BnIR: a multi-omics collaborative resource and tool platform for *Brassica napus* biological study

Zhiquan Yang[1,2,3,†], Shengbo Wang[1,2,†], Lulu Wei[1,2], Yiming Huang[1,2], Dongxu Liu[1,2], Yupeng Jia[1,2], Chengfang Luo[1,2], Yuchen Lin[1,2], Congyuan Liang[1,2], Yue Hu[1,2], Cheng Dai[1,2], Liang Guo[1,2], Qing-Yong Yang[1,2,*]

[1]National Key Laboratory of Crop Genetic Improvement, Hubei Hongshan Laboratory, Huazhong Agricultural University, Wuhan, 430070, China

[2]Hubei Key Laboratory of Agricultural Bioinformatics, College of Informatics, Huazhong Agricultural University, Wuhan, 430070, China

[3]Innovative Center of Molecular Genetics and Evolution, School of Life Sciences, Guangzhou University, Guangzhou, 510405, China

[†]These authors contributed equally: Zhiquan Yang, Shengbo Wang.

[*]Correspondence: Qing-Yong Yang (email: yqy@mail.hzau.edu.cn)

Abstract

In the post-GWAS era, multi-omics techniques have shown great power and potential for candidate gene mining and functional genomics research. However, due to the lack of effective data integration and multi-omics analysis platforms, such techniques have not still been applied widely in rapeseed, an important oil crop worldwide. Here, we constructed a rapeseed multi-omics database (BnIR; http://yanglab.hzau.edu.cn/BnIR), which provides datasets of six omics including genomics, transcriptomics, variomics, epigenetics, phenomics and metabolomics, as well as numerous "variation-gene expression-phenotype" associations by using multiple statistical methods. In addition, a series of multi-omics search and analysis tools are integrated to facilitate the browsing and application of these datasets. BnIR is the most comprehensive multi-omics database for rapeseed so far, and two case studies demonstrated its power to mine candidate genes associated with specific traits and analyze their potential regulatory mechanisms.

Introduction

For comprehensive understanding of the mechanism underlying crop trait formation, it is important to interpret the molecular intricacy and variation at multiple levels such as genome, epigenome, transcriptome, proteome, metabolome and phenome (Edwards and Batley, 2004; Li et al., 2016; Fernie and Gutierrez-Marcos, 2019; Tuggle et al., 2022). In recent years, the rapid advancement in multi-omics techniques has provided researchers with more multidimensional information to dissect complex biological systems (Ichihashi et al., 2020; Wu et al., 2021; Yang et al., 2021). Based on massive multi-omics datasets, numerous research methods have been developed to mine genetic loci and candidate genes regulating phenotypes, such as genome-wide association study (GWAS) (Yu et al., 2006), expression quantitative trait locus (eQTL) mapping(Goring et al., 2007), transcriptome-wide association study (TWAS) (Gusev et al., 2016) and summary data-based Mendelian randomization analysis (SMR) (Zhu et al., 2016). These methods have high efficiencies in mining genetic loci and candidate genes associated with certain traits and have been widely applied in crop breeding (Yu *et al.*, 2006; Wang et al., 2018; Lu et al., 2019; Liu et al., 2020; Tang et al., 2021; He et al., 2022; Zhang et al., 2022a; Zhang et al., 2022b; Zhang et al., 2022c). Multi-omics techniques are becoming increasingly important in assisting crop breeding (Yang *et al.*, 2021). For more efficient utilization of multi-omics data, multi-omics databases have been developed in some important crops, such as wheat (*Triticum aestivum* L.) (Ma et al., 2021), soybean (*Glycine max*) (Grant et al., 2010), tomato (*Solanum lycopersicum*) (Kudo et al., 2017), barley (*Hordeum vulgare* L.) (Li et al., 2022), maize (*Zea mays* L.) (Gui et al., 2020), millet (*Setaria italica* L.) (Yang et al., 2020), cotton (*Gossypium hirsutum* L.)(Yang et al., 2022b) and rice (*Oryza sativa* L.) (Sakai et al., 2013).

Rapeseed is the second most important oilseed crop widely planted in the world. The traits including yield, seed oil content (SOC), oil quality, disease resistance and stress resistance largely determine its production and economic value (Friedt et al., 2018). Due to its special origin and evolutionary history, rapeseed has a polyploid genome and

therefore a more complex sequence composition and regulatory mechanism than diploid crops, which increases the difficulty in fine mapping of the candidate loci and genes (Chalhoub et al., 2014; Friedt *et al.*, 2018). Recently, a large number of genetic loci and candidate genes associated with important traits have been efficiently identified using powerful multi-omics techniques and analysis tools (Lu *et al.*, 2019; Tang *et al.*, 2021; He *et al.*, 2022; Hu et al., 2022; Zhang *et al.*, 2022c). However, the identification is still rather time-consuming, and the regulatory mechanisms of most identified candidate loci or genes on phenotypes remain elusive. In our recent work, we have constructed and released several databases, including BnPIR (Song et al., 2021) (http://cbi.hzau.edu.cn/bnapus/), BnTIR (Liu et al., 2021a) (http://yanglab.hzau.edu.cn/BnTIR/) and BnVIR (Yang et al., 2022a) (http://yanglab.hzau.edu.cn/BnVIR/), by collecting multi-omics data including genome, transcriptome and population genetic variations and developing related omics tools. These databases have been widely applied to facilitate related research work. However, there is still a lack of platforms that integrate multi-omics data to help genetic breeding and functional genomics analysis in rapeseed.

In this study, we constructed BnIR (http://yanglab.hzau.edu.cn/BnIR/), the first *Brassica napus* multi-omics information resource database. BnIR integrates the most comprehensive multi-omics datasets involving various dimensions. In the database, multiple association analysis tools can be applied to mine the genetic loci, candidate genes and genetic variations, providing important resources, tools and reference for rapeseed breeding. In addition, BnIR provides numerous common online multi-omics data analysis tools, which can support all published rapeseed genomes and offers a convenient and efficient analysis platform. We believe that BnIR will be a valuable database for future rapeseed breeding and functional genomics research.

Results

Construction and overview of BnIR

For a comprehensive understanding of rapeseed genomics, we first acquired and integrated rapeseed multi-omics data from recent research and databases. In total, the datasets from six omics were obtained, including 29 genome assemblies, genetic variations of 2,311 accessions, transcriptome data from 2,791 libraries, 118 phenotypes, the contents of 266 metabolites and epigenome signals involving DNA methylation, histone modification, chromatin accessibility and chromatin interaction. Next, the related information was mined by processing and analysis of these datasets. In addition, multiple multi-omics analysis methods were employed to mine the genetic loci, candidate genes and genetic variations, including GWAS, eQTL, TWAS, SMR and co-localization analysis. Finally, a comprehensive rapeseed multi-omics information resource database (BnIR; http://yanglab.hzau.edu.cn/BnIR/) was constructed (Figure 1).

BnIR comprises 12 portals, including Population, Genomics, Transcriptomics, Variation, Phenotype, Epigenetics, Metabolome, Multi-omics and Network, Tools, Download and Help. Abundant and convenient visual tools are provided in these portals for browsing and comparing the germplasm resource information, genome sequences, gene structures, epigenetic signals, metabolite contents and phenotypes, so as to manage germplasm resource and assist the exploring of mechanism underlying gene regulation and evolution.

New features in omics portals

Genomics. Based on our previous BnPIR database, we added Darmor (v.10), Ningyou7 and Express617 genome assemblies as well as their genome sequences and annotations, which can be downloaded from the Genome data module. Genome synteny and gene index have been updated, which can be browsed in the Gbrowse synteny and Gene index modules. According to the phylogenetic trees of homologous genes, the Gene search module was developed for users to browse and compare the gene structures to determine whether the homologous genes have undergone functional differentiation at

the sequence level. In addition, the gene families and transcription factors (TF) of 12 *B. napus* genome assemblies are annotated, which can be browsed in the corresponding modules of the Genomics portal.

Transcriptomics. As for the Transcriptomics portal, the previous core function modules in BnTIR, including gene expression pattern query and eFP browser, have been integrated into BnIR. In eFP, the webpage has been redesigned to greatly improve the response speed by more than 150 times (from 9.348 s to 0.0611 s). In addition, the published RNA datasets from 537 RNA-seq libraries involving seven hormone treatments and six abiotic stress treatments have been added into the eFP module. Users can browse and compare the gene expression pattern between the treatment and control groups.

In order to acquire gene expression data from more accessions, the RNA-seq data of 2,791 published tissue samples were collected, processed and summarized. Finally, the gene expression data of these samples were integrated into the Expression profile (meta library) module of the Transcriptomics portal, in which users can browse the expression of interested genes in all samples. As for the breeding of *B. napus*, ecotype and seed quality improvement are two key periods. Besides, heterosis is an important and widely used phenomenon for genetic improvement. To facilitate studies of the genes involved in these improvement processes, we collected the gene expression datasets from eight representative cultivars of three ecotypes, double-high (high erucic acid and high glucosinolate) and double-low (low erucic acid and low glucosinolate) cultivars (Zhongyou821 and Zhongshuang11), and eight RNA-seq datasets of accessions with heterosis. Users can browse and compare the expression levels of interested genes between different accessions in the Expression profile (meta library)module to determine whether the gene is related to the breeding improvement process. In addition, we collected and processed the population-level gene expression datasets in leaves and seeds at 20 and 40 days after flowering (DAF). Users can browse the population-level expression patterns of the interested genes in the Population expression module. In

140 summary, these new transcriptomic information and tools will help researchers to
141 understand the function and expression features of the interested genes from a broader
142 perspective.

143 **Variation and phenotype.** The previous BnVIR database integrated 10,090,561
144 genetic variations and 21 phenotypes of 2,311 core *B. napus* accessions (Yang *et al.*,
145 2022a). On this basis, we newly collected 216 published phenotype datasets from 13
146 studies (Bus et al., 2011; Sun et al., 2016b; Sun et al., 2016a; Chen et al., 2018; Kittipol
147 et al., 2019; Wu et al., 2019; Song et al., 2020; Xuan et al., 2020; Liu et al., 2021c; Liu
148 et al., 2021b; Wang et al., 2021; Zhang *et al.*, 2022a; Zhang *et al.*, 2022b). These
149 datasets cover a variety of phenotypic traits such as agronomic traits under abiotic stress
150 conditions and seed quality traits, and have been integrated into the Phenotype and
151 Variation portals. Users can browse the phenotypic values and compare them between
152 the accessions of different ecotypes in the Phenotype portal to check whether the
153 phenotype is related to ecotype improvement. In the Variation portal, users can compare
154 the phenotypic value between accessions with different alleles or haplotypes to explore
155 the effect of genetic variation on these phenotypes (Korber et al., 2015; Korber et al.,
156 2016)

157 **Epigenome.** For epigenetic data, we collected the ChIP-seq and WGBS datasets for
158 five tissues of six accessions including six histone modifications, ATAC-seq datasets
159 for seeds of four accessions and Hi-C datasets of three accessions. After filtering and
160 processing, a total of 5,041,596 peaks of ChIP-seq and 386,313 peaks of ATAC-seq
161 were obtained, which were integrated into the Histone modification and Chromatin
162 accessibility modules, respectively. Based on the WGBS datasets, the CG, CHG and
163 CHH methylation levels of genes were obtained and integrated into the DNA
164 methylation module. Based on the Hi-C datasets, chromatin interaction frequency and
165 compartments were calculated and integrated into the Chromatin interaction module.
166 Besides, users can browse the tracks of all epigenetic data of the gene segment of
167 interest using the Jbrowser browser or query these epigenetic datasets in the

168 corresponding module of Epigenetic portal. These datasets integrated the epigenetic
169 information involving histone modification, epigenomic state, chromatin accessibility
170 and interaction, providing the first systematic epigenetic data resource in *B. napus* to
171 help identify regulatory elements in *B. napus* genome and understand their regulatory
172 mechanism.

