for efficient mapping of both candidate gene and intergenic regulatory elements
Year founded	2005	2020	2020	2020	2018	2021
URL	https://wheat.pw.usda.gov	http://guoweilong.github.io/SnpHub	https://chenym1.github.io/genetribe/	https://www.wheatgmap.org	http://wheatomics.sdau.edu.cn/	http://bioinfo.cemps.ac.cn/tritimap
Applicable platform	Web	Web and Linux	Web	Web	Web	**Web and Linux**
Input data type	/	/	Gene, gene list, or fasta file	Bulk Sequencing VCF Gene, gene list, or fasta file	Gene, gene list, or fasta file	**Fastq raw data** Bulk Sequencing VCF Gene, gene list, or fasta file
Data in website	Genome, molecular and phenotypic information of 4 wheat relative species	Genomic variation datasets of 7 wheat and its progenitors	Homology inference information of 12 Triticeae and 3 outgroup species	High-throughput BSA sequencing datasets of hexaploid wheat (Over 3500 groups)	Multi-omics data including genomes, transcriptomes, variomes, and epigenomes of multiple Triticeae species	Multi-omics data including genomes, transcriptomes, genetic variation, and epigenomes of 7 major Triticeae species
Website main function module	Blast function Genome Browsers	Raw variation data and genomic sequence retrieval	Collinear information and analysis	Gene mapping Gene and SNP annotations Transcriptional analysis Blast function Genome Browsers	Multi-omics data information Transcriptional analysis Regulatory elements analysis Functional analysis Gene identification	***De* novo gene mapping** Gene and SNP annotations Collinear information and analysis Epigenetic features & TF binding motifs **Population genomic statistics** Gene identification
Gene mapping tool	NO	NO	NO	YES	NO	**YES**

Figure 1. Triti-Map workflow.

Triti-Map, which accepts raw sequencing data (ChIP-seq, RNA-seq, or WGS data) from bulks with different traits, comprises the Interval Mapping Module (blue) for locating genomic regions associated with a target trait, the *De novo* Assembly Module (orange) for assembling trait-related sequences, and the Web-based Annotation Module (green) for locating causal variants, candidate genes, or regulatory elements based on integrated multi-omics data and information regarding Triticeae species.

Figure 2. Diagram of the Web-based Annotation Module function.

(A) Data integrated with the Web-based Annotation Module. (B) To locate causal variants and candidate genes or regulatory elements, Triti-Map integrates multi-omics data and provides different levels of analysis, including a collinearity analysis of target regions among Triticeae species as well as a functional and evolutionary characterization of SNPs, genes, or other sequences related to a target trait.

Figure 3. Optimization to address specific challenges of Triticeae gene mapping and annotation.

The major steps that were optimized are marked by the following numbers: 1: steps using specific tools or parameters that shorten the analysis time; 2: steps splitting genomes for parallel analyses; 3: steps in which candidate sequences are filtered according to the colinear regions of candidate intervals across Triticeae species; 4: steps using APIs from public databases to ensure timely updates and minimize local data storage. In each module, nodes with a colored background represent important result files, whereas nodes without a colored background represent the main analysis steps and the tools used.

Figure 4. Triti-Map case study results.

(A) Interval Mapping Module results. Upper panel: causal interval detected using the ΔSNP-index method. Lower panel: enlarged candidate region. (B) Web-based Annotation Module results. From top to bottom: SNP annotation, SNP localization, and epigenome tracks of related regions. (C) Collinearity analysis results. The regions in Triticeae species that are colinear with the detected candidate region are listed. (D) Assembly Module results. Upper panel: two newly assembled sequences (purple) highly similar to *Pm60*. Lower panel: phylogenetic tree presenting the evolutionary distance between *Pm60* and homologous genes in wheat species with a different ploidy level.

Legends for supplemental materials

Supplemental Figure 1. Distribution of SNP counts in each chromosome.

