BnVIR Database Tutorial

Overview

BnVIR (*Brassica napus* variation information resource, http://yanglab.hzau.edu.cn/BnVIR) is a comprehensive *Brassica napus* (*B. napus*) variation database, integrating omics data of population of 2,311 *B. napus* accessions for genetic variations query and discovery for candidate variations of traits in *B. napus*. BnVIR includes:

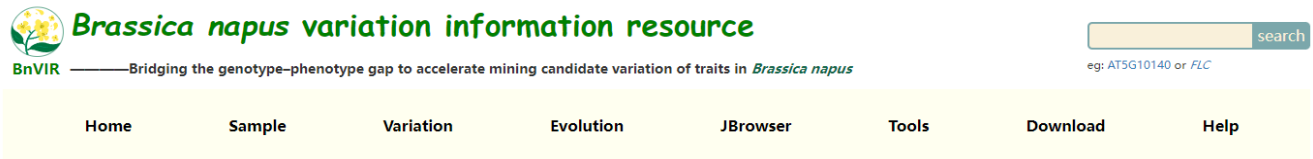
- Basic information of 2,311 *B. napus* germplasms and their re-sequencing data;
- The population structure, pedigrees and the selective signals;
- Genetic variations including single-nucleotide polymorphisms (SNPs), small insertions and deletions (InDels) and large structure variations (SVs) identified from the population;
- The phenotypic and transcriptional effect of the genetic variations and gene-gene interaction;
- GWAS associating genetic variations with phenotype;

BnVIR compromises large genetic variations of *B. napus*, which are mapped to phenotype and gene expression, providing a powerful and convenient platform for users to carry out variation analysis, mine candidate variations/genes of traits and developing valuable markers for molecular breeding in *B. napus.*

1 Homepage

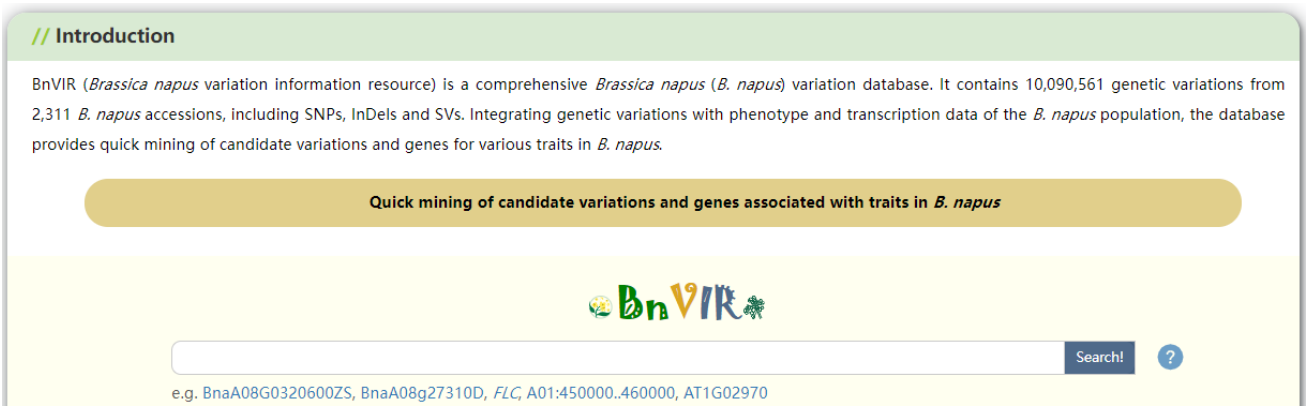
1.1 The Top Menu

The top navigation menu gathers general functions of the database, including links to different modules and tools, and quick search of genetic variations.

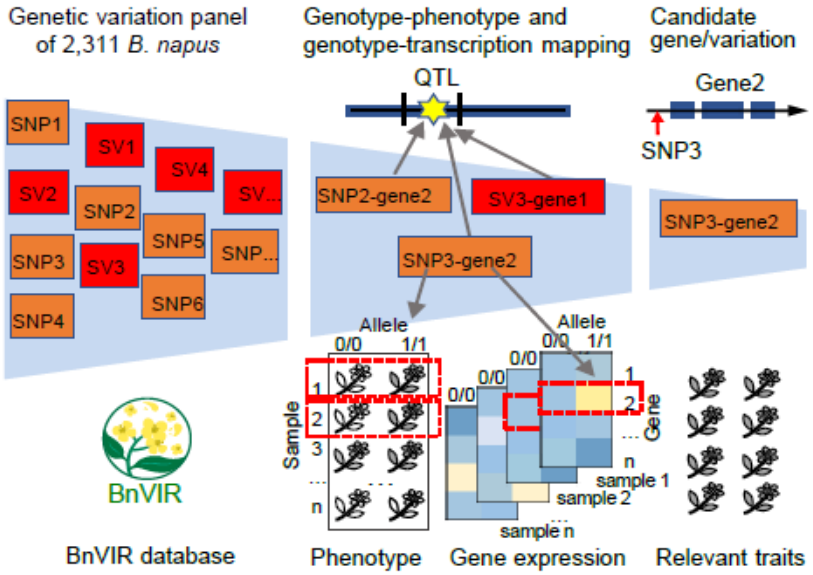


1.2 Brief introduction of BnVIR and quick start to perform candidate variations/genes mining.

For quick search of genetic variations, input of gene ID of *B. napus*, gene ID and name of Arabidopsis and genomic region are supported.

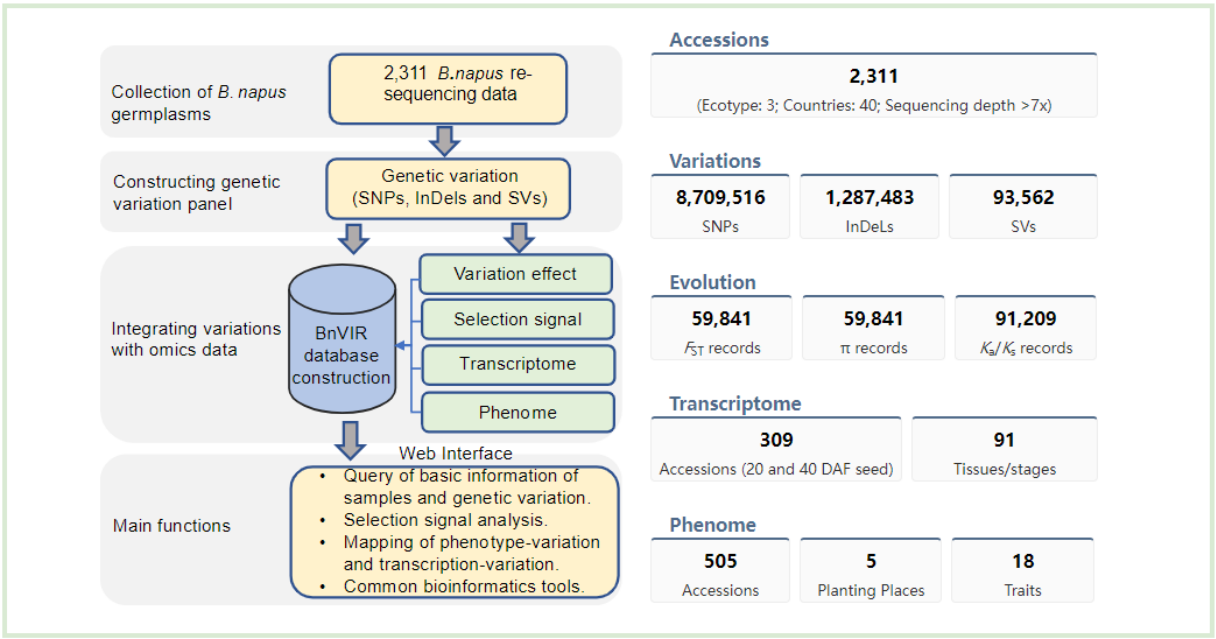


1.3 Schematic diagram of core function of BnVIR.



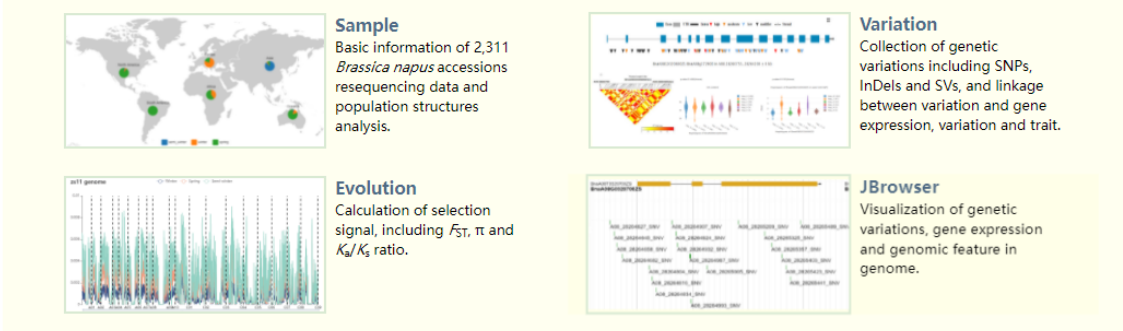
1.4 Dataflow of BnVIR

// Schematic representation



1.5 Quick entry of function module

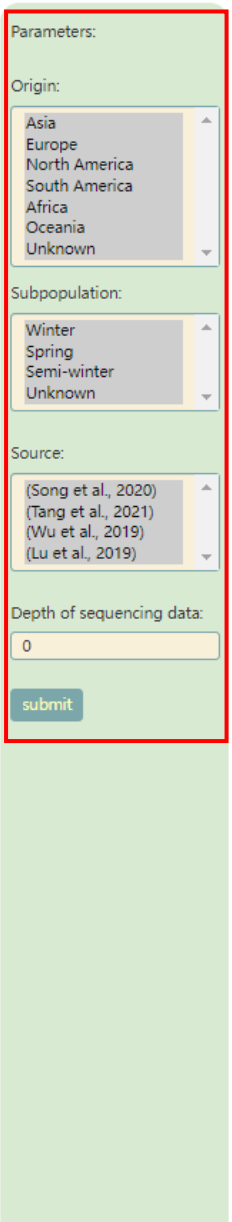
// Quick entry of function modules



2 Sample module

2.1 Accession

Accession page presents basic information of *B. napus* accessions, compromising three parts. All the figures and tables can be download by clicking the button locate in the upper right corner.
Part1: Table of basic information *B. napus* accessions, including. Fuzzy search is supported for users (users can search samples by entering any characters relevant to accessions).
Part2: Geographic distribution of *B. napus* accessions.
Part3: Parameters that can be set to choose the subset of the samples, according origin, ecotype, source and depth of the sequencing data.



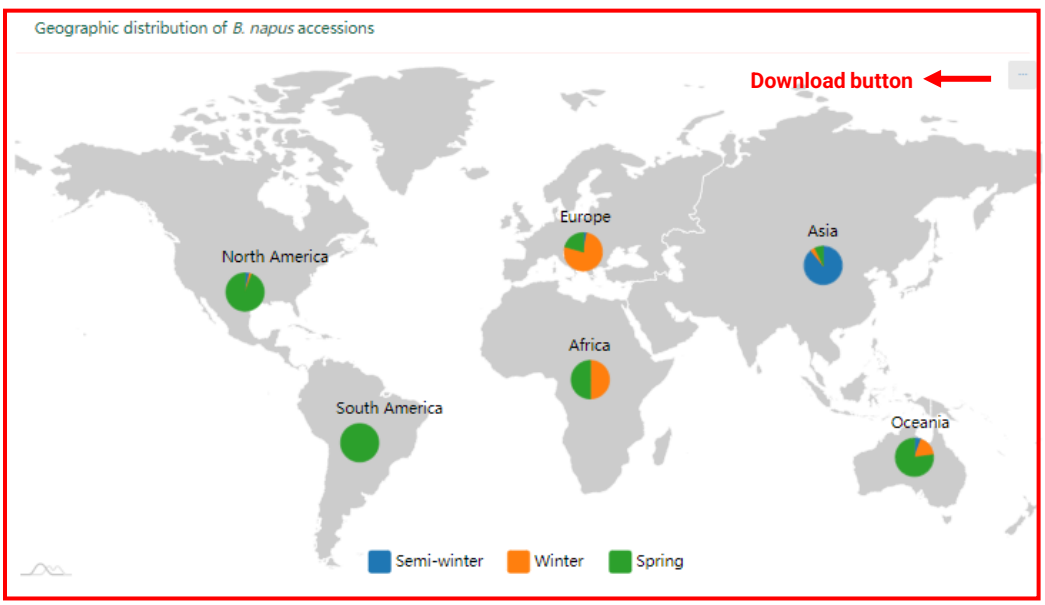
The *B. napus* population in database comprises 2,311 *B. napus* accessions, including 1,259 from Asia, 928 from Europe, 60 from North America, two from South America, 38 from Oceania and four from Africa. Three ecotypes, including spring (354 accessions), winter (756 accessions) and semi-winter (1,122 accessions), were included in the population. "Unknown" indicates that region or ecotype of accession is unknown in the previous reports.