173 **Metabolome.** In the Metabolome portal, the data about the contents of 266 metabolites
174 of 2 accessions from two studies were collected and integrated. Users can query the
175 related information of these metabolites, such as structure, molecular formula and
176 content in these accessions in the corresponding module, which provides a way to
177 quickly obtain the information of metabolites.

178 Integration and analysis of multi-omics data

179 Based on the collected population-level multi-omics data, we performed GWAS, eQTL
180 and TWAS to mine candidate genes and genetic variations associated with phenotypes.
181 A total of 379 loci were identified to be significantly associated with 55 phenotypes by
182 GWAS, involving 2,119 candidate genes. By using the population-level gene
183 expression datasets from seeds at 20 and 40 DAF of 309 accessions, 1,118,218 eQTLs
184 involving 968,375 lead SNPs(eSNPs) and 33,507 genes(eGenes) including 356,112 *cis*-
185 eQTLs and 762,106 *trans*-eQTLs. In total, 3,724 genes were identified to be
186 significantly associated with 124 phenotypes by TWAS. Next, to more precisely
187 identify the candidate genes associated with phenotypes by combining the three omics
188 datasets, co-localization and SMR analyses were performed based on the GWAS and
189 eQTL results. As a result, co-localization analysis identified 130 genes significantly
190 associated with 24 phenotypic datasets; SMR identified 298 genes significantly
191 associated with 45 phenotypic datasets. By combining TWAS, co-localization analysis
192 and SMR results, a total of 3,991 genes were found to be significantly associated with
193 131 phenotypes (Figure 2A). Known associations reported in previous studies were also
194 identified in our analysis, such as *BnaC09.MYB28* for seed glucosinolate content (SGC)
195 (Case study 1), and *BnaA05.PMT6* for SOC, suggesting that the analysis results are

reproducible. In addition, some new candidate genes were found to be related to important traits, such as *BnaC07.SRK2D* for SOC (Case study 2), suggesting that the analysis can be used to mine new candidate genes.

Based on the above results, GWAS, eQTL, TWAS, COLOC, and SMR modules in the Multi-omics portal were designed for convenient browsing and use of these datasets. For example, users can browse information about the global GWAS signals and significant variations for the phenotypes of interest in the GWAS module. Taking the erucic acid content as an example, when users select 'Erucic acid' in the Trait box and submit it (Figure 2B), the page will display the Manhattan plot of global GWAS results, showing that the locus on chromosome A08 is the most significant locus (Figure 2C). Then, the user can scroll the mouse to magnify this locus and select the region of "A08:18500000–19000000" to browse for the related information of significant variations (Figure 2D). Combined with the gene annotations, two non-synonymous mutations in the reported candidate gene (*BnaA08.FAE1*) for erucic acid content in seeds were both identified in our GWAS results (Figure 2E). In the eQTL module, users can browse the genome-wide information about *cis*-eSNPs and *trans*-eSNPs regulating gene expression levels. For example, the expression level of *BnaA09.ARF18* was reported to be associated with the silique length and thousand-seed weight (Liu et al., 2015). When users enter 'BnaA09G0559300ZS' (*BnaA09.ARF18*) and submit it (Figure 2F), the page will display a list and a circus figure of 12 eQTLs regulating its expression level (Figure 2G), including four *cis*-eSNPs near *BnaA09.ARF18* and eight *trans*-eSNPs from four chromosomes (Figure 2H). Also, users can browse the information of candidate genes associated with phenotypes identified by three methods respectively in the TWAS, SMR, and COLOC modules.

Development and application of multi-omics tools

To provide a more convenient and user-friendly platform for researchers, we developed and integrated 17 multi-omics data analysis tools in BnIR, including some common bioinformatics analysis tools, such as GO and KEGG enrichment analysis, BLAST,

multi-sequence alignment (MSA), sequence extraction, variation annotation, primer designing, electronic PCR and four graph visualization tools, which were all integrated into the Tools portal (Figure 3). Notably, all analysis tools support 28 published *Brassica* genomes, and users can conduct these analyses without switching between different platforms. For population-level multi-omics data analysis, we newly developed COLOC and SMR online analysis and visualization tools based on the rapeseed eQTL results of our database, which allow users to identify candidate genes and causal variants by uploading local GWAS results. In addition, to facilitate the identification of germplasm resources, the SNPmatch tool based on the genetic variation dataset of 2,311 accessions was integrated into the Tool portal (Pisupati et al., 2017). In the SNPmatch module, the user can identify the accession with the highest similarity to their accession by uploading the genotype of the local sample.

In summary, BnIR provides abundant and comprehensive multi-omics data resources and convenient analysis tools. As a high-efficiency data analysis platform, BnIR allows direct utilization of these data and tools to mine genetic loci, genes and variations associated with specific traits. Furthermore, BnIR can also facilitate understanding of the mechanisms for gene expression regulation and phenotypic formation in *B. napus* genome from different omics dimensions.

Case study 1: Mining of genes associated with seed glucosinolate content.

SGC is an important trait related to the oil quality of rapeseed (Gupta and Pratap, 2007; Wang *et al.*, 2018; Lu *et al.*, 2019). In order to mine the genes related to SGC, GWAS was first performed. By using the GWAS module of the Multi-omics portal, the 'Glucosinolate content' phenotype was selected for browsing significant GWAS signals (Figure 4A). As a result, a total of 20 genetic loci distributed in 10 chromosomes were identified to be associated with SGC (Figure 4B). Among them, *qGLS.C09.8* containing *BnaC09.MYB28a* and *BnaC09.MYB28b* was the most significant locus with the strongest effect (Figure 4C). *BnaC09.MYB28a* and *BnaC09.MYB28b* are members of

252 the large family of R2R3-MYB transcription factors and have been identified as
253 regulators of glucosinolate biosynthetic genes in *B. napus* (Harper et al., 2012; Wang *et*
254 *al.*, 2018). Notably, the most significant variations were distributed around
255 *BnaC09.MYB28a* and *BnaC09.MYB28b* and showed strong linkage disequilibrium (LD)
256 (Figure 4D). To further determine whether these two *BnaC09.MYB28s* are related to
257 SGC, TWAS and co-localization analysis were performed. By using the TWAS module
258 to browse the results, *BnaC09.MYB28a* and *BnaC09.MYB28b* were both found to be
259 significantly associated with SGC, suggesting that their expression levels can
260 potentially influence SGC. Then, the COLOC module was used to further test whether
261 the GWAS locus and the *cis*-eQTLs of *BnaC09.MYB28a* and *BnaC09.MYB28b* are co-
262 localized (Figure 4E). The two *BnaC09.MYB28s* and the 'Glucosinolate content'
263 phenotype were selected to query their co-localization analysis results (Figure 4E). As
264 a result, they shared the same causal variant (PPH_4 = 0.99), suggesting that
265 *BnaC09.MYB28a* and *BnaC09.MYB28b* are candidate genes associated with SGC
266 (Figure 4F and G). In addition, frameshift mutation (GCTA/-) may be the casual
267 variation, which has been analyzed in our previous research (Wang *et al.*, 2018; Yang
268 *et al.*, 2022a) (Figure 4F and G).

269 Case study 2: Identification of new genes associated with seed oil 270 content

271 SOC is one of the most important traits of rapeseed (Wang *et al.*, 2018; Tang *et al.*,
272 2021). Here, the published SOC data of 286 *B. napus* accessions were used to validate
273 the power of BnIR in identifying new candidate genes associated with traits (Wang *et*
274 *al.*, 2021). First, five significant loci were identified using GWAS, among which
275 *qOC.C07.1* was the most significant locus. Then, the GWAS module in BnIR was used
276 to browse the GWAS results in this locus (Supplementary Figure 1A). It was found that
277 significant variations were mainly distributed near *BnaC07.SRK2D*
278 (BnaC07G0388200ZS) with strong LD (Supplementary Figure 1B). *BnaC07.SRK2D*
279 encodes a member of abscisic acid (ABA)-activated SNF1-related protein kinases

(*SnRK2*), which play important roles in various seed developmental processes such as de-greening, accumulation of seed storage products, seed maturation, desiccation tolerance and germination (Nakashima and Yamaguchi-Shinozaki, 2013; Lin et al., 2021; Ali et al., 2022; Kozaki and Aoyanagi, 2022). Next, we attempted to check whether the expression of *BnaC07.SRK2D* is related to SOC through co-localization analysis. When users enter the COLOC module of the Multi-omics portal and then select the eQTL of *BnaC07.SRK2D* and the QTL of SOC in GWAS (Figure 5A), the co-localization analysis results would be presented on the page (Figure 5B-D), which showed that they shared the same causal variation, suggesting that *BnaC07.SRK2D* is a candidate gene for SOC.

Then, we used the Variation portal to further explore how genetic variation in this locus affects gene expression and ultimately influences the phenotype of the population. When entering 'BnaC07G0388200ZS' in gene box of the Single-locus module of the Variation portal and choosing the 1 kb of flanking region (Figure 5E), we obtained 13 variations in this gene region, including ten variations in the upstream or downstream of the gene and three in the genic region (Figure 5F). Next, we performed a haplotype analysis to examine their effects on the gene expression and phenotype. When selecting these variations in the Multi-locus module (Figure 5G), we obtained two haplotypes with a frequency higher than 0.05 (Figure 5H and I). We found that the accessions with Hap2 (the same as Zhongshuang 11, a representative accession with high SOC; mean SOC, 48.51%) had significantly higher gene expression and SOC than those with Hap1 (mean SOC, 43.22%) (Figure 5H and I). These results indicated that these significant variations contribute to the divergence in the expression of *BnaC07.SRK2D* and phenotype. In other words, the accessions with Hap2 had significantly higher expression levels of *BnaC07.SRK2D*, which eventually contribute to higher SOC. Such findings will provide a valuable reference for increasing SOC in future breeding.

Discussion

In this study, we first collected and processed rapeseed multi-omics data from recent research and databases, and then constructed BnIR, the first rapeseed multi-omics database. BnIR provides comprehensive multi-omics data resources, including genome sequences, annotations and expression levels of genes, epigenetic signals, metabolite contents, traits and association signals. In addition, multiple online analysis and visualization tools were developed and integrated to help researchers to mine genetic loci/genes and understand the regulatory mechanisms of gene expression and phenotypic formation. Two case studies including mining of candidate genes associated with SGC and SOC fully exhibited the power of BnIR in assisting genetic breeding.

Compared with the published rapeseed databases, BnIR has the following characteristics. 1) It integrates and optimizes the data sources and functions of BnPIR, BnTIR and BnVIR, and significantly improves the scale of data source and response speed of the webpage (Figure 6A). 2) To the best of our knowledge, BnIR comprises the most comprehensive multi-omics datasets, including 13 published genome assemblies, genetic variations of 2,311 accessions, transcriptome data from 2,791 libraries, 118 phenotypes, contents of 266 metabolites, and multiple epigenome data involving histone modification, chromatin accessibility and chromatin interaction (Figure 6B). 3) GWAS, eQTL, TWAS, SMR and co-localization analysis are applied to mine the genetic loci, candidate genes and genetic variations, providing important resources, tools and reference for rapeseed breeding (Figure 6B). 4) BnIR provides numerous common online multi-omics data analysis tools, which support all published rapeseed genomes and provide a convenient and efficient analysis platform.