Supplemental Figure 2. Distribution of ΔSNP-index in each chromosome

Supplemental Table 1. Triti-Map Web-based Annotation Module data sources

Supplemental Table 2. High-quality SNPs within candidate region function annotation using Triti-Map Web-based Annotation Module

Supplemental Table 3. Resistance-specific new sequences encoding the NB-ARC and LRR domain

Supplemental Table 4. Resistance-specific new sequences encoding the NB-ARC and LRR domain blast annotation

Input Data

Bulk A

- ChIP-seq
- RNA-seq
- WGS

Bulk B

Interval Mapping Module

1 Locate genomic regions related to target trait

BSA-based mapping

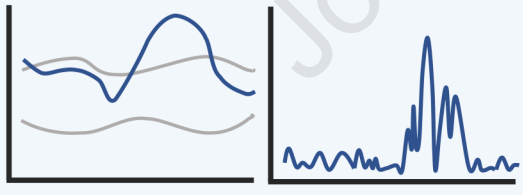
Read mapping

Variant calling

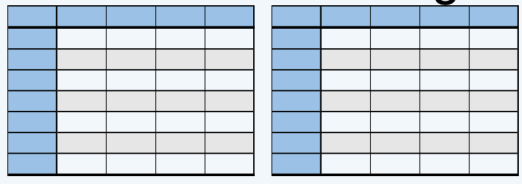
Variant filtering

Bulk variants

ΔSNP-index SNP-counts



Related SNPs and regions



Assembly Module

2 Obtain new sequences and genes related to target trait

de novo assembly

Read assembling

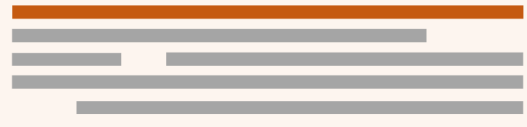
Scaffold filtering

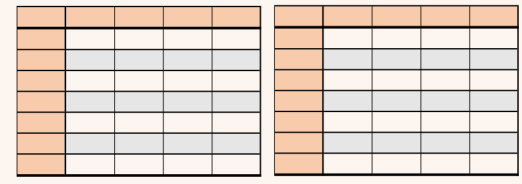
Scaffold mapping

Bulk unique scaffolds

Pfam annotation

BLAST annotation





Web-based Annotation Module

3 multi-evidence supported functional and evolutionary characterization of SNPs, genes, genomic regions and new sequences related to target trait

- SNP annotation
- New sequence annotation
- Visualization
- Homologous analysis
- Collinearity analysis
- Epigenome browser

A

Triti-Map platform data sources

- **Genome and annotation data**
  - *Triticum aestivum*
  - *Triticum dicoccoides*
  - *Aegilops tauschii*
  - *Hordeum vulgare*
  - *Triticum urartu* — mbkbase
  - **10 + genome** — 10WheatGenomes