Search: Fuzzy search

Download button ← Download

Accession name	Germplasm	Origin	Subpopulation	Depth	Source
1-C2	Canard	Europe	Unknown	7.00	(Song et al., 2020)
1-C3	Brutor	Europe	Spring	5.00	(Song et al., 2020)
10	Yan81-2	Asia	Semi-winter	10.00	(Tang et al., 2021)
100	Cy18Pxw-62	Asia	Semi-winter	8.00	(Tang et al., 2021)
1000	Y15	Asia	Semi-winter	8.00	(Tang et al., 2021)
1002	Y16	Asia	Semi-winter	8.00	(Tang et al., 2021)
1004	Y17	Asia	Semi-winter	9.00	(Tang et al., 2021)
1006	L508	Asia	Semi-winter	8.00	(Tang et al., 2021)
1010	Ningyou8	Asia	Semi-winter	8.00	(Tang et al., 2021)
1012	Ningyou6	Asia	Semi-winter	9.00	(Tang et al., 2021)

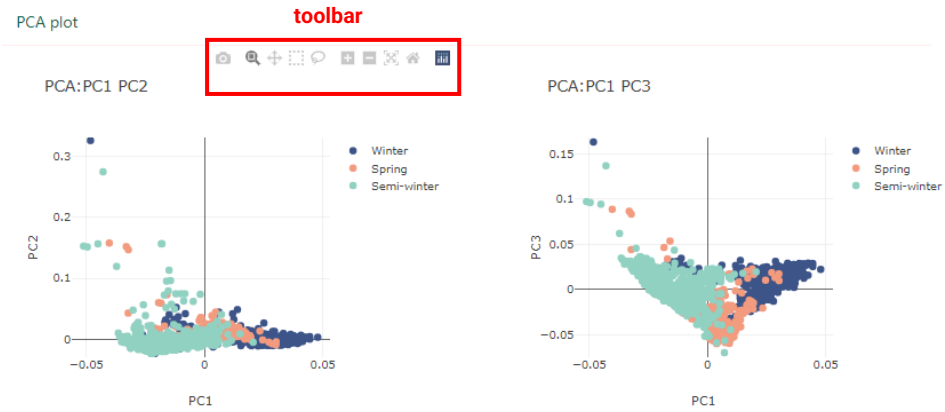
Showing 1 to 10 of 2,311 entries Previous 1 2 3 4 5 ... 232 Next



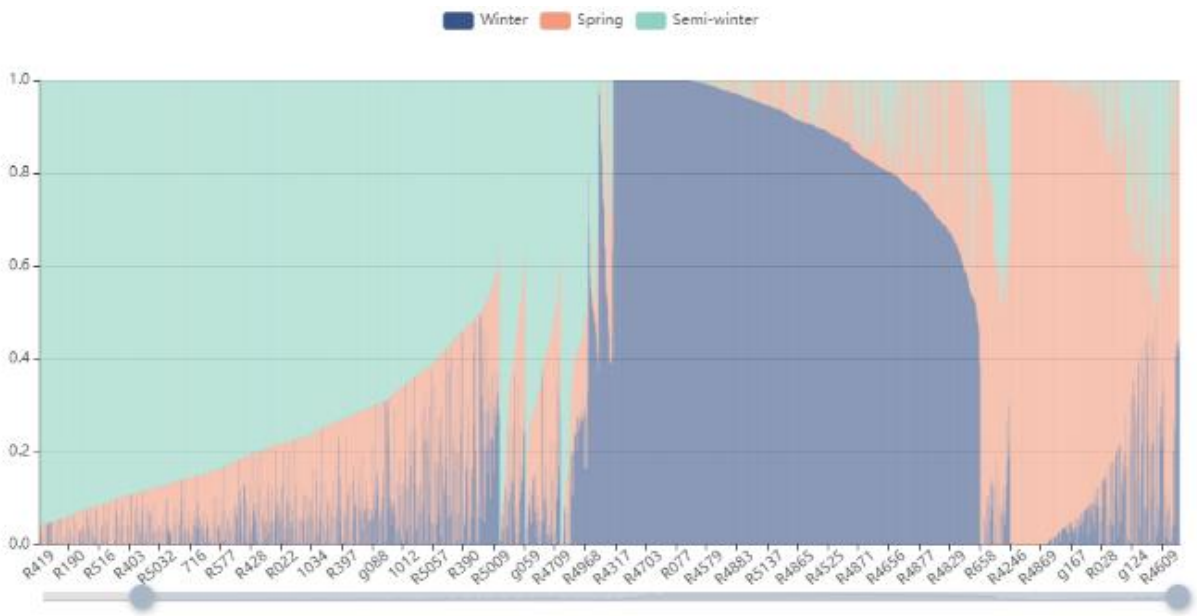
2.2 Population

This page shows the population analysis of the *B. napus* analysis, including PAC, Population structure and phylogenetics.

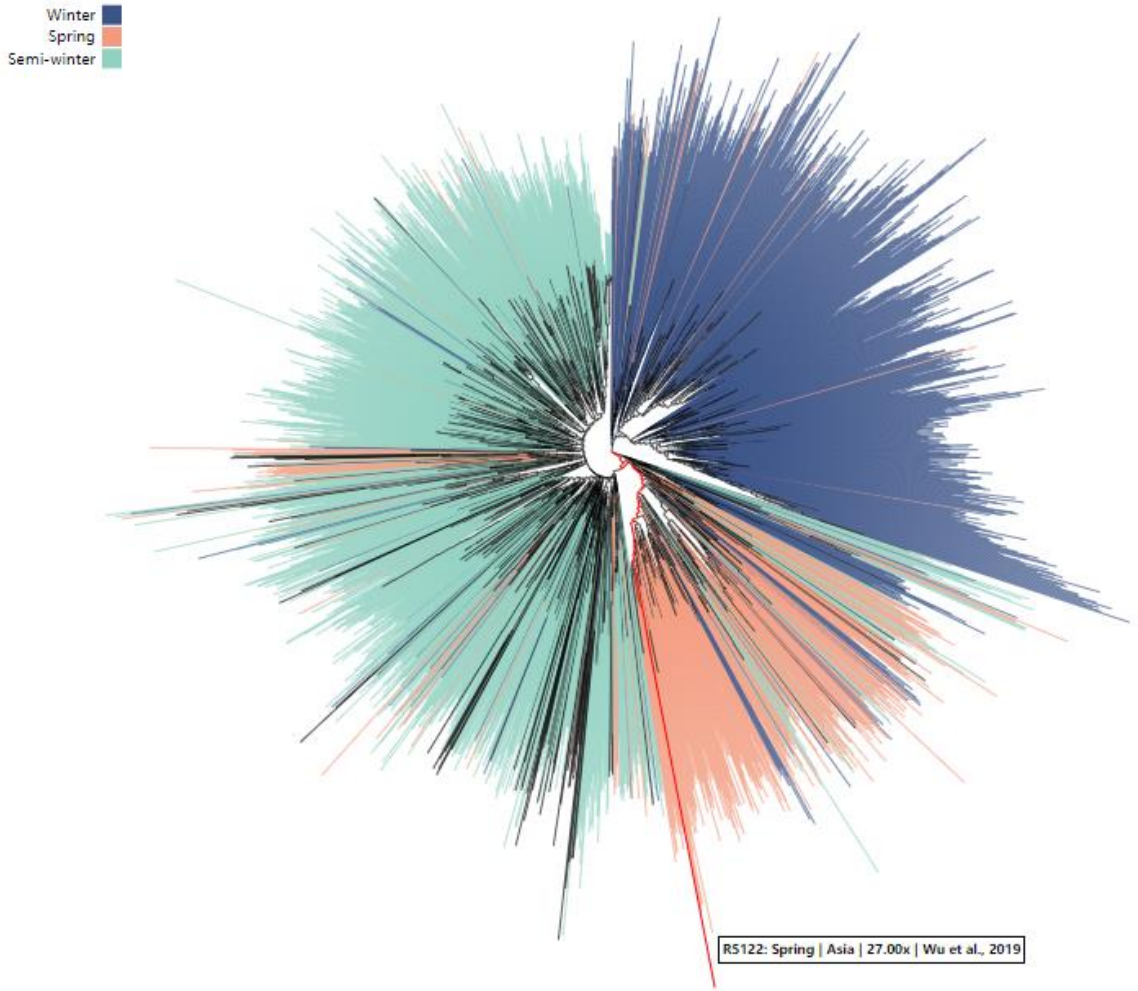
Part1: PCA analysis for B. napus population. More interactive features can be used via toolbar in the upper right hand corner.



Part 2: Population structure of the *B. napus* population. Ecotypes of spring, winter and semi-winter were marked in red, blue and green color. When mouse over the bar, information of the accessions are shown.



Part 3: Phylogenetic tree of *B. napus* accessions. Ecotypes of spring, winter and semi-winter were marked in red, blue and green color. When mouse over the branches, information of the accessions are shown.
Part 4: Parameters that can be set to choose the subset of the samples, according origin, ecotype, source and depth of the sequencing data.

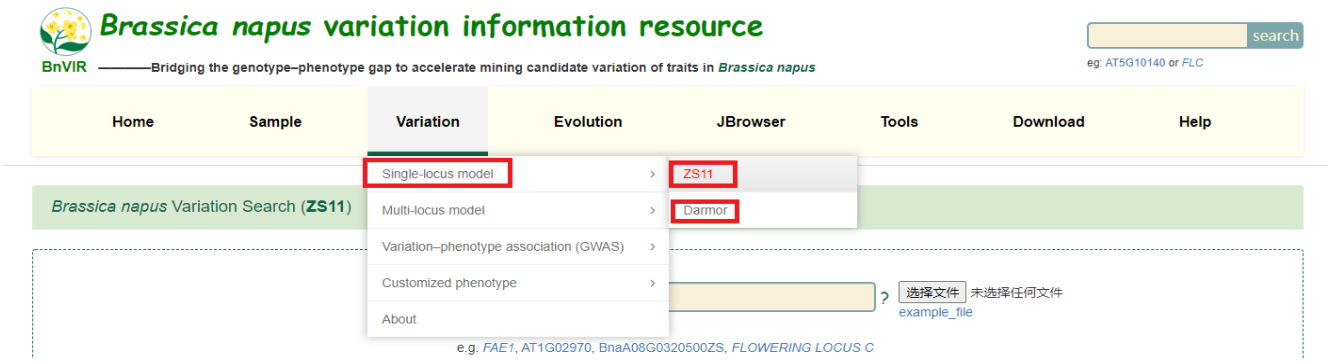


2.3 About

More information about Sample module.