In summary, we provide a valuable database for future rapeseed breeding and functional genomics research. In the future, with the release of more multi-omics datasets, we will update the database annually. In addition, more newly developed high-efficiency multi-omics methods will be applied to mine more candidate loci and genes

associated with traits. These efforts will make BnIR a more useful platform in rapeseed research.

Materials and Methods

Data sources

To obtain comprehensive rapeseed multi-omics datasets, we mined and integrated the data from genomics, transcriptomics, genetic variations, phenotypic data, epigenetics and metabolomics (Figure 1). In total, 29 published *Brassica* genome assemblies (including 14 *B. napus* genome assemblies) and 2,290,519 genes in them were collected. As for the Transcriptome portal, 2,791 RNA-seq libraries of 272 individuals for 23 tissues were collected from published studies and databases (including 91 libraries from BnTIR) (Liu *et al.*, 2021a). As for the phenotypic data, the data of 118 phenotypes of 2,512 accessions were collected from 13 published studies (Figure 2). The epigenetics data are summarized in Download portal of BnIR. Metabolome datasets were retrieved from two previous studies (Geng et al., 2017; Clement et al., 2018). Genetic variation datasets of 2,311 accessions from BnVIR were also integrated into BnIR (Yang *et al.*, 2022a).

Genome annotation

As for three newly collected *B. napus* genome assemblies, the repeat sequences were predicted as previously described and then integrated to the Gbrowse browser of Genome syn module (McKay et al., 2010; Song *et al.*, 2020). The gene families, transcription factors, GO terms and KEGG pathways of all genome assemblies were predicted according to their homology to the *A. thaliana* genome and the annotations of *A. thaliana* downloaded from TAIR (https://www.arabidopsis.org/). GO and KEGG enrichment analyses were performed using the “clusterProfiler” R package (Yu et al., 2012).

Comparative genome analysis

Genome sequences were compared between ZS11 reference genome and other 11 *B. napus* and two diploid (Z1 and HDEM) genome assemblies using the NUCmer program (v.4.0.0beta2) with parameters "nucmer --mum –noextend" in MUMmer4 (Marcais et al., 2018). After filtering of the one-to-one alignments with a minimum alignment

length of 50 bp using the show-diff program from MUMmer4, the remaining alignment blocks were used for genome browser visualization. For genome browser visualization, the Dotplot module in JBrowse2 and Genome synteny module in Gbrowser were embedded in BnIR (McKay *et al.*, 2010; Hofmeister and Schmitz, 2018).

A total of 1,185,955 genes from 12 *B. napus* genome assemblies were used to construct the gene clusters. Firstly, protein sequences of every pair from 12 genome assemblies were aligned using DIAMOND (v.0.9.14.115, http://github.com/bbuchfink/diamond). Then, gene synteny was detected by McScan (python version). The genes with synteny were grouped into one cluster. Finally, these genes were grouped into 231,106 gene clusters.

Transcriptome analysis

After clipping adaptor sequences and removing low-quality reads by fastp (v. 0.23.0), the clean RNA-seq data from accessions were mapped to the ZS11 reference genome using Hisat2 (v.2.1.2) with default parameters (Kim et al., 2015). Gene expression levels were normalized using the number of transcripts per kilobase million reads (TPM) by the StringTie software (v.1.3.5) with default settings (Pertea et al., 2015). The co-expression network was obtained by calculating the Pearson correlation coefficient of pairwise gene expression levels, and the gene modules including the gene pairs with a Pearson correlation coefficient higher than 0.8 were retained as a co-expression network.

Epigenome analysis

The methods used to remove adaptor sequences and filter low-quality reads were the same as those in RNA-seq data analysis. As for ChIP-seq and ATAC-seq, the clean data from accessions were mapped to the ZS11 reference genome using bowtie2 (v.2.3.2) with default parameters (Langmead and Salzberg, 2012). PCR duplicate reads were removed using Picard (v.2.19, http://broadinstitute.github.io/picard/). Peaks were called using the callpeak module of MACS2 software (v.2.1.2) with the parameters " --broad -f BAM -g 1000000000 -B -p 0.00001 --nomodel --extsize 147 " (Feng et al., 2012).

As for Hi-C, the clean reads of each accession were mapped to the ZS11 reference genome using BWA-MEM with default parameters (Li and Durbin, 2009; Song *et al.*, 2020). Then, the Hi-C interaction matrix was created using the Juicer pipeline (Durand et al., 2016). KR-normalized matrix was extracted from Hi-C format files at the resolutions of 10 kb, 50 kb and 100 kb using the Juicer_tools (v.1.7.6) for JBrowser (Rao et al., 2014; Buels et al., 2016), and then compartments were called using Juicer_tools at the resolution of 100 kb.

As for the BS-seq data, clean data of each accession were mapped to the ZS11 reference genome using Bismark (v.0.13.0) with parameter settings "-N 1, -L 30" (Krueger and Andrews, 2011). Bigwig files of all epigenome data analysis can be visited by JBrowser in the Tools portal (Buels *et al.*, 2016).

GWAS

The SNPs and InDels with a minor allele frequency (MAF) lower than 0.05 were filtered for GWAS performed for all traits using the GEMMA (v.0.98.1) (Zhou and Stephens, 2012). The population structure was controlled by including the first three principal components as covariates, as well as an IBS kinship matrix derived from all variants (SNPs and InDels) calculated by GEMMA. The cutoff for determining significant associations was set as $-\log_{10}(1/n)$, where n represents the total number of variations. Then, genetic loci and lead variation were identified with the following steps. 1) Two significant variations were regarded as the same locus if the distance between them is less than 0.5 Mb and; 2) As for a genetic locus, the most significant variation associated with the phenotype was identified as the lead SNP.

eQTL mapping

Gene expression values were taken as the values of the phenotype for eQTL mapping. Only those genes expressed in more than 95% of the accessions were defined as expressed genes for eQTL mapping. Variations with MAF > 5% were used to perform GWAS for each gene by using GEMMA to detect the associations for variations and genes (Zhou and Stephens, 2012). The cutoff for determining significant associations

422 was set as $-\log_{10}(1/n)$, where n represents the total number of variations. Based on the
423 distance between eQTLs and targeted-genes, we subdivided all eQTLs into *cis*-eQTLs
424 if the variation was found within 1 Mb of the transcription start site or transcription end
425 site of the target gene, otherwise as *trans*-eQTLs. In BnIR, the regulatory pairs of
426 eQTLs and eGenes are visualized using BioCircos.js (Cui et al., 2016).

427 **TWAS**

428 TWAS was used to integrate GWAS and gene expression datasets to identify gene-trait
429 associations (Gusev *et al.*, 2016). TWAS was conducted by the EMMAX module using
430 the gene expression data of seeds at 20 DAF and 40 DAF with the data of the
431 phenotypes from the same accessions. Models were considered as 'transcriptome-wide
432 significant' if they passed the Bonferroni correction for all genes.

433 **SMR analysis**

434 SMR analysis integrated the summary-level data from GWAS with eQTL data to
435 identify genes associated with a complex trait considering their pleiotropy. The *cis*-
436 eQTL signals of expressed genes and GWAS signals of the phenotype were used to
437 perform SMR analysis and HEIDI test by SMR software (v.1.03) (Gusev *et al.*, 2016).
438 Then, the gene was defined as a candidate gene of the phenotype when the $-\log_{10}$(P-
439 value) of SMR was lower than 1/n (n is the number of all expressed protein-coding
440 genes) and P-value of HEIDI test was higher than 1.57×10^{-3}.

441 **Co-localization analysis**

442 GWAS loci associated with phenotypes and candidate genes in these loci were used for
443 co-localization analysis performed using the "COLOC" R package with default
444 parameters (Hormozdiari et al., 2016). The variants in *cis*-eQTLs of genes and QTLs
445 of phenotypes were defined as co-localized when the posterior probability of a co-
446 localized signal (PPH_4) value was higher than 0.5 and there was at least one shared
447 significant variation.

448 **Implementation**

449 BnIR (http://yanglab.hzau.edu.cn/BnIR) was constructed based on the ThinkPHP

(v.5.0.24) framework with jQuery (v.3.6.0) as the JavaScript library, and runs on the Apache 2 web server (v.2.4.53) with MySQL (v.8.0.29) as its database engine. The database is available online without registration and optimized for Chrome (recommended), Opera, Firefox, Windows Edge and macOS Safari.

DATA AVAILABILITY

Sources of all datasets are described in supplemental materials and methods. All datasets are made available at http://yanglab.hzau.edu.cn/BnIR/download.

FUNDING

This research was supported by the National Natural Science Foundation of China (32070559), the National Key Research and Development Plan of China (2021YFF1000100), China Postdoctoral Science Foundation (2022M710875), Hubei Hongshan Laboratory (2021HSZD004) and Developing Bioinformatics Platform in Hainan Yazhou Bay Seed Lab (No. JBGS-B21HJ0001).

Conflict of interest statement. None declared.

ACKNOWLEDGEMENTS

We thank bioinformatics computing platform of the National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, managed by Hao Liu.

Author contributions: Qing-Yong Yang designed the project. Qing-Yong Yang and Zhiquan Yang managed and coordinated the project. Zhiquan Yang, Shengbo Wang, Yiming Huang, Chengfang Luo, Yupeng Jia, Yuchen Lin and Yue Hu collected datasets and performed the bioinformatics analysis. Dongxu Liu and Congyuan Liang designed the database; Shengbo Wang, Yiming Huang, Lulu Wei and Dongxu Liu constructed the database. Qing-Yong Yang, Zhiquan Yang, Shengbo Wang and Yiming Huang wrote the manuscript. Qing-Yong Yang, Zhiquan Yang, Dongxu Liu and Congyuan Liang directed the project.

Reference

Ali, F., Qanmber, G., Li, F., and Wang, Z. (2022). Updated role of ABA in seed maturation, dormancy, and germination. J. Adv. Res. **35**:199-214.

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., et al. (2016). JBrowse: a dynamic web platform for genome visualization and analysis. Genome Biol. **17**:66.

Bus, A., Korber, N., Snowdon, R.J., and Stich, B. (2011). Patterns of molecular variation in a species-wide germplasm set of *Brassica napus*. TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik **123**:1413-1423.

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I.A., Tang, H., Wang, X., Chiquet, J., Belcram, H., Tong, C., Samans, B., et al. (2014). Plant genetics. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. Science **345**:950-953.

Chen, L., Wan, H., Qian, J., Guo, J., Sun, C., Wen, J., Yi, B., Ma, C., Tu, J., Song, L., et al. (2018). Genome-wide association study of cadmium accumulation at the seedling stage in rapeseed (*Brassica napus* L.). Front. Plant Sci. **9**:375.

Clement, G., Moison, M., Soulay, F., Reisdorf-Cren, M., and Masclaux-Daubresse, C. (2018). Metabolomics of laminae and midvein during leaf senescence and source-sink metabolite management in *Brassica napus* L. leaves. J. Exp. Bot. **69**:891-903.

Cui, Y., Chen, X., Luo, H., Fan, Z., Luo, J., He, S., Yue, H., Zhang, P., and Chen, R. (2016). BioCircos.js: an interactive Circos JavaScript library for biological data visualization on web applications. Bioinformatics **32**:1740-1742.

Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016). Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. **3**:95-98.

Edwards, D., and Batley, J. (2004). Plant bioinformatics: from genome to phenome. Trends Biotechnol. **22**:232-237.