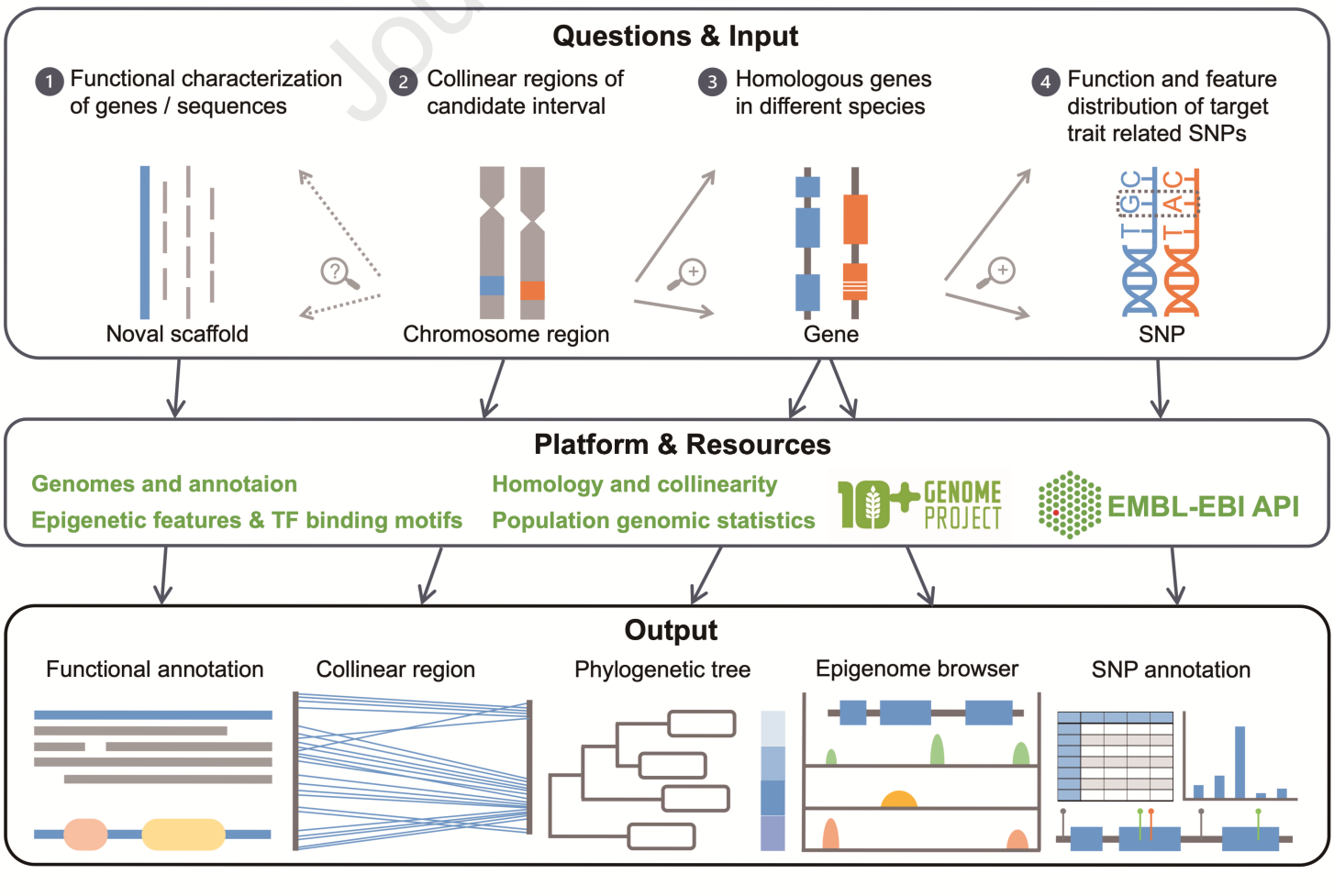
  (*Triticum aestivum*, *Triticum dicoccoides*, *Aegilops tauschii*, *Hordeum vulgare* — Plants Ensembl)
- **Epigenomics data**
  - *Triticum aestivum* — GEO GSE121903
  - *Aegilops tauschii* — SRA PRJNA353049
  - *Hordeum vulgare* — GEO GSE128434
- **EBI variation data**
  - *Triticum aestivum*
  - *Hordeum vulgare*