3 Variation

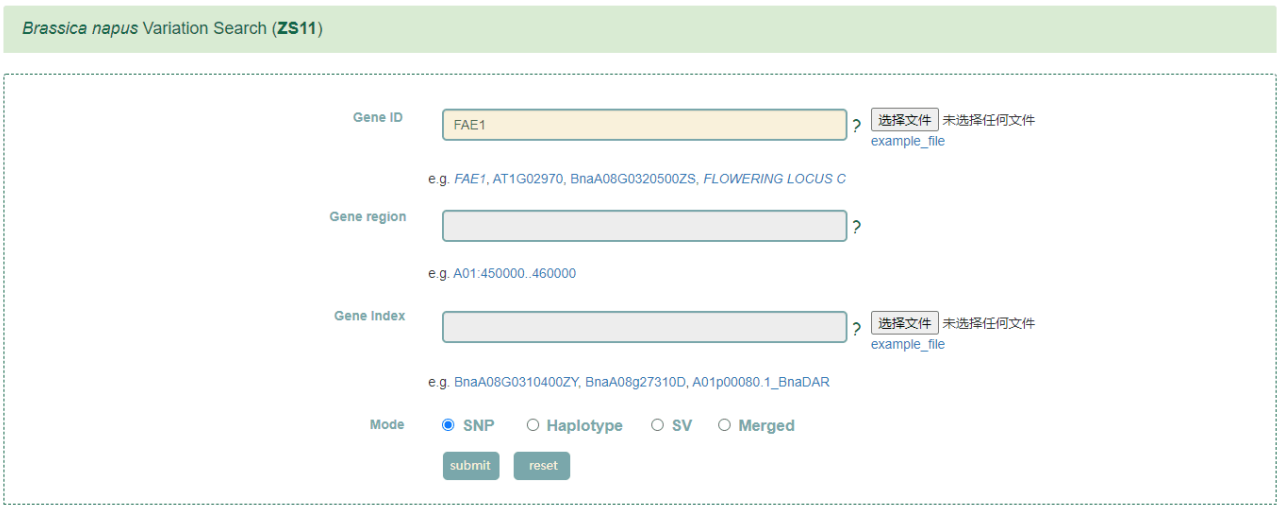
Variation module integrates genetic variations, phenotype and transcription data of the *B. napus* population, providing variation query and association between genotype and phenotype in mining candidate variations/genes od traits. There are four functional interfaces in Variation module, including Single-locus model, Multiple-locus model, GWAS and Customized phenotype. Information about variations, including annotation, allele frequency, distribution, description of gene expression and phenotypic values, are provide in Single-locus model, enabling user to screen candidate variations/genes of traits. Multi-locus model enables users to combine the multiple loci cross genes to check gene-gene interactions and their contribution to phenotype. GWAS interface provides browse of GWAS result for the *B. napus* population. In Customized phenotype, users can upload their own phenotype data to inspect the candidate variations/genes of traits. Users can choose ZS11 and Darmor-*bzh* genome as reference in each function page (except for Multiple-locus model) in secondary Menu.



3.1 Single-locus model

3.1.1 Input

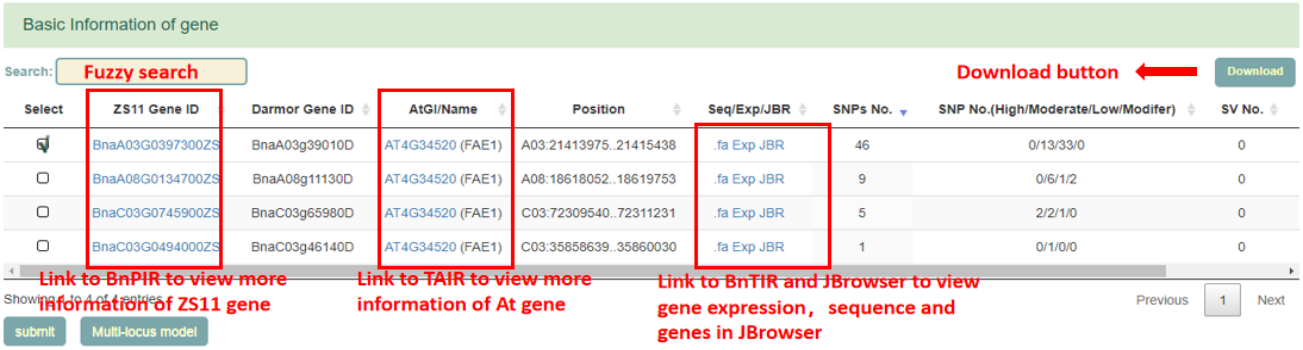
We will demostrate the uses of Variation module based of ZS11 reference.
Gene ID: gene ID of ZS11 reference, gene ID and gene name of Arabidopsis.
Genomic region: genomic region of ZS11 genome in a format of "chr:start..end".
Gene index: gene ID of any published B. napus genomes.
Users can choose SNP, Haplotype, SV or all the three types of variations.



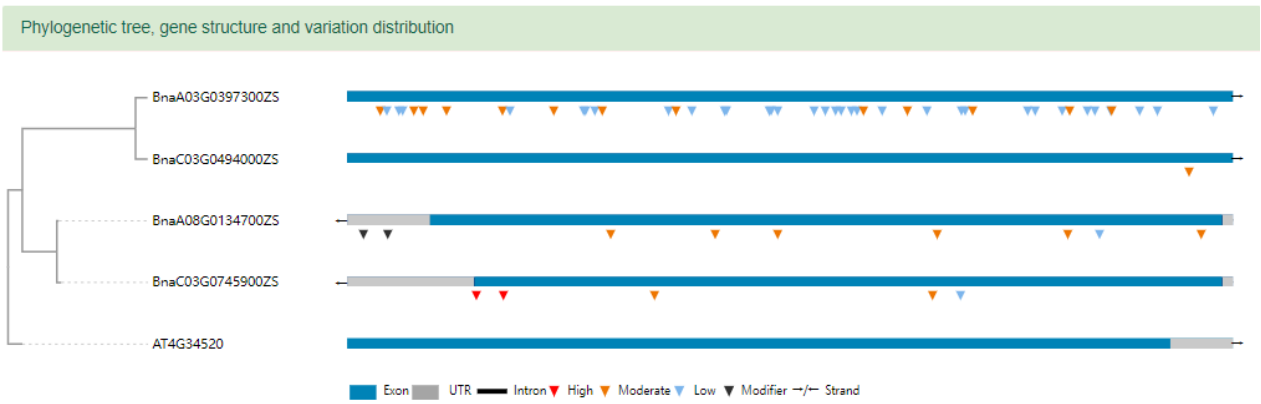
3.1.2 Result page of SNP mode

We demonstrate the search results with the *FAE1* gene. After entering "*FAE1*"and choose "SNP" in search bar, the result page will shown in minute. The result page compromise 11 parts. Figures and tables in the result page are interactive and downloadable by click the button locate in the upper right hand corner.

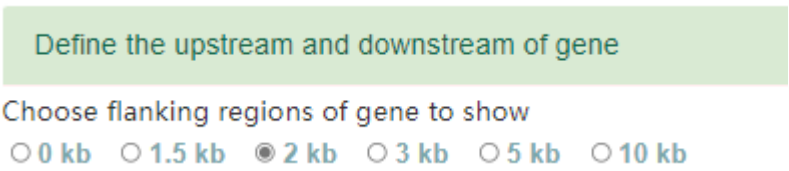
Part 1: Table of target genes and their basic information. Orthologs of *FAE1* in ZS11 are shown in the table. Uses can click the blue column to get more information about the genes. User can choose one gene for further analysis. BnaA08G0134700ZS is selected here.



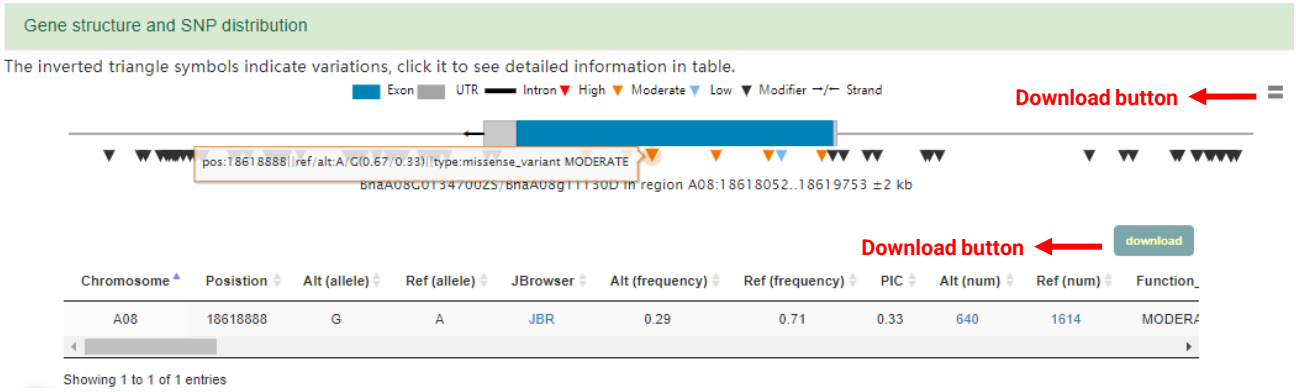
Part 2: Phylogenetic tree of orthologs genes of *B. napus* and Arabidopsis. For each gene in tree, the gene structure is shown with SNPs or short InDels at the bottom indicated by colored triangle. Mouse over to see brief information of the variation.



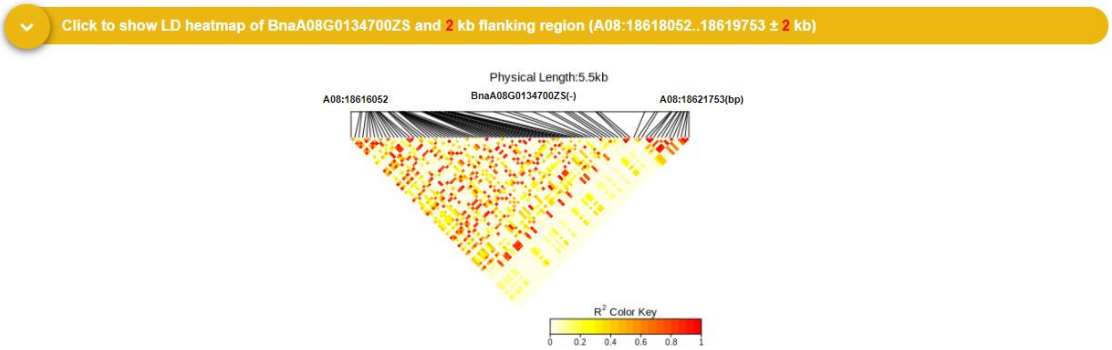
Part 3: Define the upstream and downstream of the target gene. "2 kb" is selected here.



Part 4: Gene structure and distribution of the variations. Variations are represented by triangles with different colors. Mouse over to see brief information of the variation. Click the triangle to see annotation of variation in the table at bottom.

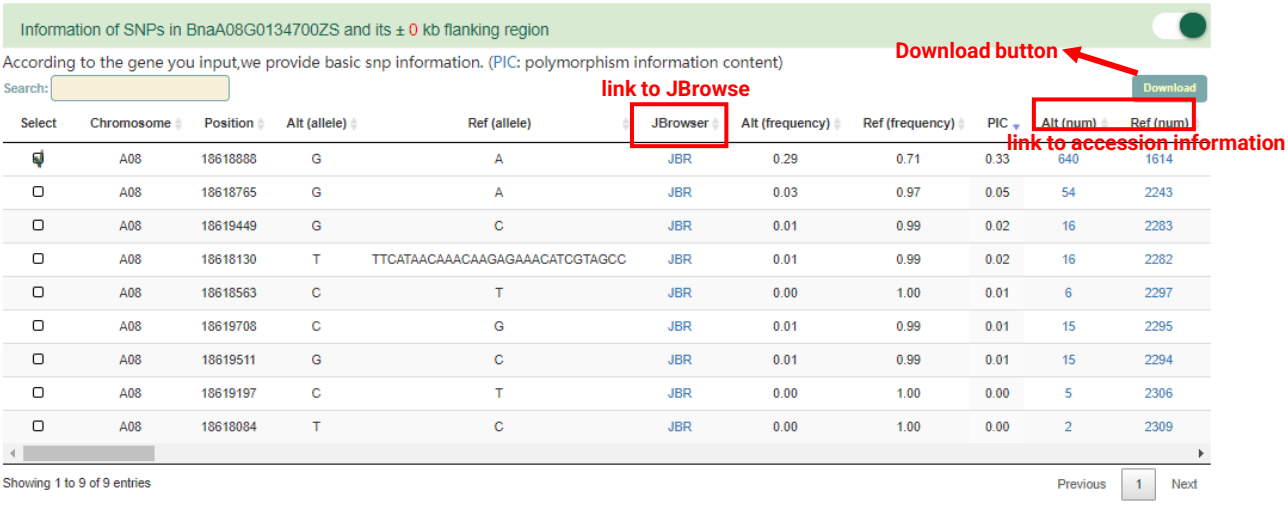


Part 5: LD heatmap of the gene and its up- and downstream.