Feng, J., Liu, T., Qin, B., Zhang, Y., and Liu, X.S. (2012). Identifying ChIP-seq enrichment using MACS. Nat. Protoc. **7**:1728-1740.

Fernie, A.R., and Gutierrez-Marcos, J. (2019). From genome to phenome: genome-wide association studies and other approaches that bridge the genotype to phenotype gap. Plant J. **97**:5-7.

Friedt, W., Tu, J., and Fu, T. (2018). Academic and economic importance of *Brassica napus* rapeseed. In The Brassica napus Genome, S. Liu and R. Snowdon and B. Chalhoub, eds. (Springer International Publishing: Cham), pp. 1-20.

Geng, S., Yu, B., Zhu, N., Dufresne, C., and Chen, S. (2017). Metabolomics and proteomics of *Brassica napus* guard cells in response to Low CO_2. Front. Mol. Biosci. **4**:51.

Goring, H.H., Curran, J.E., Johnson, M.P., Dyer, T.D., Charlesworth, J., Cole, S.A., Jowett, J.B., Abraham, L.J., Rainwater, D.L., Comuzzie, A.G., et al. (2007).

Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. Nat. Genet. **39**:1208-1216.

Grant, D., Nelson, R.T., Cannon, S.B., and Shoemaker, R.C. (2010). SoyBase, the USDA-ARS soybean genetics and genomics database. Nucleic Acids Res. **38**:D843-846.

Gui, S., Yang, L., Li, J., Luo, J., Xu, X., Yuan, J., Chen, L., Li, W., Yang, X., Wu, S., et al. (2020). ZEAMAP, a Comprehensive Database Adapted to the Maize Multi-Omics Era. iScience **23**:101241.

Gupta, S.K., and Pratap, A. (2007). History, Origin, and Evolution. In Advances in Botanical Research, (Academic Press: pp. 1-20.

Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B.W.J.H., Jansen, R., de Geus, E.J.C., Boomsma, D.I., Wright, F.A., et al. (2016). Integrative approaches for large-scale transcriptome-wide association studies. Nat. Genet. **48**:245-252.

Harper, A.L., Trick, M., Higgins, J., Fraser, F., Clissold, L., Wells, R., Hattori, C., Werner, P., and Bancroft, I. (2012). Associative transcriptomics of traits in the polyploid crop species *Brassica napus*. Nat. Biotechnol. **30**:798-802.

He, Y., Yang, Z., Tang, M., Yang, Q.Y., Zhang, Y., and Liu, S. (2022). Enhancing canola breeding by editing a glucosinolate transporter gene lacking natural variation. Plant Physiol. **188**:1848-1851.

Hofmeister, B.T., and Schmitz, R.J. (2018). Enhanced JBrowse plugins for epigenomics data visualization. BMC Bioinformatics **19**:159.

Hormozdiari, F., van de Bunt, M., Segre, A.V., Li, X., Joo, J.W.J., Bilow, M., Sul, J.H., Sankararaman, S., Pasaniuc, B., and Eskin, E. (2016). Colocalization of GWAS and eQTL signals detects target genes. Am. J. Hum. Genet. **99**:1245-1260.

Hu, J., Chen, B., Zhao, J., Zhang, F., Xie, T., Xu, K., Gao, G., Yan, G., Li, H., Li, L., et al. (2022). Genomic selection and genetic architecture of agronomic traits during modern rapeseed breeding. Nat. Genet. **54**:694-704.

Ichihashi, Y., Date, Y., Shino, A., Shimizu, T., Shibata, A., Kumaishi, K., Funahashi, F., Wakayama, K., Yamazaki, K., Umezawa, A., et al. (2020). Multi-omics analysis on an agroecosystem reveals the significant role of organic nitrogen to increase agricultural crop yield. Proc. Natl. Acad. Sci. U. S. A. **117**:14552-14560.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods **12**:357-360.

Kittipol, V., He, Z., Wang, L., Doheny-Adams, T., Langer, S., and Bancroft, I. (2019). Data in support of genetic architecture of glucosinolate variations in *Brassica napus*. Data Brief **25**:104402.

Korber, N., Bus, A., Li, J., Parkin, I.A., Wittkop, B., Snowdon, R.J., and Stich, B. (2016). Agronomic and deed quality traits dissected by genome-wide association mapping in *Brassica napus*. Front. Plant Sci. **7**:386.

Korber, N., Bus, A., Li, J., Higgins, J., Bancroft, I., Higgins, E.E., Parkin, I.A., Salazar-Colqui, B., Snowdon, R.J., and Stich, B. (2015). Seedling development

traits in *Brassica napus* examined by gene expression analysis and association mapping. BMC Plant Biol. **15**:136.

Kozaki, A., and Aoyanagi, T. (2022). Molecular aspects of seed development controlled by gibberellins and abscisic acids. Int. J. Mol. Sci. **23**:1876.

Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics **27**:1571-1572.

Kudo, T., Kobayashi, M., Terashima, S., Katayama, M., Ozaki, S., Kanno, M., Saito, M., Yokoyama, K., Ohyanagi, H., Aoki, K., et al. (2017). TOMATOMICS: A Web Database for Integrated Omics Information in Tomato. Plant & cell physiology **58**:e8.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods **9**:357-359.

Li, D., Huang, Z., Song, S., Xin, Y., Mao, D., Lv, Q., Zhou, M., Tian, D., Tang, M., Wu, Q., et al. (2016). Integrated analysis of phenome, genome, and transcriptome of hybrid rice uncovered multiple heterosis-related loci for yield increase. Proc. Natl. Acad. Sci. U. S. A. **113**:E6026-E6035.

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

Li, T., Bian, J., Tang, M., Shangguan, H., Zeng, Y., Luo, R., Sun, H., Ke, Q., Nie, X., Li, Y., et al. (2022). BGFD: an integrated multi-omics database of barley gene families. BMC Plant Biol. **22**:454.

Lin, Z., Li, Y., Wang, Y., Liu, X., Ma, L., Zhang, Z., Mu, C., Zhang, Y., Peng, L., Xie, S., et al. (2021). Initiation and amplification of SnRK2 activation in abscisic acid signaling. Nat. Commun. **12**:2456.

Liu, D., Yu, L., Wei, L., Yu, P., Wang, J., Zhao, H., Zhang, Y., Zhang, S., Yang, Z., Chen, G., et al. (2021a). BnTIR: an online transcriptome platform for exploring RNA-seq libraries for oil crop Brassica napus. Plant Biotechnol J **19**:1895-1897.

Liu, H., Wang, J., Zhang, B., Yang, X., Hammond, J.P., Ding, G., Wang, S., Cai, H., Wang, C., Xu, F., et al. (2021b). Genome-wide association study dissects the genetic control of plant height and branch number in response to low-phosphorus stress in *Brassica napus*. Ann. Bot. **128**:919-930.

Liu, H., Li, X., Zhang, Q., Yuan, P., Liu, L., King, G.J., Ding, G., Wang, S., Cai, H., Wang, C., et al. (2021c). Integrating a genome-wide association study with transcriptomic data to predict candidate genes and favourable haplotypes influencing *Brassica napus* seed phytate. DNA Res. **28**:dsab011.

Liu, J., Hua, W., Hu, Z., Yang, H., Zhang, L., Li, R., Deng, L., Sun, X., Wang, X., and Wang, H. (2015). Natural variation in *ARF18* gene simultaneously affects seed weight and silique length in polyploid rapeseed. Proc. Natl. Acad. Sci. U. S. A. **112**:E5123-5132.

Liu, S., Huang, H., Yi, X., Zhang, Y., Yang, Q., Zhang, C., Fan, C., and Zhou, Y. (2020). Dissection of genetic architecture for glucosinolate accumulations in leaves and seeds of *Brassica napus* by genome-wide association study. Plant

Biotechnol. J. **18**:1472-1484.

Lu, K., Wei, L., Li, X., Wang, Y., Wu, J., Liu, M., Zhang, C., Chen, Z., Xiao, Z., Jian, H., et al. (2019). Whole-genome resequencing reveals *Brassica napus* origin and genetic loci involved in its improvement. Nat. Commun. **10**:1154.

Ma, S., Wang, M., Wu, J., Guo, W., Chen, Y., Li, G., Wang, Y., Shi, W., Xia, G., Fu, D., et al. (2021). WheatOmics: A platform combining multiple omics data to accelerate functional genomics studies in wheat. Mol. Plant **14**:1965-1968.

Marcais, G., Delcher, A.L., Phillippy, A.M., Coston, R., Salzberg, S.L., and Zimin, A. (2018). MUMmer4: A fast and versatile genome alignment system. PLoS Comput. Biol. **14**:e1005944.

McKay, S.J., Vergara, I.A., and Stajich, J.E. (2010). Using the generic synteny browser (GBrowse_syn). Curr. Protoc. Bioinformatics **Chapter 9**:Unit-9.12.

Nakashima, K., and Yamaguchi-Shinozaki, K. (2013). ABA signaling in stress-response and seed development. Plant Cell Rep. **32**:959-970.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. **33**:290-295.

Pisupati, R., Reichardt, I., Seren, U., Korte, P., Nizhynska, V., Kerdaffrec, E., Uzunova, K., Rabanal, F.A., Filiault, D.L., and Nordborg, M. (2017). Verification of Arabidopsis stock collections using SNPmatch, a tool for genotyping high-plexed samples. Sci. Data **4**:170184.

Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell **159**:1665-1680.

Sakai, H., Lee, S.S., Tanaka, T., Numa, H., Kim, J., Kawahara, Y., Wakimoto, H., Yang, C.C., Iwamoto, M., Abe, T., et al. (2013). Rice Annotation Project Database (RAP-DB): an integrative and interactive database for rice genomics. Plant & cell physiology **54**:e6.

Song, J.M., Liu, D.X., Xie, W.Z., Yang, Z., Guo, L., Liu, K., Yang, Q.Y., and Chen, L.L. (2021). BnPIR: *Brassica napus* pan-genome information resource for 1689 accessions. Plant Biotechnol. J **19**:412-414.

Song, J.M., Guan, Z., Hu, J., Guo, C., Yang, Z., Wang, S., Liu, D., Wang, B., Lu, S., Zhou, R., et al. (2020). Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. Nat. Plants **6**:34-45.

Sun, C., Wang, B., Yan, L., Hu, K., Liu, S., Zhou, Y., Guan, C., Zhang, Z., Li, J., Zhang, J., et al. (2016a). Genome-wide association study provides insight into the genetic control of plant height in rapeseed (*Brassica napus* L.). Front. Plant Sci. **7**:1102.

Sun, C., Wang, B., Wang, X., Hu, K., Li, K., Li, Z., Li, S., Yan, L., Guan, C., Zhang, J., et al. (2016b). Genome-wide association study dissecting the genetic architecture underlying the branch angle trait in rapeseed (*Brassica napus* L.). Sci.

Rep. **6**:33673.

Tang, S., Zhao, H., Lu, S., Yu, L., Zhang, G., Zhang, Y., Yang, Q.Y., Zhou, Y., Wang, X., Ma, W., et al. (2021). Genome- and transcriptome-wide association studies provide insights into the genetic basis of natural variation of seed oil content in Brassica napus. Mol. Plant **14**:470-487.

Tuggle, C.K., Clarke, J., Dekkers, J.C.M., Ertl, D., Lawrence-Dill, C.J., Lyons, E., Murdoch, B.M., Scott, N.M., and Schnable, P.S. (2022). The Agricultural Genome to Phenome Initiative (AG2PI): creating a shared vision across crop and livestock research communities. Genome Biol. **23**:3.