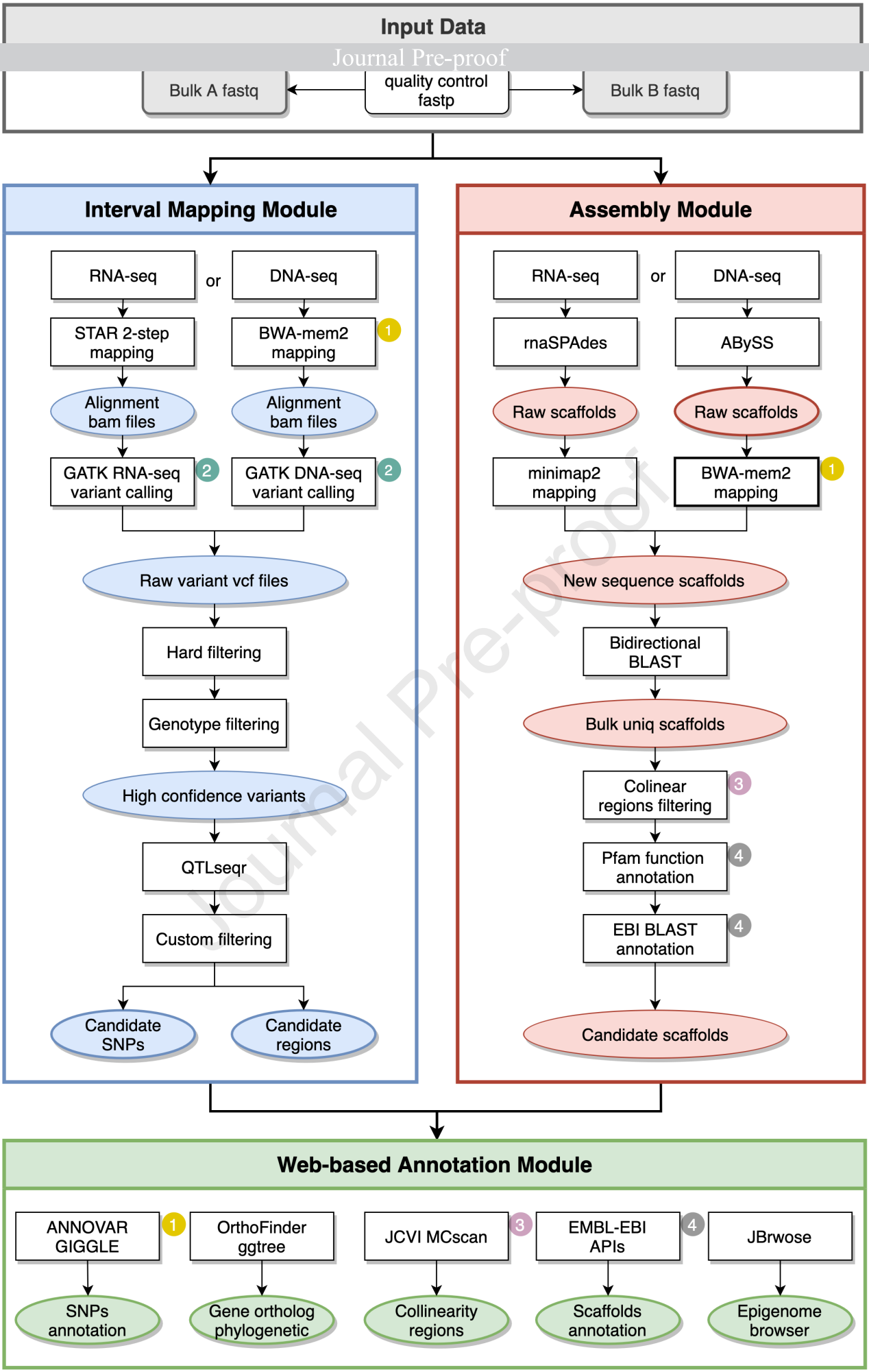
  (Plants Ensembl)
- **Genetic variation map**
  - diploid Triticum
  - diploid Ae. tauschii
  - tetraploid accessions
  - hexaploid accessions

  (BIGD GVM000082)
- **EBI RESTful API**
  - hmmscan
  - blast

  (EBI web service)