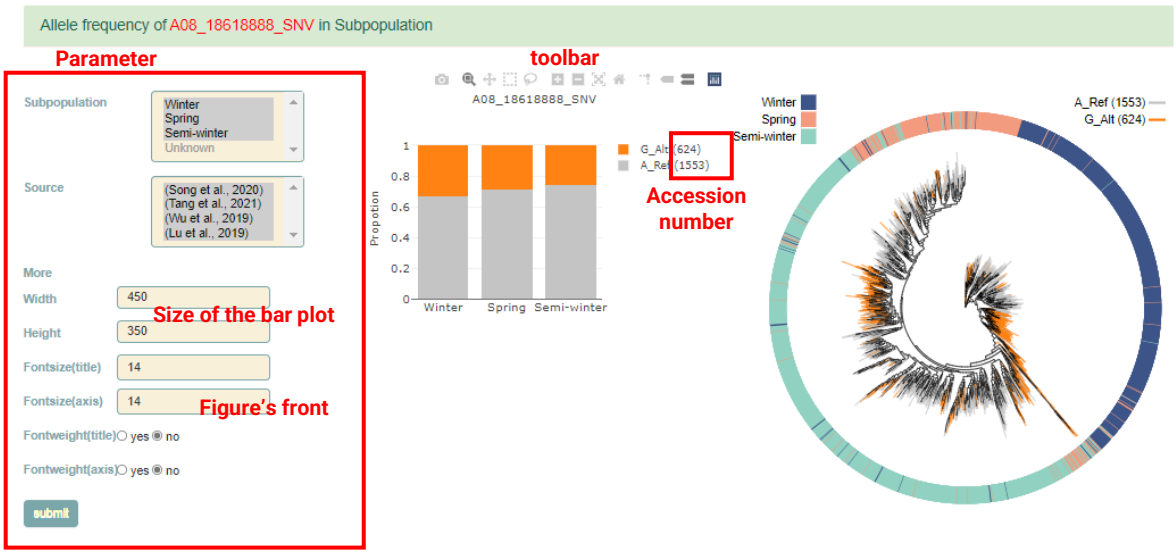


Part 6: Statistics of the annotation result of variations.

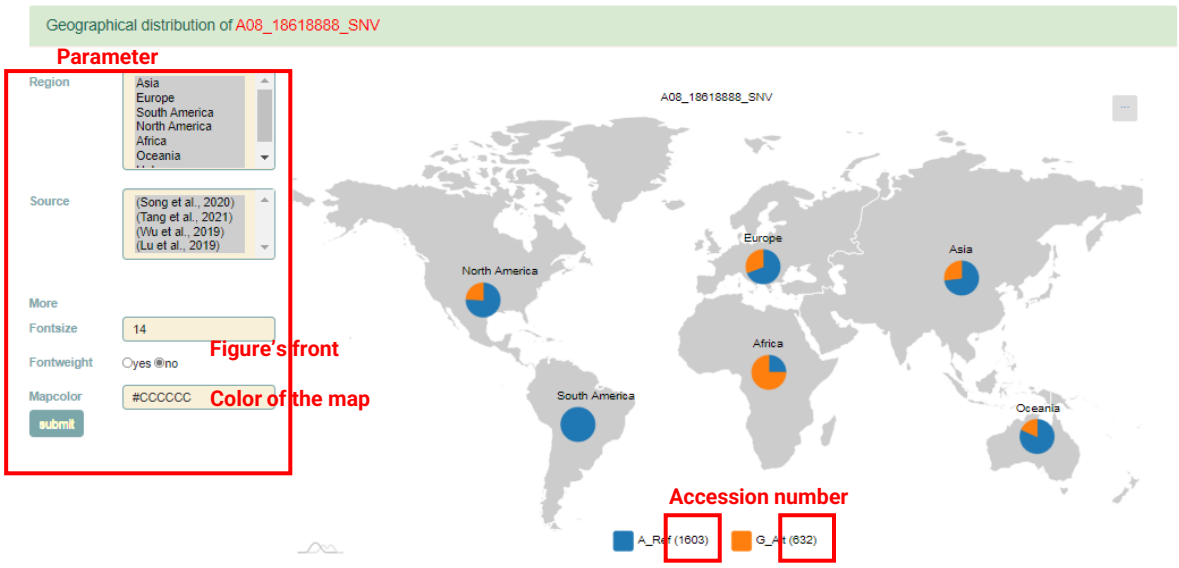
Part 7: Table showing the detail information of the variations. Users can search variations by entering any characters relevant to variation. Click the “JBR” column will bring users to JBRowser module for visualization of the variation. Click number of “Alt” and “Ref” column will bring users another view presenting a table of sample information.



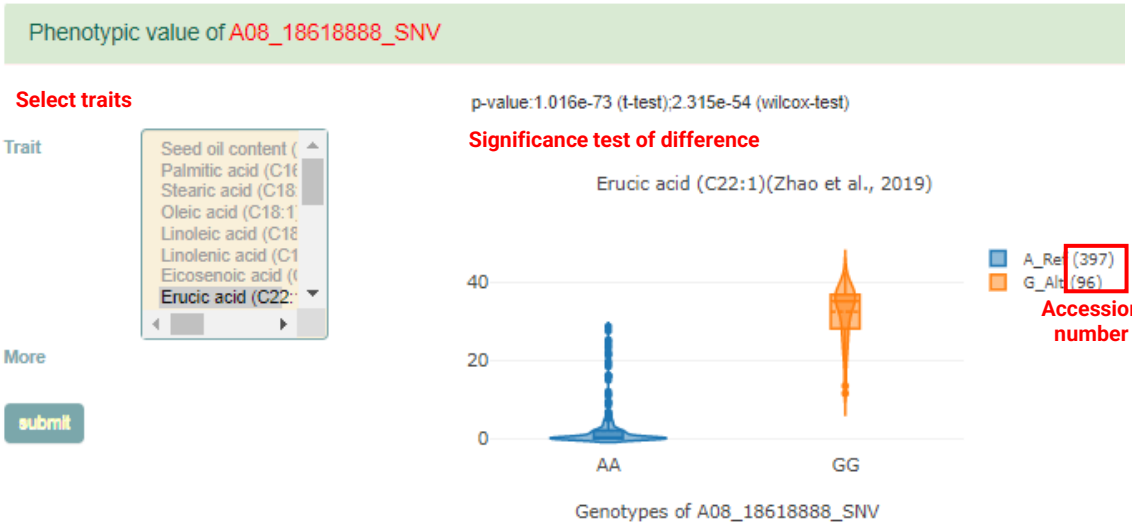
Part 8: Allele frequency of the selected variation. Bar plot and phylogenetic tree shows the allele frequency for each subpopulation. Mouse over the bar plot to get statistics result, and more interactive features are shown in toolbar at the top. Mouse over the branch in tree to see accessions information. Users can choose subpopulation and change style of the figure by setting the parameter in the left.



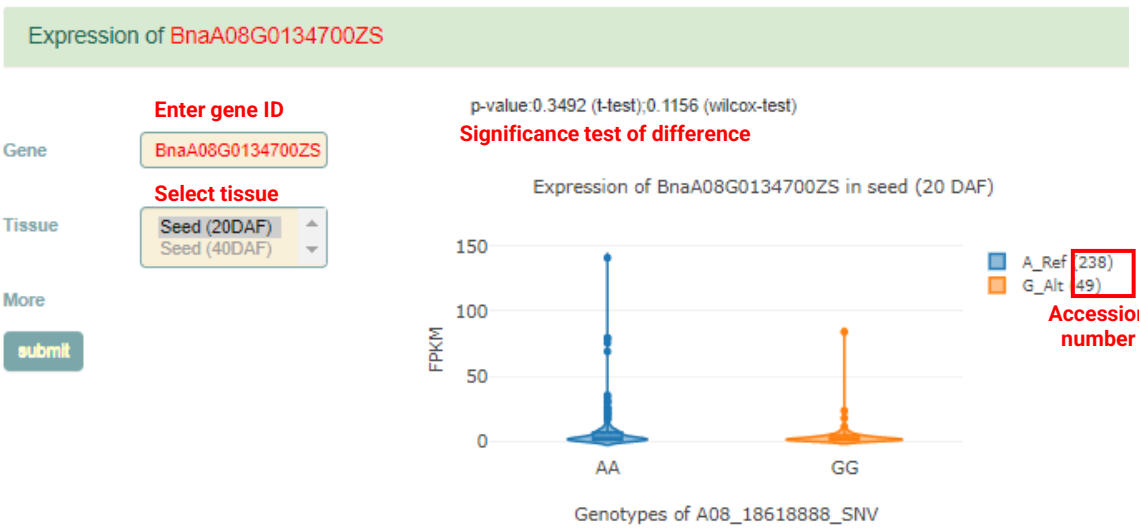
Part 9: Geographic distribution of the accessions with different alleles. Users can choose subpopulation and change style of the figure by setting the parameter in the left.



Part 10: Phenotypic effect of the selected variation. Violin plot shows the distribution of phenotype value of the *B. napus* accessions with different alleles. Number of accessions of "Alt" and "Ref" are shown in legend. Differenced determined by Student's t-test and Wilcoxon test are shown at the top. 18 traits are provided to be choose in left hand.



Part 11: Gene expression of *B. napus* accessions with different alleles. Violin plot shows the distribution of gene expression level of the B. napus accessions with different alleles. Number of accessions of "Alt" and "Ref" are shown in legend. Differenced determined by Student's t-test and Wilcoxon test are shown at the top. Users can enter gene ID and choose tissues in left hand.

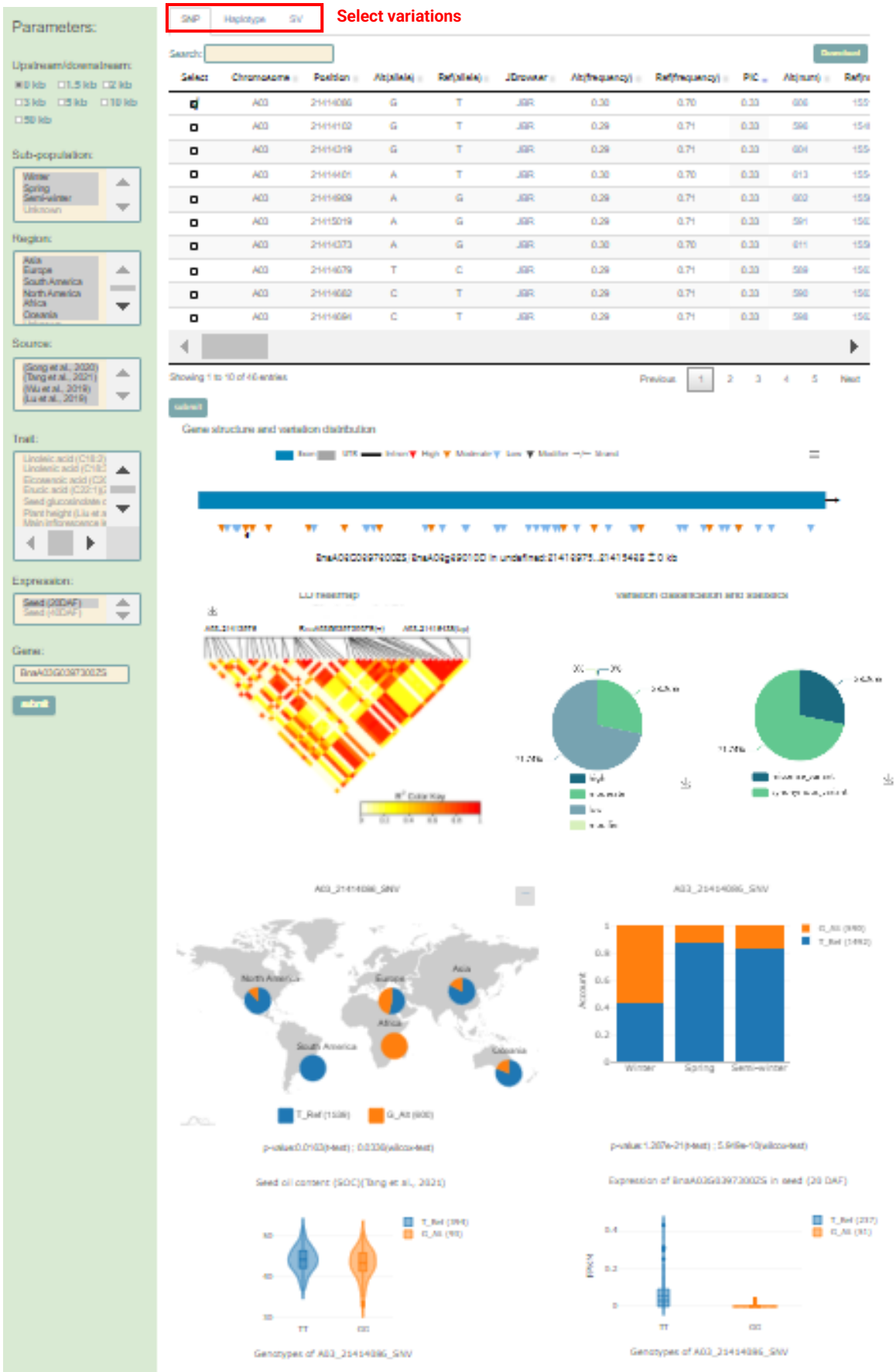


3.1.3 result page of the Haplotype and SV mode

The result of Haplotype and SV mode are similar to that of SNP mode. Identification of haplotypes and SVs was description in Methods.