Wang, B., Wu, Z., Li, Z., Zhang, Q., Hu, J., Xiao, Y., Cai, D., Wu, J., King, G.J., Li, H., et al. (2018). Dissection of the genetic architecture of three seed-quality traits and consequences for breeding in *Brassica napus*. Plant Biotechnol. J. **16**:1336-1348.

Wang, H., Wang, Q., Pak, H., Yan, T., Chen, M., Chen, X., Wu, D., and Jiang, L. (2021). Genome-wide association study reveals a patatin-like lipase relating to the reduction of seed oil content in *Brassica napus*. BMC Plant Biol. **21**:6.

Wu, D., Liang, Z., Yan, T., Xu, Y., Xuan, L., Tang, J., Zhou, G., Lohwasser, U., Hua, S., Wang, H., et al. (2019). Whole-genome resequencing of a worldwide collection of rapeseed accessions reveals the genetic basis of ecotype divergence. Mol. Plant **12**:30-43.

Wu, L., Han, L., Li, Q., Wang, G., Zhang, H., and Li, L. (2021). Using interactome big data to crack genetic mysteries and enhance future crop breeding. Mol. Plant **14**:77-94.

Xuan, L., Yan, T., Lu, L., Zhao, X., Wu, D., Hua, S., and Jiang, L. (2020). Genome-wide association study reveals new genes involved in leaf trichome formation in polyploid oilseed rape (*Brassica napus* L.). Plant Cell Environ. **43**:675-691.

Yang, Y., Saand, M.A., Huang, L., Abdelaal, W.B., Zhang, J., Wu, Y., Li, J., Sirohi, M.H., and Wang, F. (2021). Applications of multi-omics technologies for crop improvement. Front. Plant Sci. **12**:563953.

Yang, Z., Liang, C., Wei, L., Wang, S., Yin, F., Liu, D., Guo, L., Zhou, Y., and Yang, Q.Y. (2022a). BnVIR: bridging the genotype-phenotype gap to accelerate mining of candidate variations underlying agronomic traits in Brassica napus. Mol. Plant. **15**:779-782.

Yang, Z., Zhang, H., Li, X., Shen, H., Gao, J., Hou, S., Zhang, B., Mayes, S., Bennett, M., Ma, J., et al. (2020). A mini foxtail millet with an Arabidopsis-like life cycle as a C4 model system. Nat. Plants **6**:1167-1178.

Yang, Z., Wang, J., Huang, Y., Wang, S., Wei, L., Liu, D., Weng, Y., Xiang, J., Zhu, Q., Yang, Z., et al. (2022b). CottonMD: a multi-omics database for cotton biological study. Nucleic Acids Res. **51**:D1446-D1456.

Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS **16**:284-287.

Yu, J., Pressoir, G., Briggs, W.H., Vroh Bi, I., Yamasaki, M., Doebley, J.F.,

McMullen, M.D., Gaut, B.S., Nielsen, D.M., Holland, J.B., et al. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. **38**:203-208.

Zhang, G., Peng, Y., Zhou, J., Tan, Z., Jin, C., Fang, S., Zhong, S., Jin, C., Wang, R., Wen, X., et al. (2022a). Genome-wide association studies of salt-alkali tolerance at seedling and mature stages in *Brassica napus*. Front. Plant Sci. **13**:857149.

Zhang, G.F., Zhou, J.Z., Peng, Y., Tan, Z.D., Li, L., Yu, L.Q., Jin, C., Fang, S., Lu, S.P., Guo, L., et al. (2022b). Genome-wide association studies of salt tolerance at seed germination and seedling stages in *Brassica napus*. Front. Plant Sci. **12**:772708.

Zhang, Y., Zhang, H., Zhao, H., Xia, Y., Zheng, X., Fan, R., Tan, Z., Duan, C., Fu, Y., Li, L., et al. (2022c). Multi-omics analysis dissects the genetic architecture of seed coat content in *Brassica napus*. Genome Biol. **23**:86.

Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. Nat. Genet. **44**:821-824.

Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M.R., Powell, J.E., Montgomery, G.W., Goddard, M.E., Wray, N.R., Visscher, P.M., et al. (2016). Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat. Genet. **48**:481-487.

Supplemental information

Supplementary Figures

Supplementary Figure 1: Mining the genes associated with SOC using GWAS module.

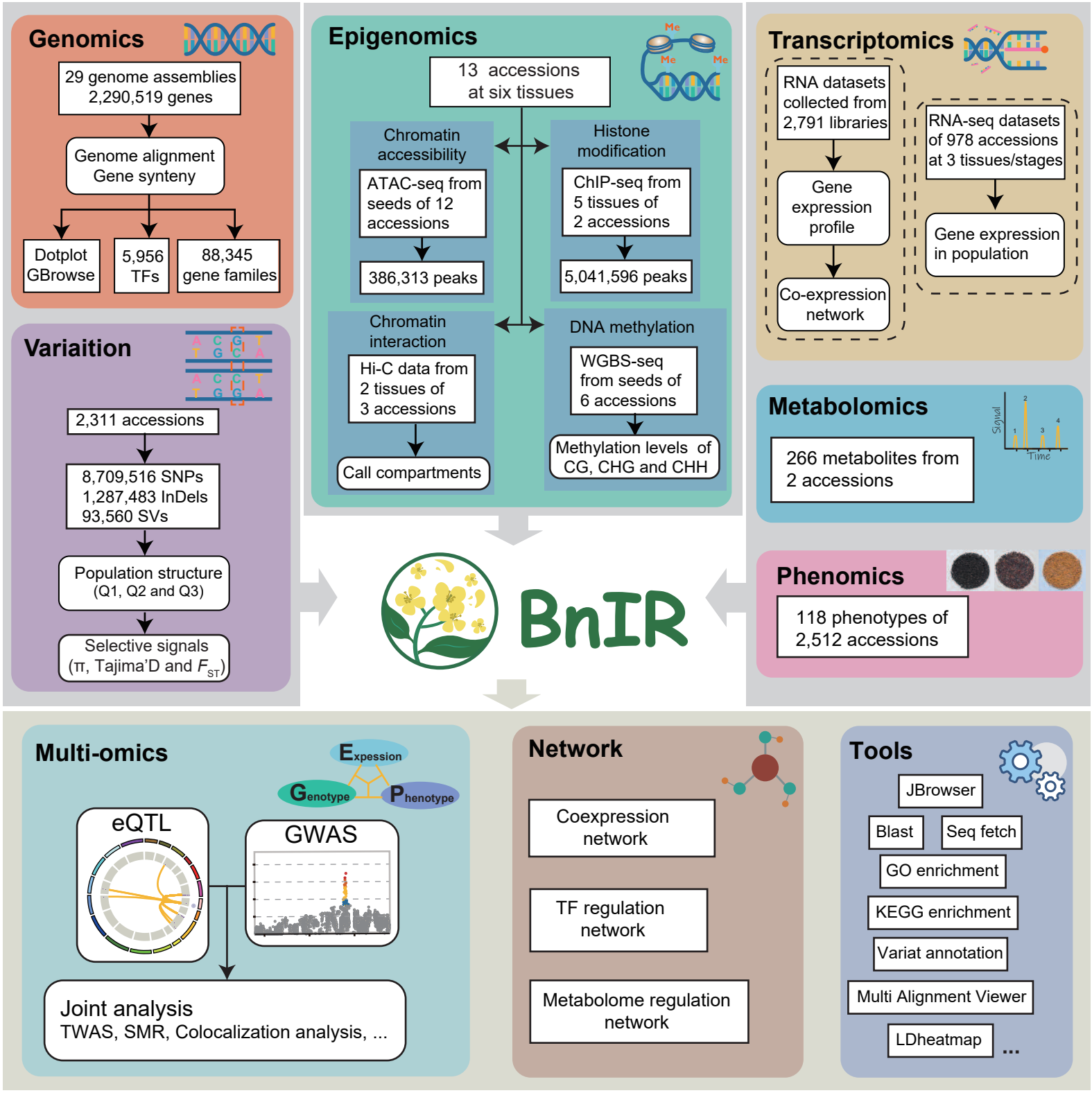


Figure 1. Construction pipelines and overview of BnIR.