B





1 using specific tools or parameters that shorten the analysis time

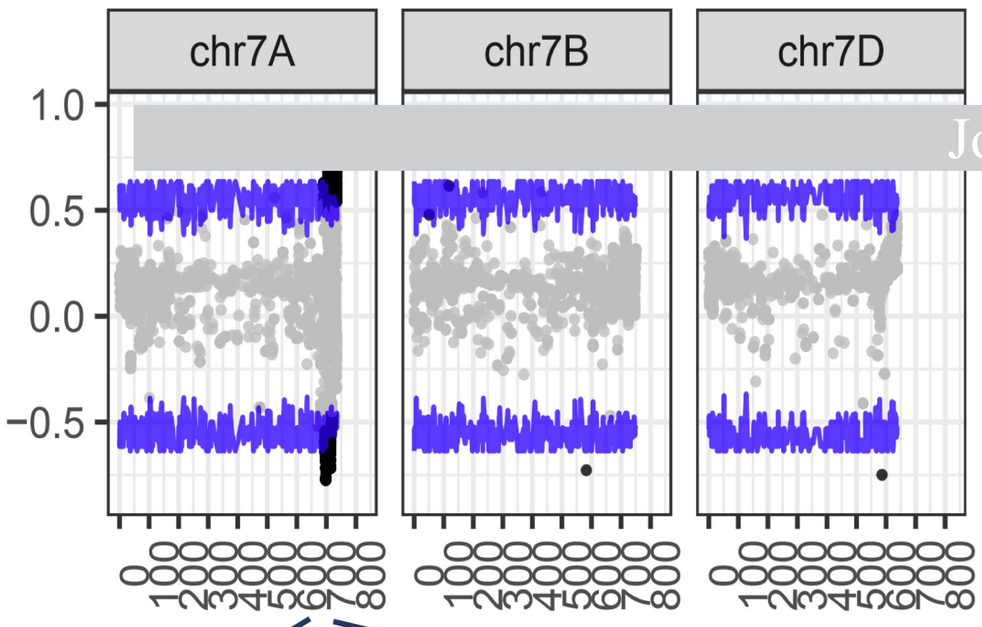
2 splitting genomes for parallel analyses

3 filtering candidate sequences according to the colinear regions of candidate intervals across Triticeae species

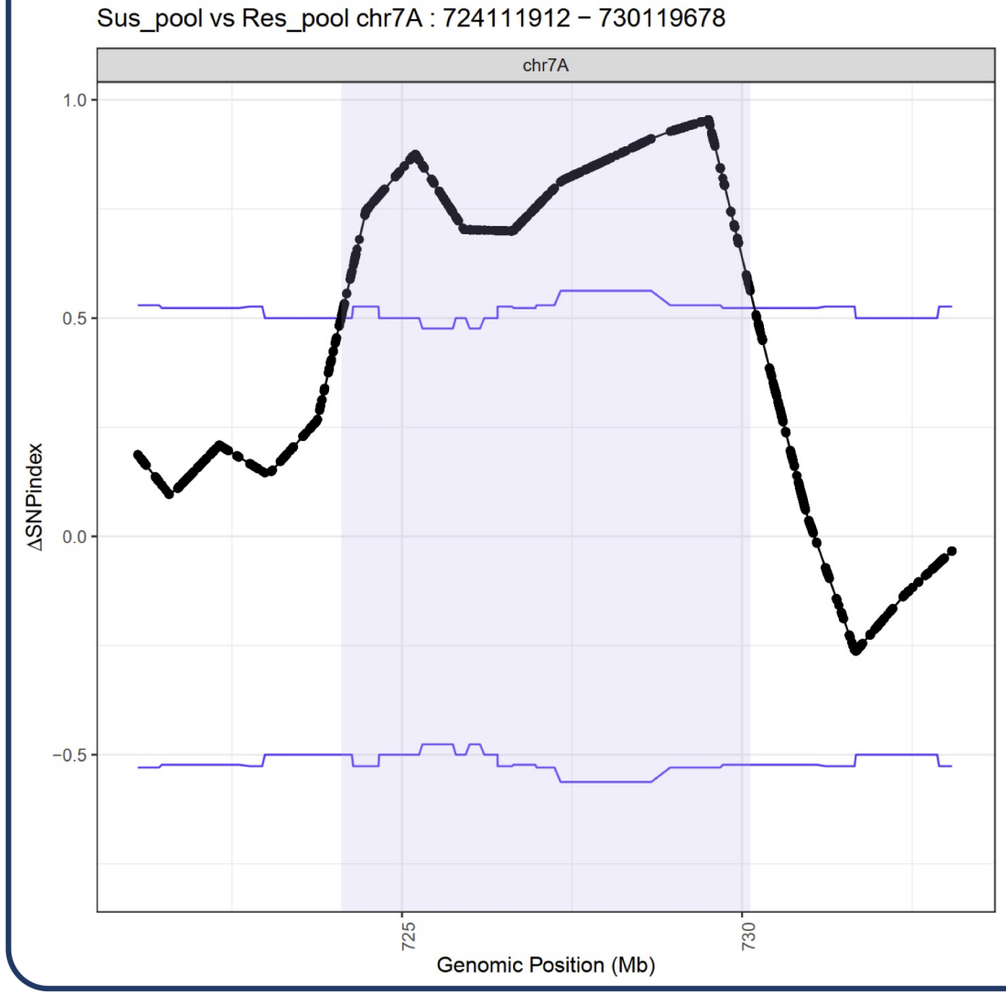
4 using APIs from public databases to ensure timely updates and minimize local data storage