3.1.3 Result page of the merge mode

All the figures and table are arranged compactly. All the parameters are in the left hand. Users can quickly switch the result page of SNP, haplotype and SV by click them at the top.

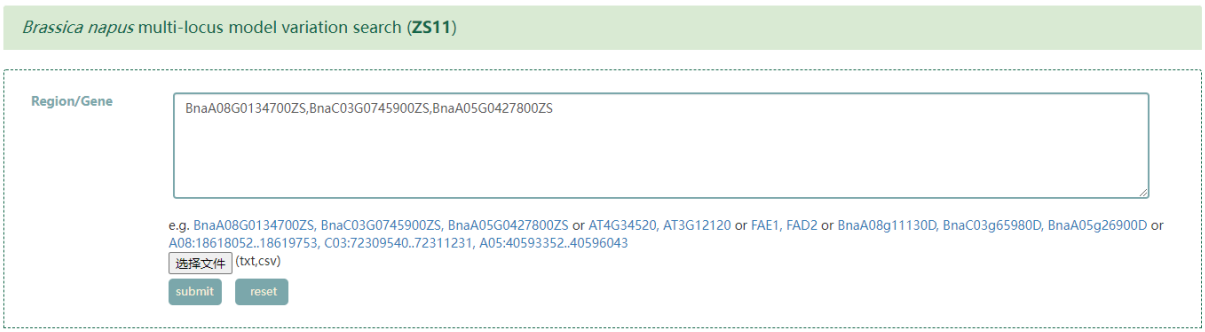


3.2 Multi-locus model

This page enables users to combine the multiple loci cross genes to check gene-gene interactions and their phenotypic effect.

3.2.1 Input

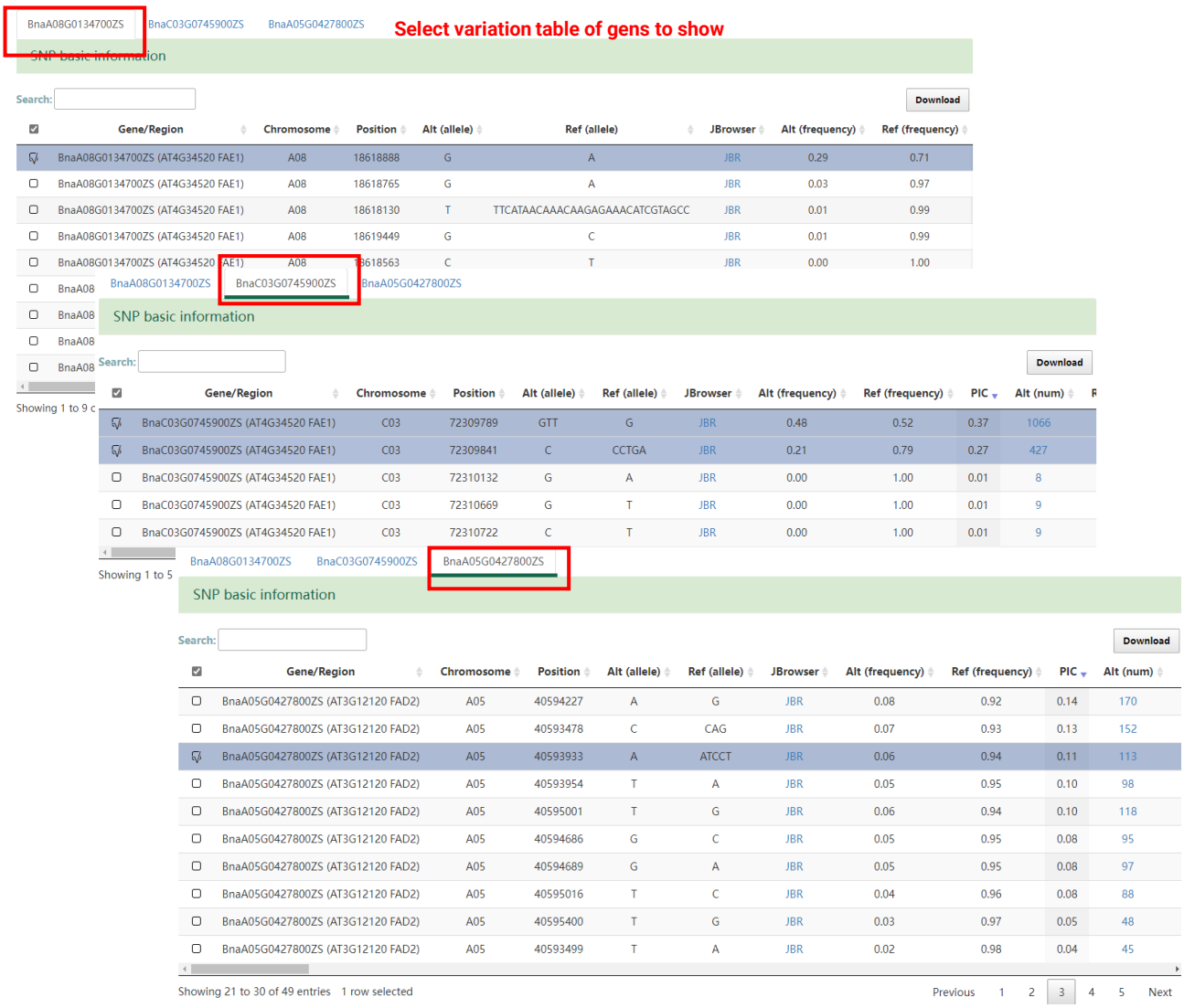
Users can enter gene list (separated by comma), or genomic regions. Gene number should be not more than 10.



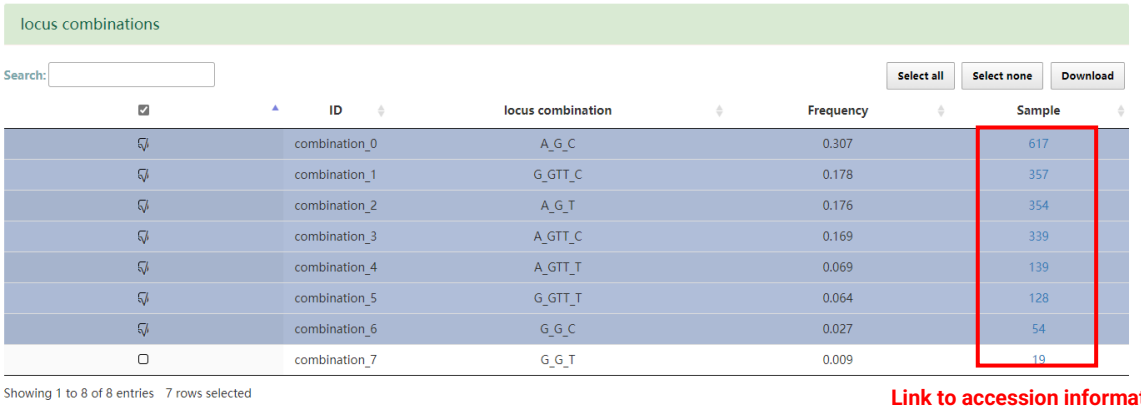
3.2.1 Result page

The result page including four parts. We show the result of the input of "BnaA08G0134700ZS,BnaC03G0745900ZS,BnaA05G0427800ZS" here.

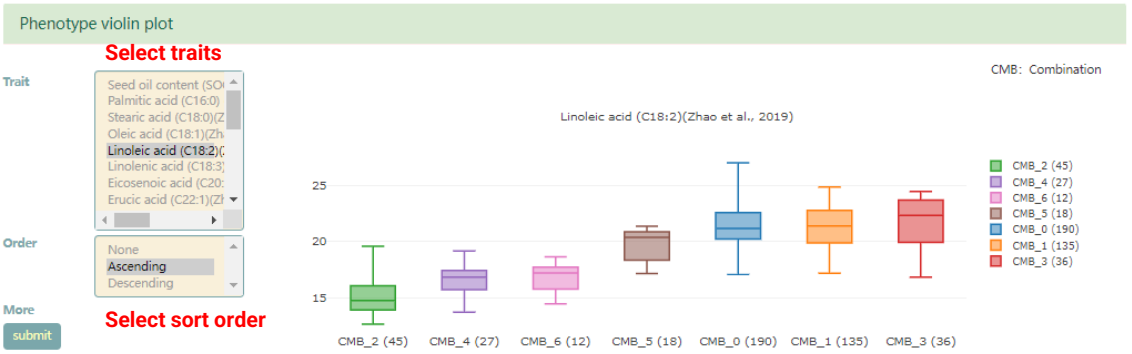
Part 1: Variation information of each gene. Information of the variation table is described above. Click gene ID at the top of the result page and select interested variations for each genes and then submit. Four variations are selected here.



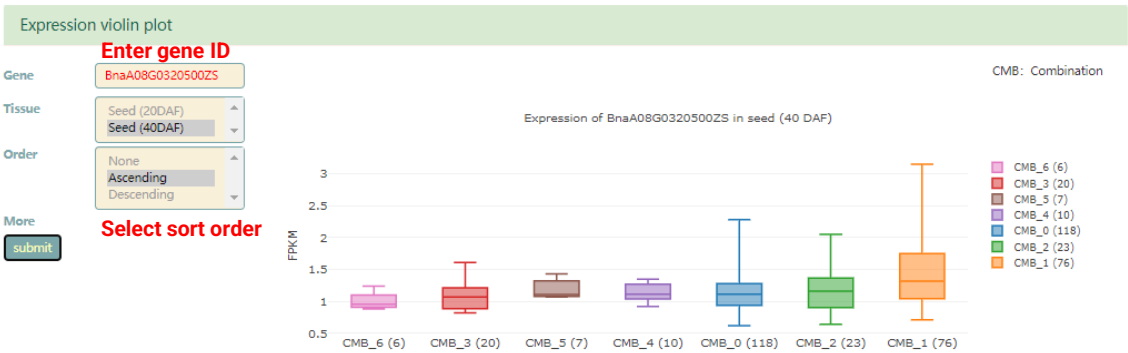
Part 2: Detail information of the different four-alleles combinations, including allelic combination and their accession number, frequency. Click the number of the accession will bring users to the table with accession information.



Part 3: Phenotypic effect of the locus combinations. Violin plot shows the distribution of phenotype value of the *B. napus* accessions with different combinations. Number of accessions of each combination is shown in legend. 18 traits are provided to choose in left hand.



Part 4: Gene expression of *B. napus* accessions with different combinations. Violin plot shows the distribution of gene expression level of the *B. napus* accessions with different combinations. Number of each combination is shown in legend. Users can enter gene ID and choose tissues in left hand.

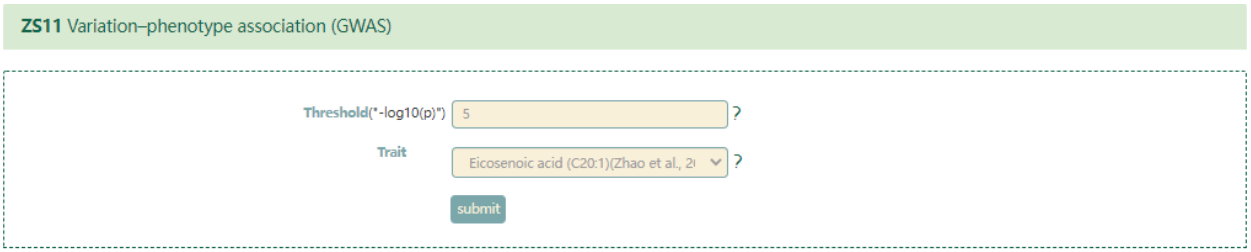


3.3 GWAS

GWAS interface provides browse of GWAS result for 18 traits in the *B. napus* population.