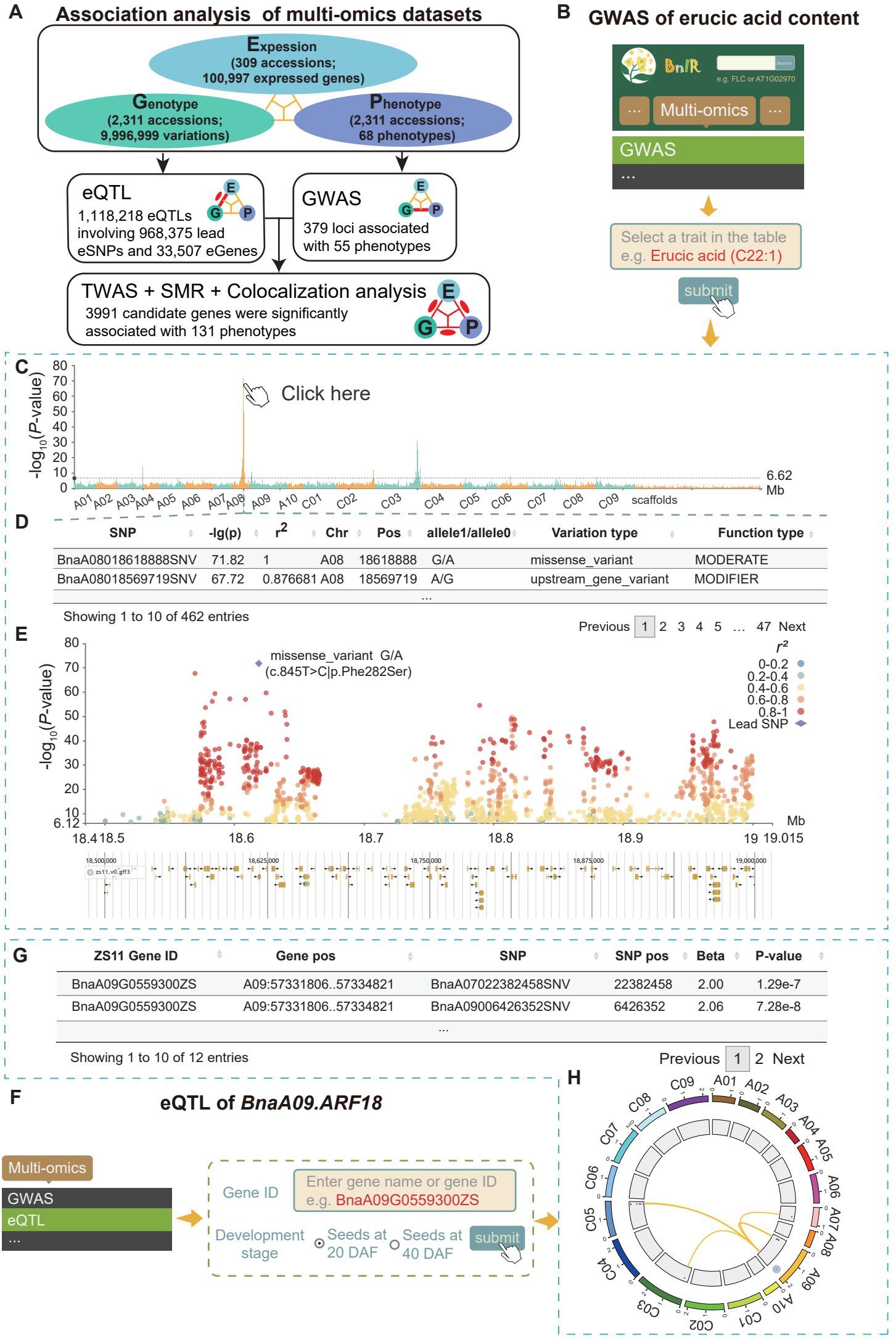

A
Association analysis of multi-omics datasets
Expression
(309 accessions;
100,997 expressed genes)
Genotype
(2,311 accessions;
9,996,999 variations)
Phenotype
(2,311 accessions;
68 phenotypes)
eQTL
1,118,218 eQTLs
involving 968,375 lead
eSNPs and 33,507 eGenes
GWAS
379 loci associated
with 55 phenotypes
TWAS + SMR + Colocalization analysis
3991 candidate genes were significantly
associated with 131 phenotypes
B
GWAS of erucic acid content
BnIR
e.g. FLC or AT1G02970
Multi-omics
GWAS
Select a trait in the table
e.g. Erucic acid (C22:1)
submit
C
-log10(P-value)
Click here
6.62
Mb
A01 A02 A03 A04 A05 A06 A07 A08 A09 A10 C01 C02 C03 C04 C05 C06 C07 C08 C09 scaffolds
D
SNP | -lg(p) | r2 | Chr | Pos | allele1/allele0 | Variation type | Function type
BnaA08018618888SNV | 71.82 | 1 | A08 | 18618888 | G/A | missense_variant | MODERATE
BnaA08018569719SNV | 67.72 | 0.876681 | A08 | 18569719 | A/G | upstream_gene_variant | MODIFIER
...
Showing 1 to 10 of 462 entries
Previous 1 2 3 4 5 ... 47 Next
E
missense_variant G/A
(c.845T>C|p.Phe282Ser)
r2
0-0.2
0.2-0.4
0.4-0.6
0.6-0.8
0.8-1
Lead SNP
-log10(P-value)
6.12
18.4 18.5 18.6 18.7 18.8 18.9 19 19.015
Mb
G
ZS11 Gene ID | Gene pos | SNP | SNP pos | Beta | P-value
BnaA09G0559300ZS | A09:57331806..57334821 | BnaA07022382458SNV | 22382458 | 2.00 | 1.29e-7
BnaA09G0559300ZS | A09:57331806..57334821 | BnaA09006426352SNV | 6426352 | 2.06 | 7.28e-8
...
Showing 1 to 10 of 12 entries
Previous 1 2 Next
F
eQTL of BnaA09.ARF18
Multi-omics
GWAS
eQTL
...
Gene ID
Enter gene name or gene ID
e.g. BnaA09G0559300ZS
Development stage
Seeds at 20 DAF
Seeds at 40 DAF
submit
H
C09 A01 A02 A03 A04 A05 A06 A07 A08 A09 A10 C01 C02 C03 C04 C05 C06 C07 C08

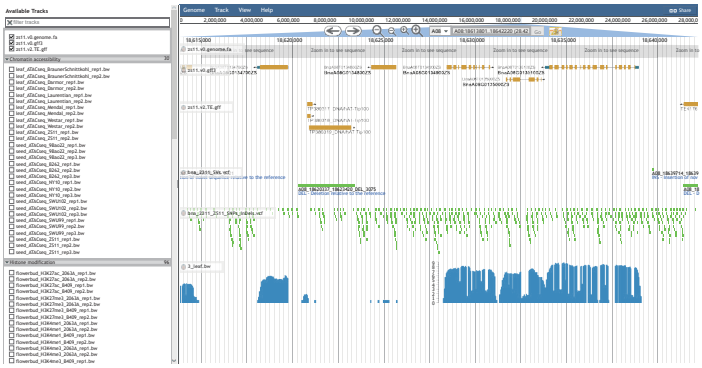
Figure 2 Integration of multi-omics data. (A) Pipeline of integrating population-level multi-omics datasets. **(B-E)** Usage of the GWAS module. **(B)** Pipeline of searching the GWAS signals of seed erucic acid content. **(C)** Global GWAS signals of seed erucic acid content. **(D)** List of significant variations in A08:18500000–19000000. **(E)** Local Manhattan plot of A08:18500000-19000000. Purple diamond represents the lead variation. Green to red dots indicate weak to strong linkage disequilibrium coefficients with the lead variation. **(F-H)** Usage of eQTL module. **(F)** Pipeline of searching the eQTL signals of *BnaA09.ARF18*. **(G)** List of eSNPs of *BnaA09.ARF18*. **(H)** Circos plot of the global eQTLs of *BnaA09.ARF18*. Purple dot in the outer circle represents the genomic position of *BnaA09.ARF18*. Purple dots in the median circle represent the *P*-values of eSNPs. Yellow lines represent the associations between eSNPs and *BnaA09.ARF18*.

Tools

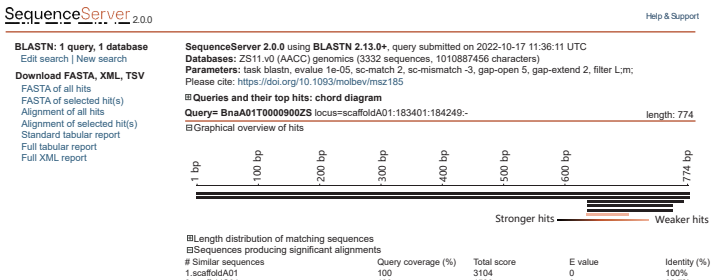
A: JBrowse B: BLAST C: MSA D: GO E: KEGG F: COLOC G: SMR H: SNPmatch I: Primer3 J: e-PCR

Seq fetch, Heatmap, LDheatmap, Data2geomap, Data2heatmap, Domain search, ...

A: JBrowse

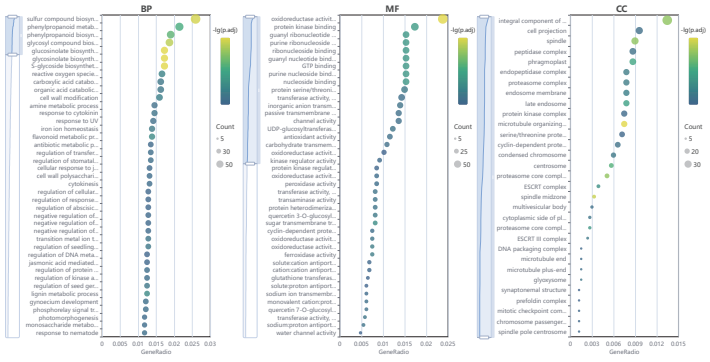


B: BLAST

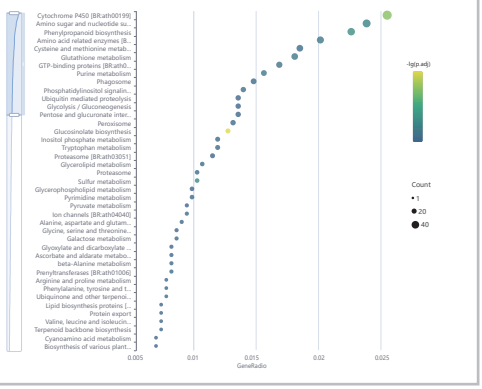


C: MSA

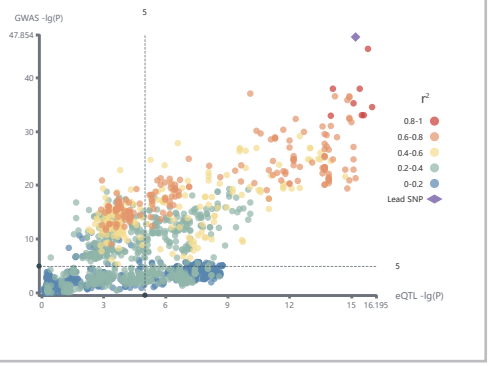
D: GO enrichment



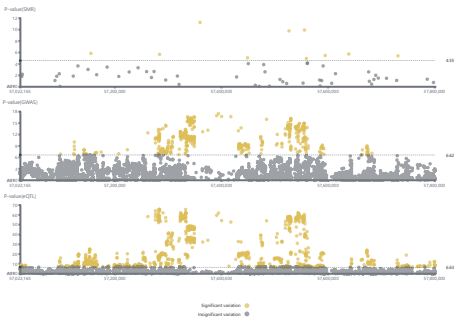
E: KEGG enrichment



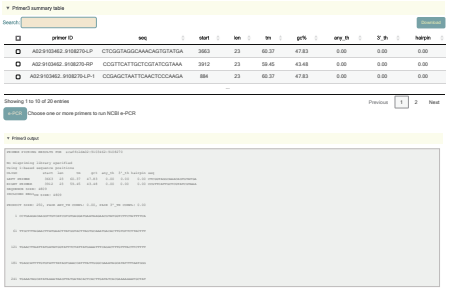
F: COLOC



G: SMR



H: SNPmatch



I: Primer3

J: e-PCR

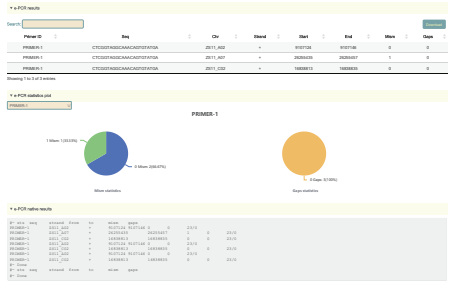
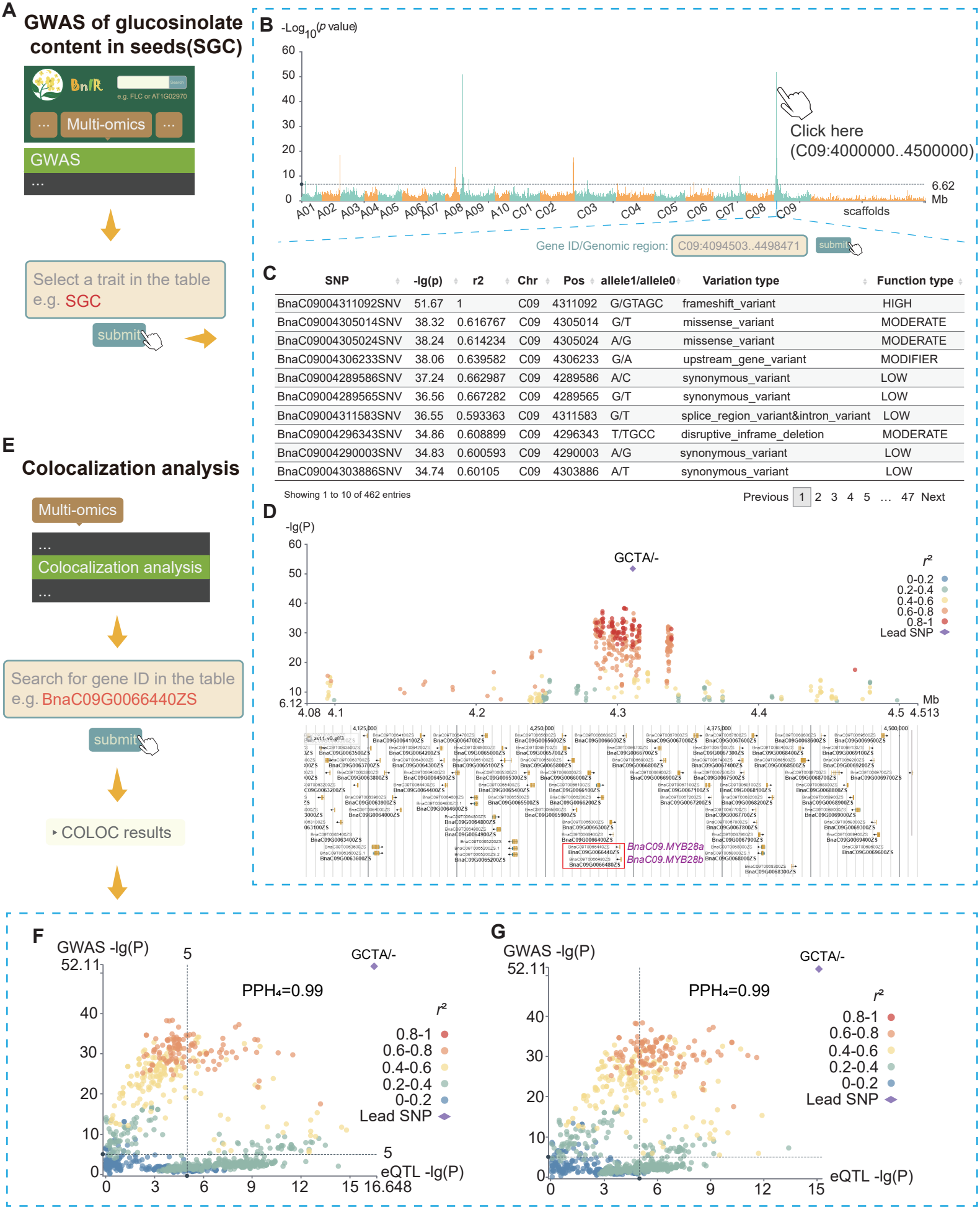