A **Interval Mapping Module Results**

Chromosome7 ΔSNP-index



Candidate region ΔSNP-index

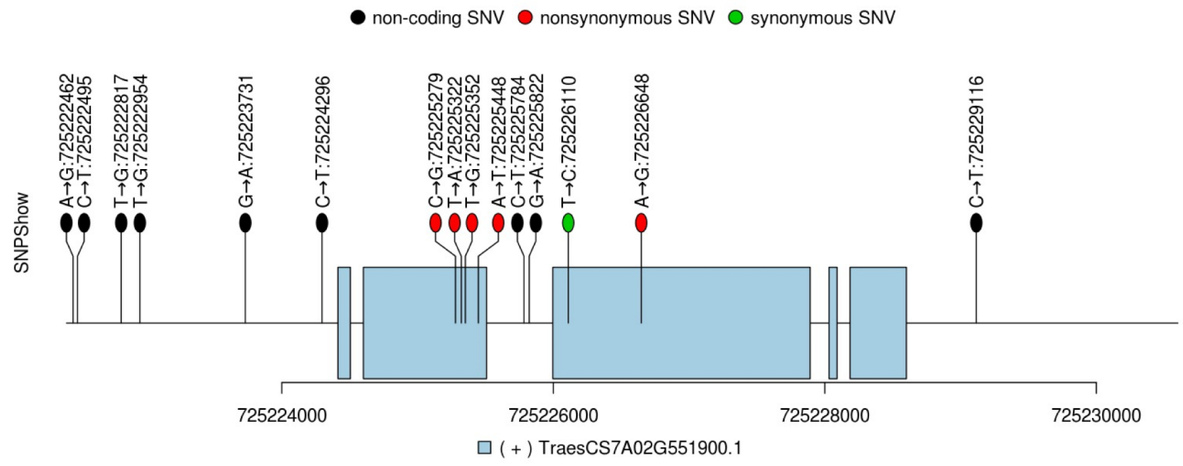


B **Web-based Annotation Module Results**

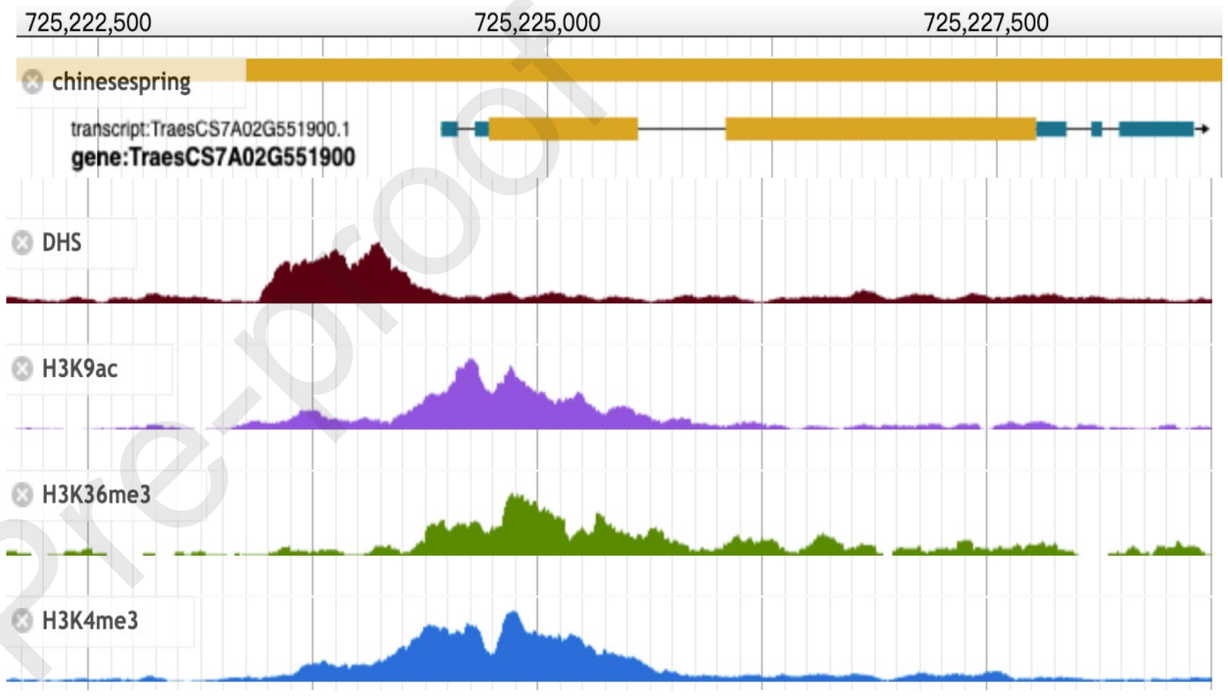
TraesCS7A02G551900 SNP annotation table

Chr	Start	Ref	Alt	Feature	Function	Epigenetic	Motif
7A	725223731	G	A	upstream	NA	H3K9ac;DHS	NA
7A	725224296	C	T	upstream	NA	H3K36me3;H3K4me3;H3K9ac;DHS	NA
7A	725225352	T	G	exonic	nonsynonymous SNV	H3K36me3;H3K4me3;H3K9ac	NA
7A	725225448	A	T	exonic	nonsynonymous SNV	H3K36me3;H3K4me3;H3K9ac	NA
7A	725225784	C	T	intronic	NA	H3K36me3;H3K4me3;H3K9ac	NA
7A	725225822	G	A	intronic	NA	H3K36me3;H3K4me3;H3K9ac	NA
7A	725226110	T	C	exonic	synonymous SNV	H3K4me3	NA
7A	725226648	A	G	exonic	nonsynonymous SNV	H3K4me3	NA
7A	725229116	C	T	downstream	NA	NA	ABF1;bHLH69