3.3.1 Setting

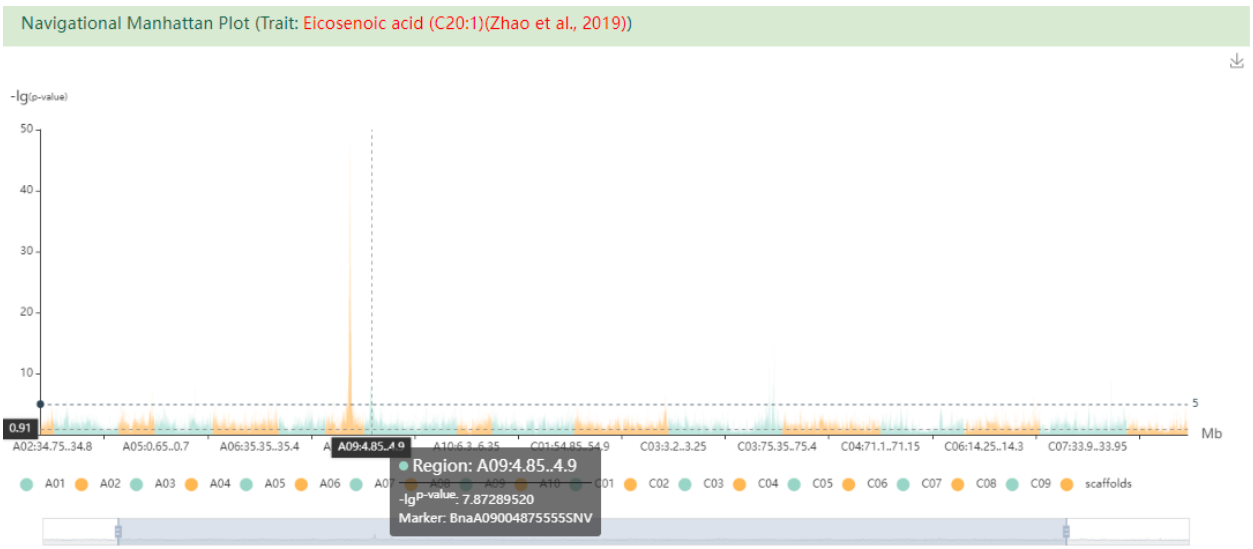
Firstly, users should set the threshold of the significant associations, and select the trait.



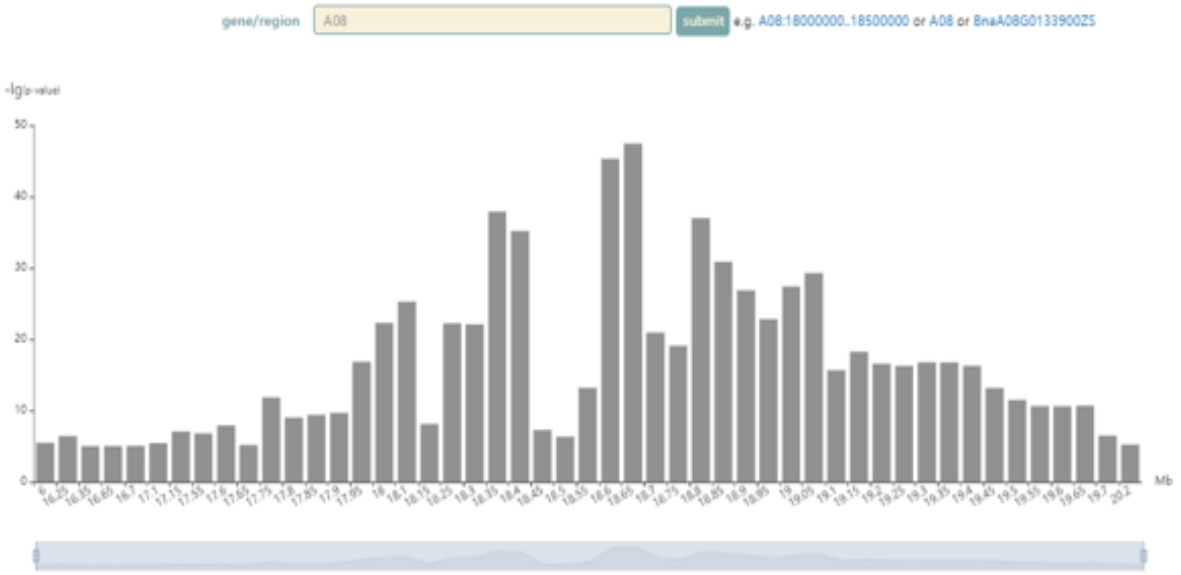
3.3.2 Result page

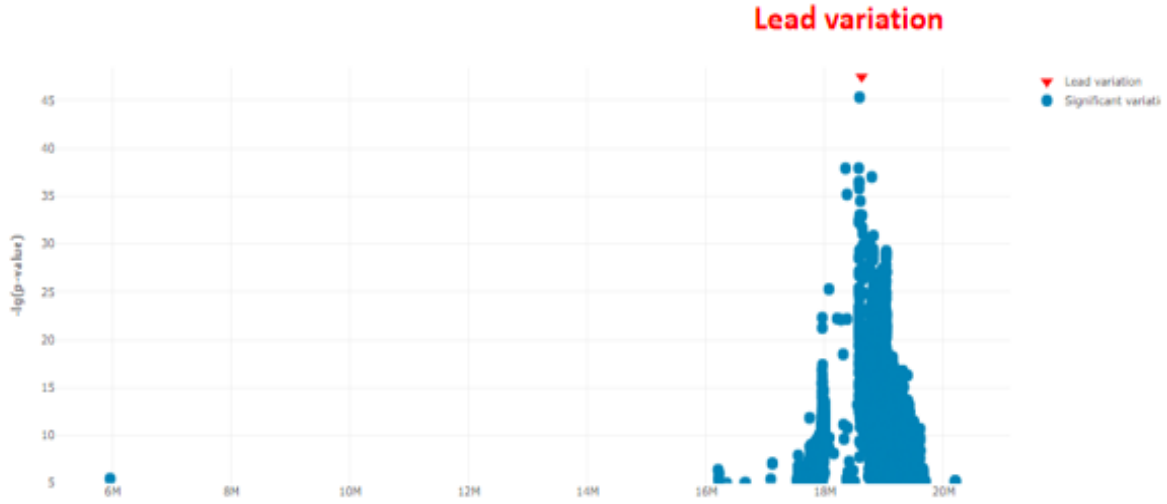
The result page provides global Manhattan plot and query of significance of the local region or gene.

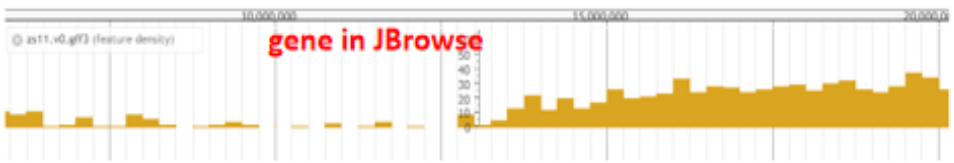
Part 1: Manhattan plot of the genome. Users can zoom in and out the plot by scrolling the mouse. Users can also mouse over the bar to get the information of the region, including genome location, -lg(p value), variation ID.



Part 2: Local Manhattan plot. Users can check specific genomic region by entering genomic region or gene ID in the search bar at the top. Top: Bar plot of the 50-kb. Middle: scatter plot the significant variations. Lead variations are indicated by red triangle. Bottom: JBrowse of the genome and genes. Users can mouse over these interactive plots to get the corresponding information.





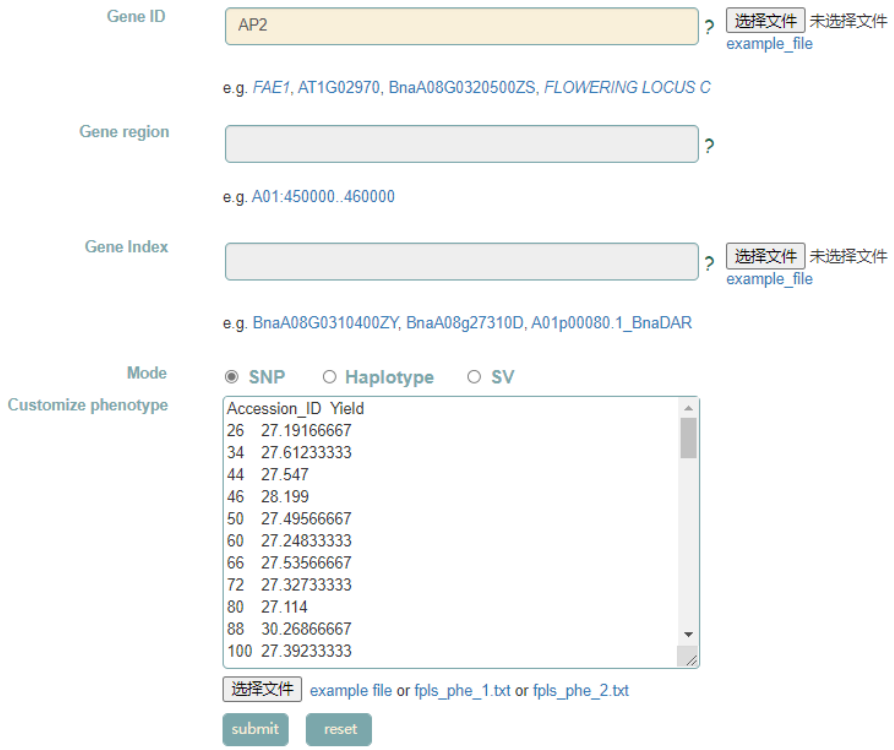


3.4 Customized phenotype

User can upload their own phenotype data to screen candidate variations/genes of the traits in the page.

3.4.1 Input

User can upload or copy the phenotype file in which rows represent the accession ID and columns represent the traits. Other input is same as the in Single-locus model. Note that although accession names overlap with those in the database, the seed stocks might be sourced from different origins with potentially different degrees of purity, or maintained in different collections with the possibility of outcrossing, which lead to an unreliable result.



3.4.2 Result page

The result page is similar to that of Single-locus model.

3.5 About

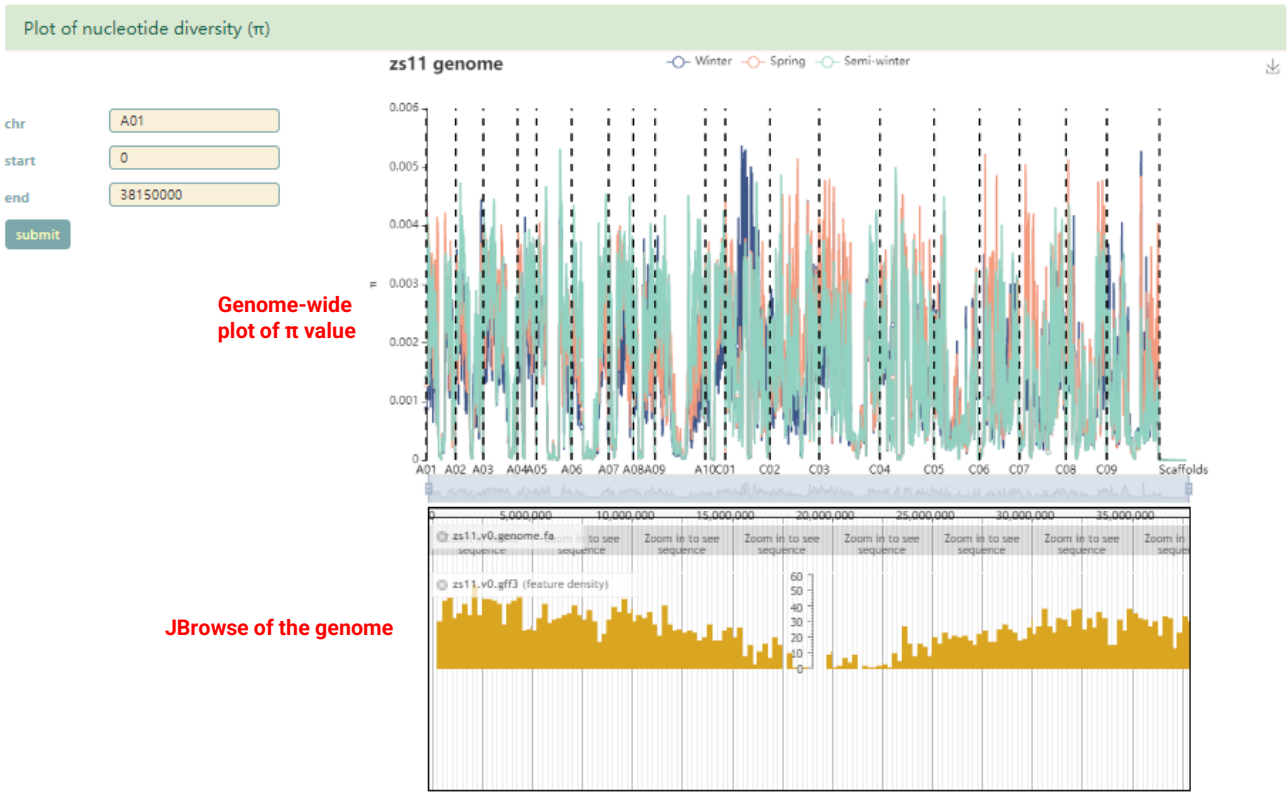
More information about Variation module.