Figure 3 Multi-omics tools in the Tools portal of BnIR. A total of 17 tools were integrated into the Tools portal, including Jbrowse **(A)**, BLAST **(B)**, MSA **(C)**, GO enrichment analysis **(D)**, KEGG enrichment analysis **(E)**, co-localization analysis (COLOC) **(F)**, SMR **(G)**, germplasm identification tool SNPmatch **(H)**, primer design tool Primer3 **(I)** and electronic PCR tool e-PCR **(J)**.



SNP	-lg(p)	r2	Chr	Pos	allele1/allele0	Variation type	Function type
BnaC09004311092SNV	51.67	1	C09	4311092	G/GTAGC	frameshift_variant	HIGH
BnaC09004305014SNV	38.32	0.616767	C09	4305014	G/T	missense_variant	MODERATE
BnaC09004305024SNV	38.24	0.614234	C09	4305024	A/G	missense_variant	MODERATE
BnaC09004306233SNV	38.06	0.639582	C09	4306233	G/A	upstream_gene_variant	MODIFIER
BnaC09004289586SNV	37.24	0.662987	C09	4289586	A/C	synonymous_variant	LOW
BnaC09004289565SNV	36.56	0.667282	C09	4289565	G/T	synonymous_variant	LOW
BnaC09004311583SNV	36.55	0.593363	C09	4311583	G/T	splice_region_variant&intron_variant	LOW
BnaC09004296343SNV	34.86	0.608899	C09	4296343	T/TGCC	disruptive_inframe_deletion	MODERATE
BnaC09004290003SNV	34.83	0.600593	C09	4290003	A/G	synonymous_variant	LOW
BnaC09004303886SNV	34.74	0.60105	C09	4303886	A/T	synonymous_variant	LOW

Figure 4 Case study of mining the genes associated with SGC. (A) Pipeline of searching SGC in GWAS module. **(B)** Genome-wide Manhattan plot of SGC in the GWAS module. **(C)** List of 402 significant variations associated with SGC identified by GWAS. **(D)** Local Manhattan plot of SGC at C09: 4.0–4.5 Mb. Two purple texts represent *BnaC09.MYB28a* and *BnaC09.MYB28b*. **(E)** Pipeline of searching the co-localization analysis results of *BnaC09.MYB28*s in BnIR. **(F-G)** Co-localization analysis of *cis*-eQTLs of two *BnaC09.MYB28*s and GWAS loci. *Cis*-eQTL signals of *BnaC09.MYB28a* **(F)** and *BnaC09.MYB28b* **(G)** co-localized with the *qGLS.C09.8* locus in GWAS. In **(D)**, **(F)** and **(G)**, purple diamonds represent the lead variations (one 4-bp frameshift InDel of GCTA/-), and green to red dots indicate weak to strong linkage disequilibrium coefficients with the lead variation.

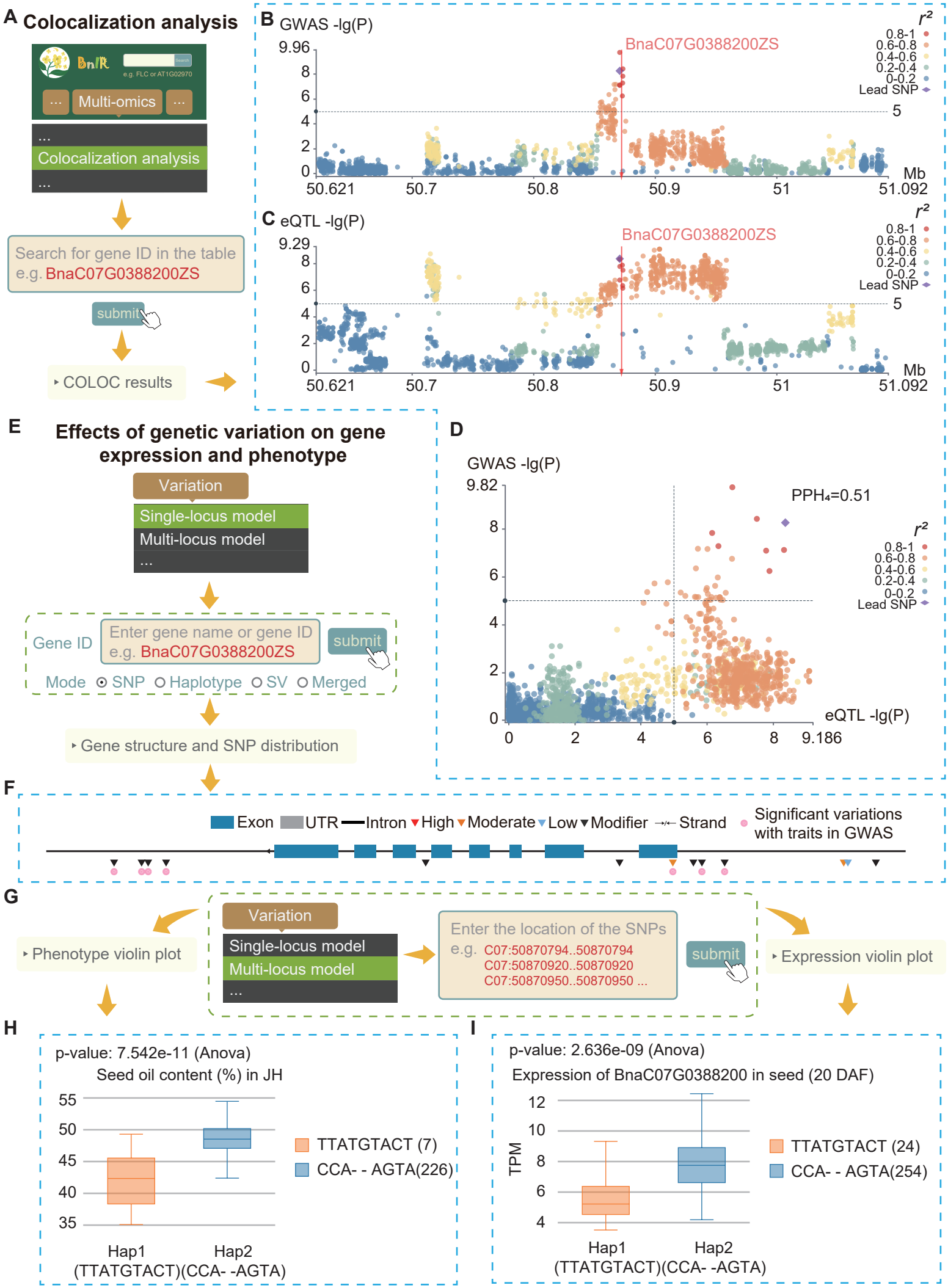

A Colocalization analysis
BnIR
e.g. FLC or AT1G02970
Multi-omics
Colocalization analysis
Search for gene ID in the table
e.g. BnaC07G0388200ZS
submit
COLOC results
B GWAS -lg(P)
BnaC07G0388200ZS
r²
0.8-1
0.6-0.8
0.4-0.6
0.2-0.4
0-0.2
Lead SNP
Mb
50.621
50.7
50.8
50.9
51
51.092
C eQTL -lg(P)
9.29
E Effects of genetic variation on gene expression and phenotype
Variation
Single-locus model
Multi-locus model
Gene ID
Enter gene name or gene ID
e.g. BnaC07G0388200ZS
submit
Mode SNP Haplotype SV Merged
Gene structure and SNP distribution
D
GWAS -lg(P)
9.82
PPH4=0.51
eQTL -lg(P)
9.186
F
Exon UTR Intron High Moderate Low Modifier Strand
Significant variations with traits in GWAS
G
Phenotype violin plot
Variation
Single-locus model
Multi-locus model
Enter the location of the SNPs
e.g. C07:50870794..50870794
C07:50870920..50870920
C07:50870950..50870950 ...
submit
Expression violin plot
H
p-value: 7.542e-11 (Anova)
Seed oil content (%) in JH
TTATGTACT (7)
CCA- - AGTA(226)
Hap1 (TTATGTACT)
Hap2 (CCA- -AGTA)
I
p-value: 2.636e-09 (Anova)
Expression of BnaC07G0388200 in seed (20 DAF)
TPM
TTATGTACT (24)
CCA- - AGTA(254)
Hap1 (TTATGTACT)
Hap2 (CCA- -AGTA)

Figure 5 Case study of identifying new genes associated with SOC. (A) Pipeline of searching the co-localization analysis results of *BnaC07.SnRK2D* in BnIR. **(B-D)** Co-localization analysis of *cis*-eQTL of *BnaC07.SnRK2D* and GWAS loci. **(B-C)** Local Manhattan plots of SOC in GWAS **(B)** and *BnaC07.SnRK2D* in eQTL mapping **(C)**. **(D)** *Cis*-eQTL signals of *BnaC07.SnRK2D* co-localized with *qOC.C07.1* locus in GWAS. Purple diamond represents the lead variation. Green to red dots indicate weak to strong linkage disequilibrium coefficients with the lead variation. **(E)** Pipeline of searching variations in *BnaC07.SnRK2D* using the Single-locus module. **(F)** Variations in *BnaC07.SnRK2D* and its upstream and downstream variation within 1 kb. Red, orange, blue and black inverted triangles represent variations with high, moderate, low and modifier effects. Pink dots represent significant variations with SOC in GWAS. **(G)** Pipeline of haplotype analysis using the Multi-locus module. **(H-I)** Violin plots of SOC **(H)** and expression levels of *BnaC07.SnRK2D* **(I)** in accessions with Hap1 (orange boxes) and Hap2 (blue boxes).