TraesCS7A02G551900 SNP visualization



TraesCS7A02G551900 Epigenome Browser

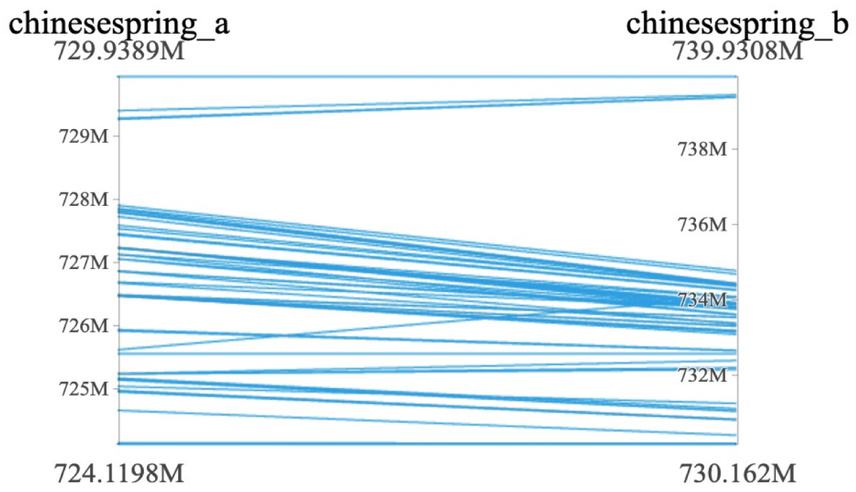


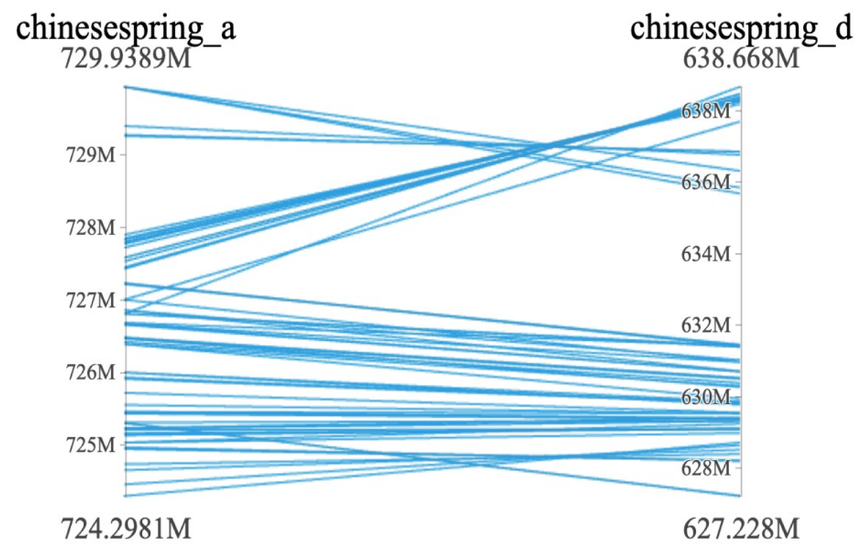
C **Collineartiy analysis Results**

Triticeae collinearity table

Subgenome	Chr	Start	End
barley	chr7H	651897277	654419599
chinesespring_b	7B	730160026	739934199
chinesespring_d	7D	627227193	638668735
goatgrass	7D	631506508	639718055
svevo_a	7A	715672754	721003899
svevo_b	7B	707926003	715263791
urartu	Tu7	705240019	710582462
zavitan2_a	7A	733327835	739176412
zavitan2_b	7B	748029123	765133814

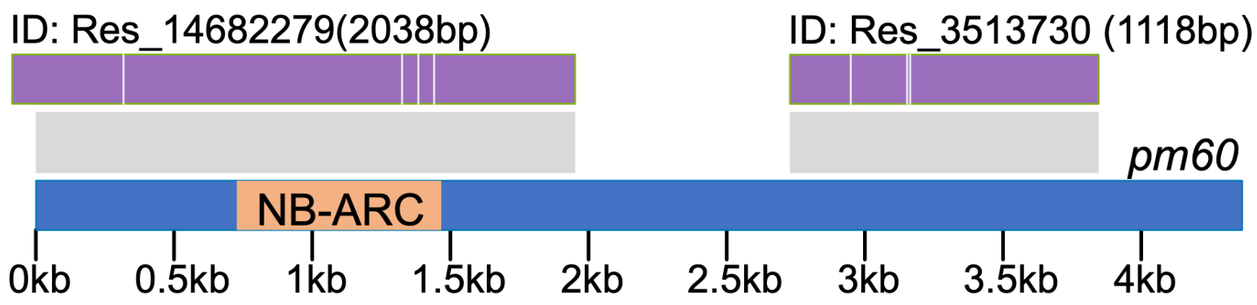
Collinear visualization





D **Assembly Module Results**

Resistance bulk new sequence



Pm60 Phylogenetic tree