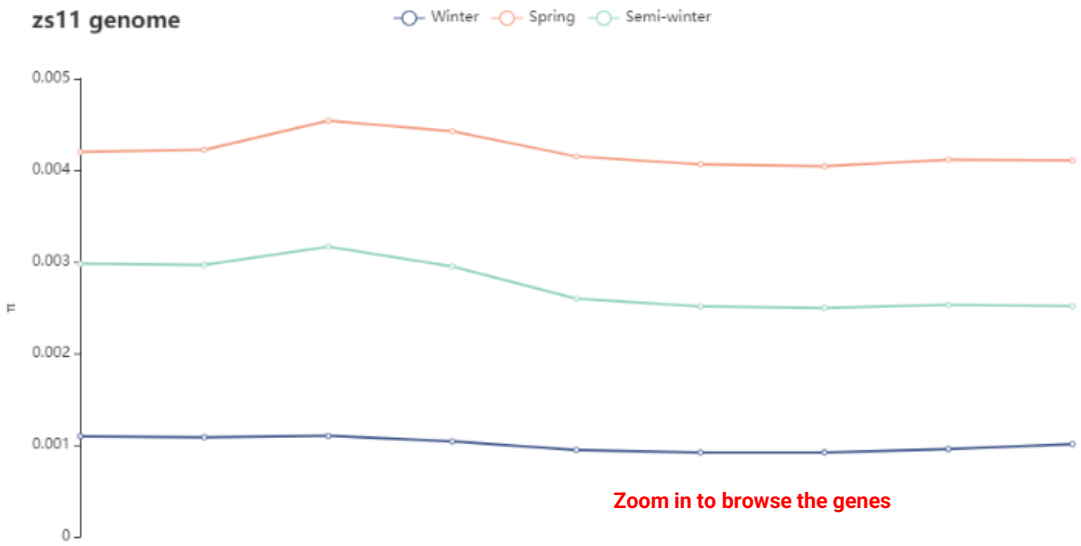
4 Evolution

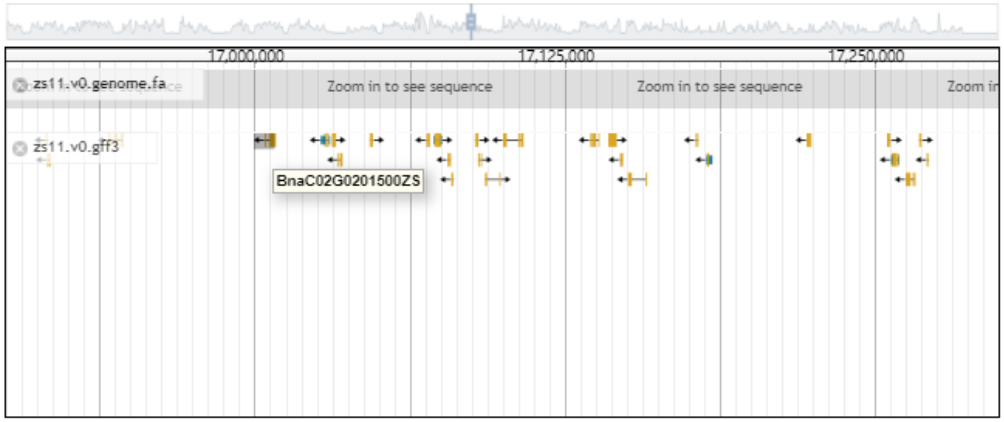
The Evolution module provides selection signals, including nucleotide diversity (π), fixation index (F_{ST}) and K_a/K_s ratio, for the detection of genomic regions or genes under selection.

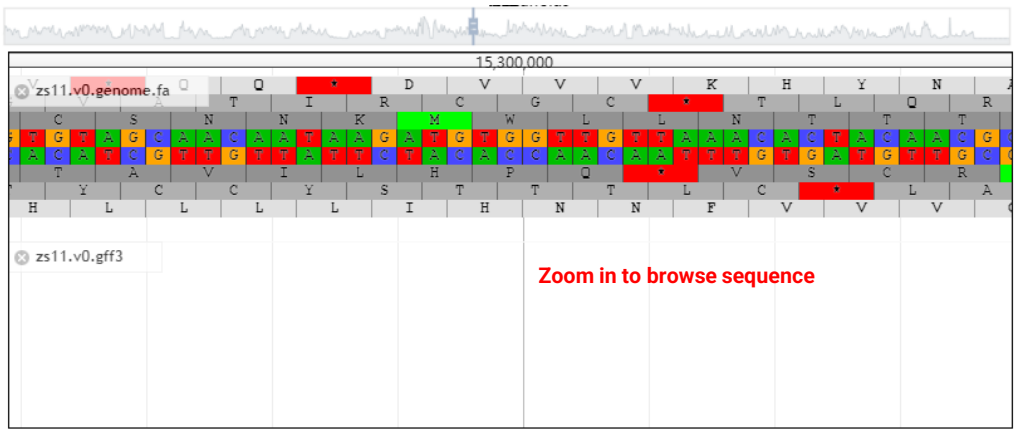
4.1 π

Users can zoom in and out the genome-wide plot to screen the π value of the specific region by scrolling the mouse. In line plot of π value at the top, three ecotypes are shown, respectively. Gene density, genes module and genome sequence will be shown in JBrowse at the bottom when gradually zoom out the line plot. Mouse over the line plot to view the π value and genome position.

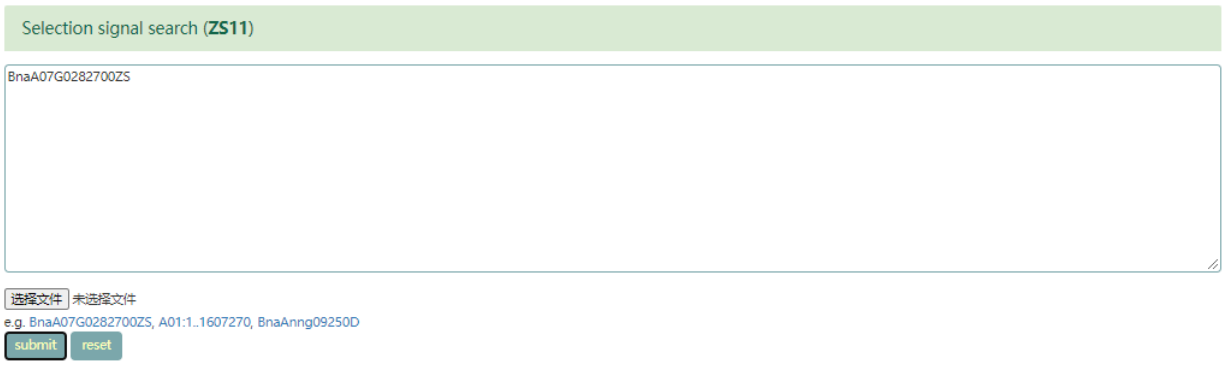


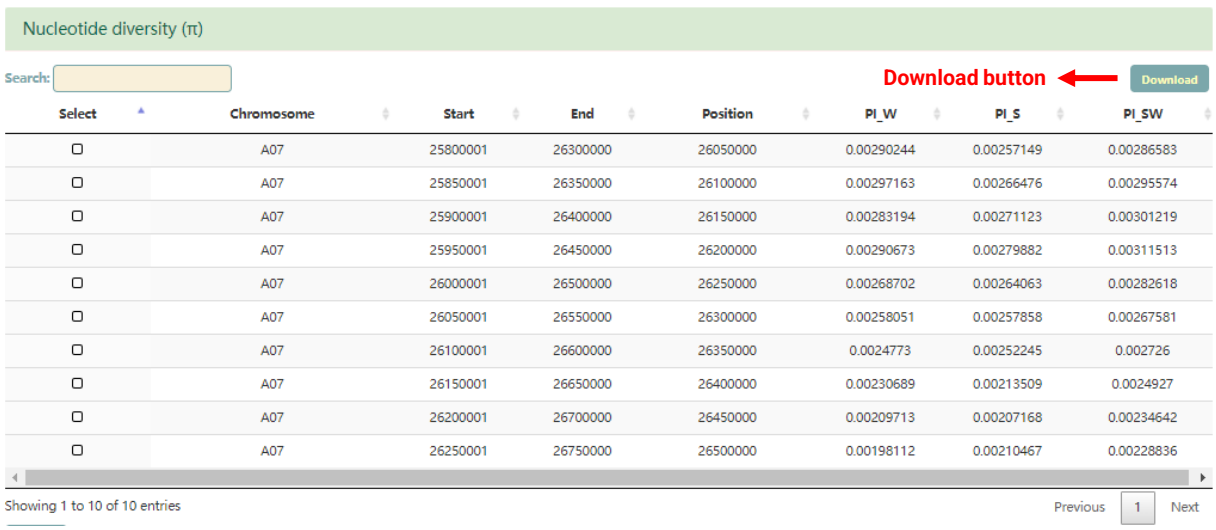

zs11 genome
Winter
Spring
Semi-winter
0.005
0.004
0.003
0.002
0.001
0
π
Zoom in to browse the genes

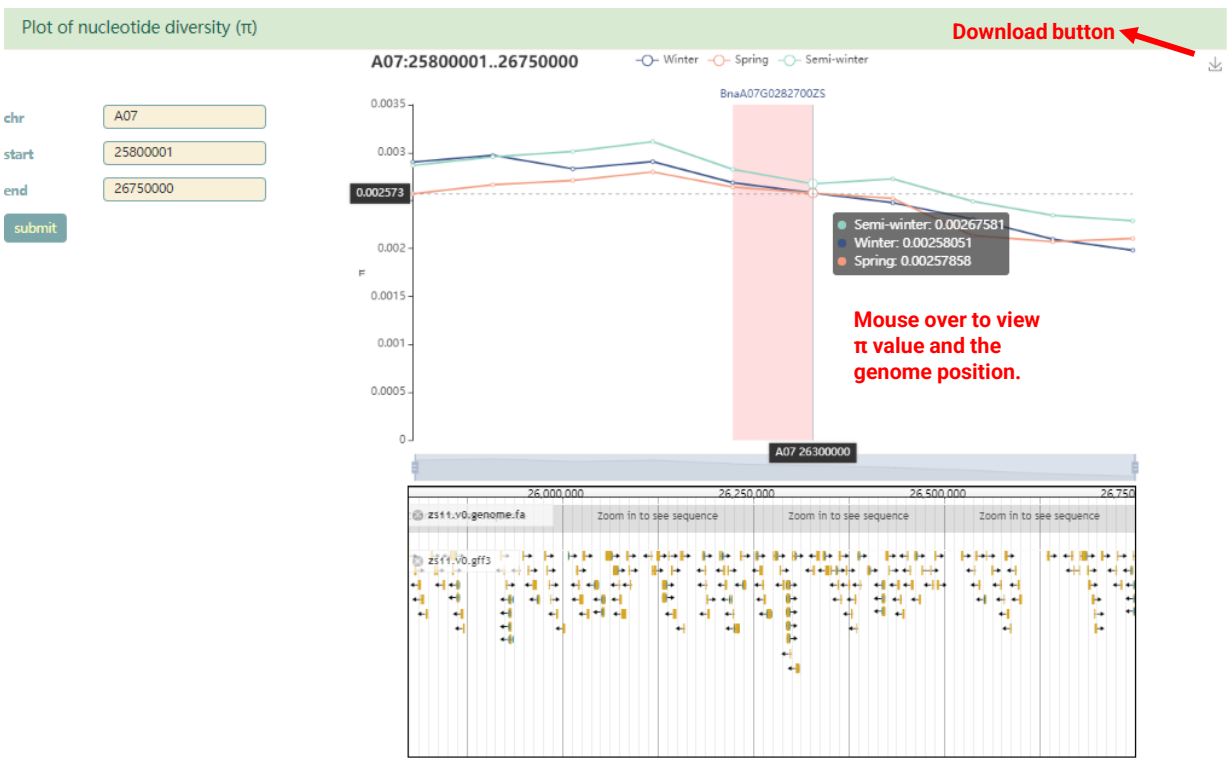

17,000,000
17,125,000
17,250,000
zs11.v0.genome.fa
Zoom in to see sequence
Zoom in to see sequence
zs11.v0.gff3
BnaC02G0201500ZS


15,300,000
zs11.v0.genome.fa
zs11.v0.gff3
Zoom in to browse sequence

Users can also screen the π value of the specific region by entering genomic region or gene ID of *B. napus*. "BnaA07G0282700ZS" is entered here. In the result page, table with π value, line plot and JBrowse of target region are available for download or browse.