A

Upgrade and integration of BnPIR, BnTIR and BnVIR

Data source	BnIR		
	Portal	**New dataset**	**Improved function**
BnPIR 9 genome assemblies	Genome	+**20** genome assemblies. Totally, **2,290,519** genes; **88,423** gene index; **21,508** gene clusters; **179** gene families; **58** TF families	Optimized user experience and response speed New funciton: Gene search, Gene family, Gene cluster, Transcription factor.
BnTIR 91 RNA-seq libraries	Transcriptome	+**2,700** RNA-seq libraries, including **192** from ZS11 at 7 hormone treatments, **345** from ZS11 at 6 abiotic stress treatments and **978** from population-level assessions of 3 tissues and **1,185** from other studies	Greatly impoved response speed of eFP (**9.348 s** to **0.0611 s)** New funciton: Meta-library, Population expression, eFP (multi-gene module).
BnVIR 2,311 accessions 10,090,561 variations 21 phenotypes	Population, Variation, Phenotype, Multiomics	+**97** phenotypes of **2,512** accessions	New funciton: Phenotype search, eQTL, TWAS, SMR, Colocalization analysis.

New Portal	New dataset	Function description
Epigenome	**99** ChIP-seq libraries; **8** ATAC-seq libraries; **12** WGBS-seq libraries; **3** Hi-C libraries	Searching the peaks of ChIP-seq and ATAC-seq, the CG, CHG and CHH methylation levels, Hi-C contact frequencies and features; Browse epigenetic signals in Jbrowser
Metabolome	**266** metabolites of **2** accessions	Searching meyabolite contents
Multiomics	GWAS, eQTL, TWAS, SMR and colocalization analysis results using **55** phenotypes of **2,311** accessions	Browse the signals of GWAS, eQTL, TWAS, SMR and colocalization analysis results

B

Comparison of BnIR with the published *B. napus* database

Database	Major function
BnIR	Genome information, gene expression, epigenetic signal, metabolite content, traits, multi-omics association information, analysis and visualization tools.
BrassicaEDB	Comprehensive gene expression profile information and a user-friendly visualization interface.
BnaSNPDB	An interactive web-based platform and set of analytic tools for effcient retrieve and analysis of SNPs among 1,007 accessions.
BnaGVD	A large number of variations and search tools for accelerating studies on the functional genomics and the screening of molecular markers.

Database	Genome	Transcriptome (meta-library)	Transcriptome (population)	Epigenome	Variation	Phenotype	Metabolome
Unit	genome assemblies	libraries	libraries	libraries	accessions	phenotypes	metabolites
BnIR	**29**	**2,791**	**978**	**128**	**2,311**	**118**	**544**
BrassicaEDB	**1**	**837**	**-**	**-**	**-**	**-**	**-**
BnaSNPDB	**-**	**-**	**-**	**-**	**1,007**	**-**	**-**
BnaGVD	**3**	**-**	**-**	**-**	**1,007**	**10**	**-**

Database	Variation-Phenotype	Variation-Expression	Network analysis	Association analysis
BnIR	√	√	√	**GWAS, eQTL, TWAS, SMR, colocalization analysis**
BrassicaEDB	**-**	**-**	**-**	**-**
BnaSNPDB	**-**	**-**	**-**	**-**
BnaGVD	**-**	**-**	**-**	**GWAS**

Figure 6 New features of BnIR. (A) Upgrade of BnIR relative to BnPIR, BnTIR and BnVIR. Red bold texts represent new datasets in BnIR. **(B)** Comparison of BnIR with other *B. napus* omics databases. Green to yellow cells represent fewer to more omics datasets.

Supplemental information

BnIR: a multi-omics collaborative resource and tool platform for *Brassica napus* biological study

Zhiquan Yang[1,2,3,†], Shengbo Wang[1,2,†], Lulu Wei[1,2], Yiming Huang[1,2], Dongxu Liu[1,2], Yupeng Jia[1,2], Chengfang Luo[1,2], Yuchen Lin[1,2], Congyuan Liang[1,2], Yue Hu[1,2], Cheng Dai[1,2], Liang Guo[1,2], Qing-Yong Yang[1,2,*]

[1]National Key Laboratory of Crop Genetic Improvement, Hubei Hongshan Laboratory, Huazhong Agricultural University, Wuhan, 430070, China

[2]Hubei Key Laboratory of Agricultural Bioinformatics, College of Informatics, Huazhong Agricultural University, Wuhan, 430070, China

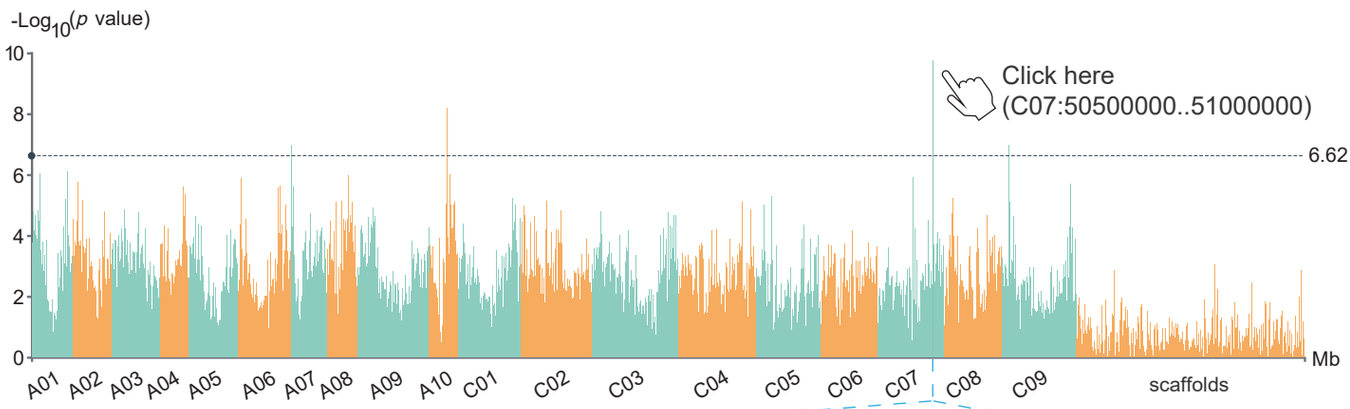
[3]Innovative Center of Molecular Genetics and Evolution, School of Life Sciences, Guangzhou University, Guangzhou, 510405, China

[†]These authors contributed equally: Zhiquan Yang, Shengbo Wang.

[*]Correspondence: Qing-Yong Yang (email: yqy@mail.hzau.edu.cn)

A **GWAS of glucosinolate content in seeds(SGC)**

B



Gene ID/Genomic region: C07:50500000..51000000 submit

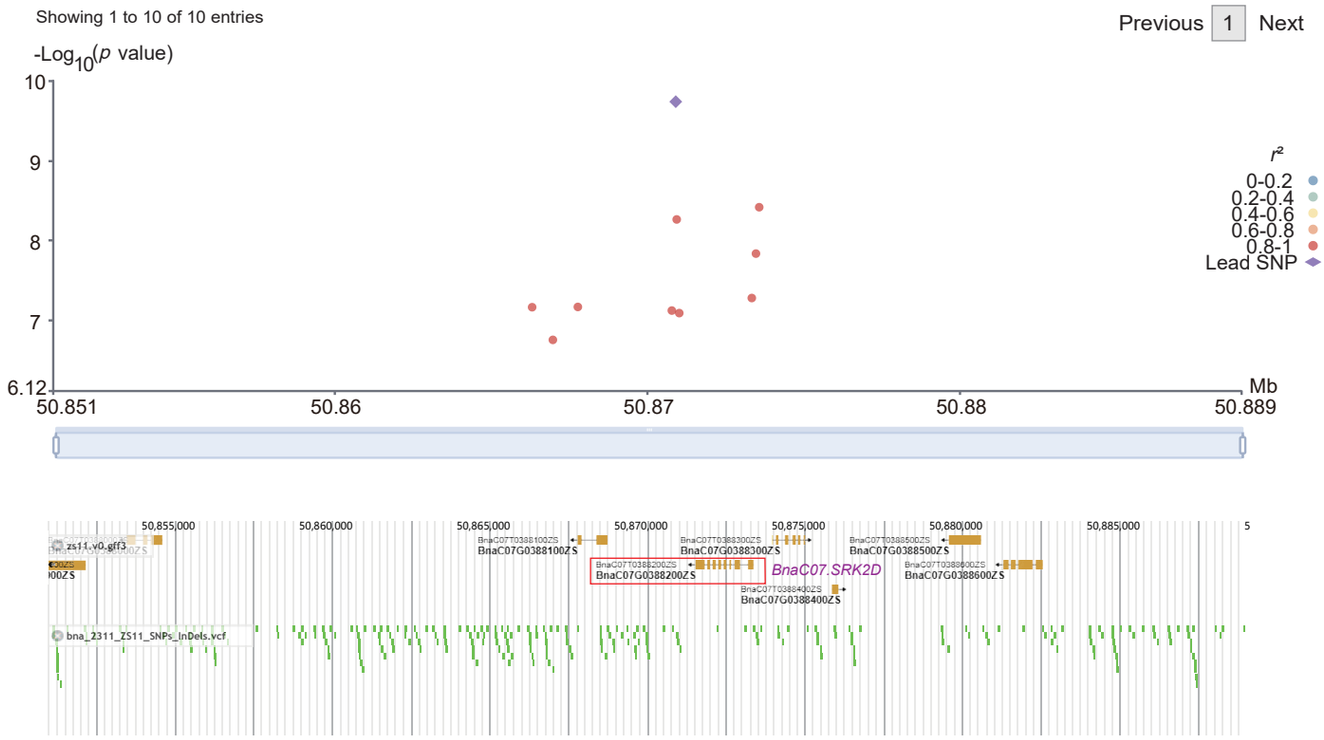
C

SNP	-lg(GWAS_p)	LD_r2 (to lead SNP)	PVE(r2)	Beta	SE	Chr	Pos	allele1/allele0	Variation type	Function type
BnaC07050870920SNV	9.74	1.0000	0.0958	-2.661518	0.4021322	C07	50870920	T/C	upstream_gene_variant	MODIFIER
BnaC07050871032SNV	7.09	0.9411	0.0716	-2.30125	0.4176682	C07	50871032	T/A	upstream_gene_variant	MODIFIER
BnaC07050870950SNV	8.26	0.9406	0.0785	-2.422521	0.4024712	C07	50870950	ATG/A	upstream_gene_variant	MODIFIER
BnaC07050870794SNV	7.12	0.9292	0.0667	-2.220025	0.4018871	C07	50870794	T/C	upstream_gene_variant	MODIFIER
BnaC07050873484SNV	7.83	0.8171	0.0800	-2.445226	0.4189096	C07	50873484	C/T	upstream_gene_variant	MODIFIER
BnaC07050873588SNV	8.42	0.7994	0.0801	-2.586631	0.4252047	C07	50873588	T/A	upstream_gene_variant	MODIFIER
BnaC07050873351SNV	7.28	0.7915	0.0751	-2.472915	0.4420629	C07	50873351	A/G	missense_variant	MODERATE
BnaC07050866327SNV	7.16	0.7075	0.0710	-2.290309	0.4132066	C07	50866327	G/A	downstream_gene_variant	MODIFIER
BnaC07050867789SNV	7.17	0.6877	0.0646	-2.082158	0.3755485	C07	50867789	A/C	synonymous_variant	LOW
BnaC07050866990SNV	6.75	0.6759	0.0655	-2.152067	0.4016859	C07	50866990	A/C	downstream_gene_variant	MODIFIER

Showing 1 to 10 of 10 entries

Previous 1 Next

D



Supplementary Figure 1: Mining the genes associated with SOC using GWAS module. (A) Pipeline of searching SOC in GWAS module. **(B)** Genome-wide Manhattan plot of SOC in the GWAS module. **(C)** List of 10 significant variations associated with SOC identified by GWAS. **(D)** Local Manhattan plot of SOC at C07: 50.5–51.0 Mb. The purple text represents *BnaC07.SRK2D*.