4.2-4.3 F_{ST} and K_a/K_s

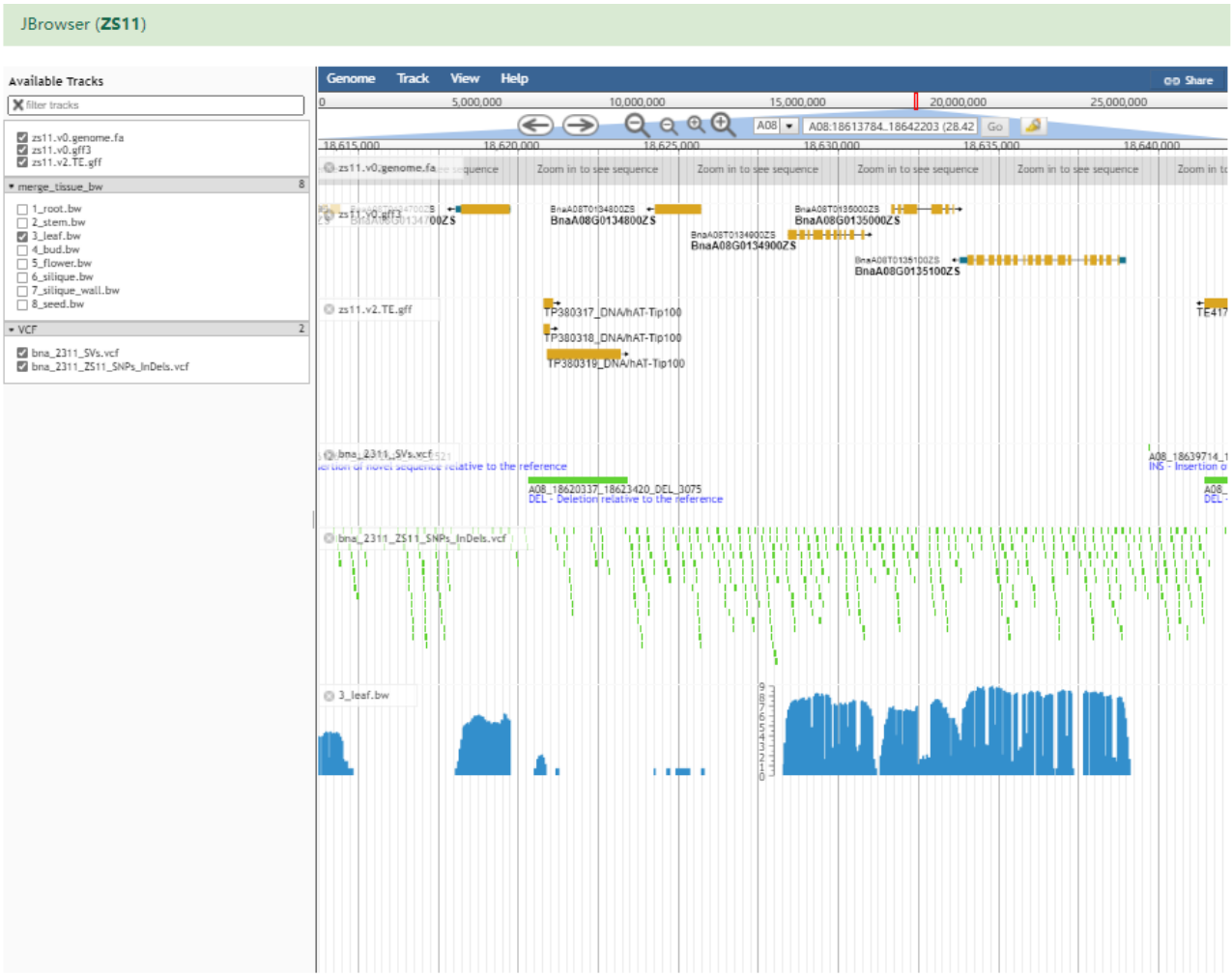
Input and the result pages of F_{ST} and K_a/K_s is similar to that of π.

4.4 About

More information about Evolution module.

5 JBrowser

This module enable users visualize genetic variations, gene features and transcription data in ZS11 and Darmor-*bzh* genome, respectively. There are 13 tracks can be selected to display, including nucleotide sequence, genes, transposable elements (TEs), expression profile data of different tissues and variation (SNPs, InDels and SVs). Description of the files in JBrowse are shown in the following table.



Description of the files in JBrowser are shown in the following table. And the generation of the file are description in About and Methods of website.

Reference	File	Description
ZS11	zs11.v0.genome.fa	DNA sequence of ZS11 genome (Song et al., 2020)
	zs11.v0.gff3	Gene annotation of ZS11 genome (Song et al., 2020)
	zs11.v2.TE.gff	TE annotation of ZS11 genome (Song et al., 2020)
	1_root.bw	mRNA read coverage of root by mapping to ZS11
	2_stem.bw	mRNA read coverage of stem by mapping to ZS11
	3_leaf.bw	mRNA read coverage of leaf by mapping to ZS11
	4_bud.bw	mRNA read coverage of bud by mapping to ZS11
	5_flower.bw	mRNA read coverage of flower by mapping to ZS11
	6_silique.bw	mRNA read coverage of silique by mapping to ZS11
	7_silique_wall.bw	mRNA read coverage of silique wall by mapping to ZS11
	8_seed.bw	mRNA read coverage of seed by mapping to ZS11
	bna_2311_SVs.vcf	Information of SV identified in population of 2,311 *B. napus* accessions based on ZS11 genome
	bna_2311_SNPs_InDels.vcf	Information of SNPs and InDels identified in population of 2,311 *B. napus* accessions ZS11 genome
Darmor-*bzh* v4	Darmor-bzh.v4.genome.fa	DNA sequence of Darmor-*bzh* genome (Chalhoub et al., 2014)
	Darmor.v5.gff3	Gene annotation of Darmor-*bzh* genome (Chalhoub et al., 2014)
	Darmor.TE.gff	TE annotation of Darmor-*bzh* genome (Chalhoub et al., 2014)
	1_root.bw	mRNA read coverage of root by mapping to Darmor-bzh
	2_stem.bw	mRNA read coverage of stem by mapping to Darmor-bzh
	3_leaf.bw	mRNA read coverage of leaf by mapping to Darmor-bzh
	4_bud.bw	mRNA read coverage of bud by mapping to Darmor-bzh
	5_flower.bw	mRNA read coverage of flower by mapping to Darmor-bzh
	6_silique.bw	mRNA read coverage of silique by mapping to Darmor-bzh
	7_silique_wall.bw	mRNA read coverage of silique wall by mapping to Darmor-bzh
	8_seed.bw	mRNA read coverage of seed by mapping to Darmor-bzh
	bna_2311_SVs.vcf	Information of SV identified in population of 2,311 B. napus accessions Darmor-bzh genome
	bna_2311_SNPs_InDels.vcf	Information of SNPs and InDels identified in population of 2,311 B. napus accessions Darmor-bzh genome

6 Tools

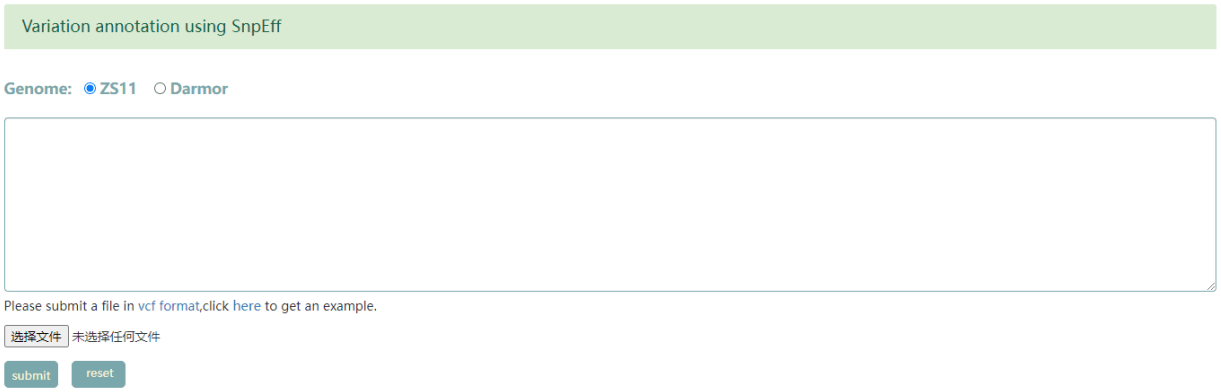
Some useful bioinformatic tools are provided in this module.

6.1 Variation annotation

This page provides handy way for user to annotation the variation of their own with SnpEff based on ZS11 and Darmor-*bzh* genome.

6.1.1 Input

Users can upload vcf file of their own. Users can click “vcf Format” to know about the detail of vcf file. A example file is provide for test.



6.1.2 Output

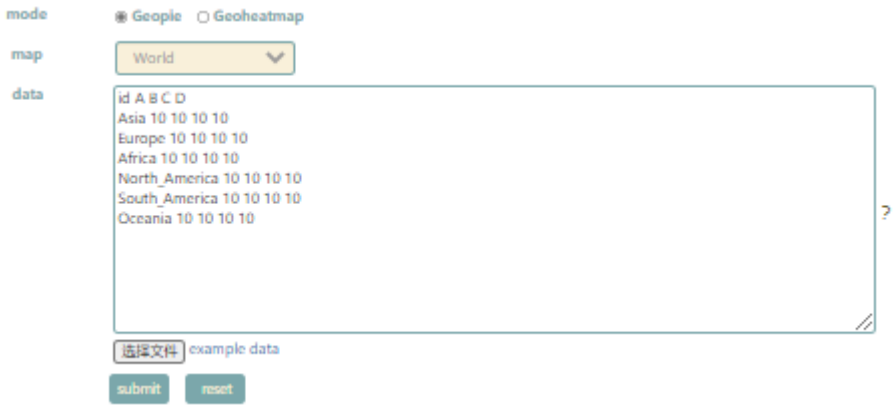
The output of vcf file and html file are available for download.

6.2 Data2geomap

This function allows users to draw a geomap for data of their own.

6.2.1 Input

Users can upload sample data of their own. Example data are provided for test. Map of the world, China and US and two types of geomap, geopie and geoheatmap, can be chosen.



6.2.2 Output

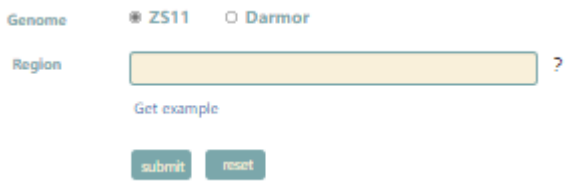
Users can download the figure by click the button in the upper right corner.

6.3 LD heatmap

Users can obtain a LD heatmap estimated from variation panel of based on ZS11 and Darmor-*bzh* genome in database for a specific genomic region.

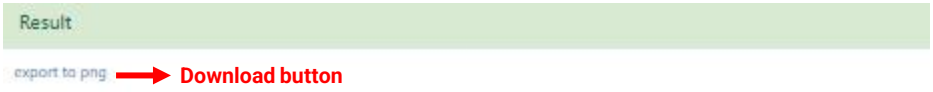
6.3.1 Input

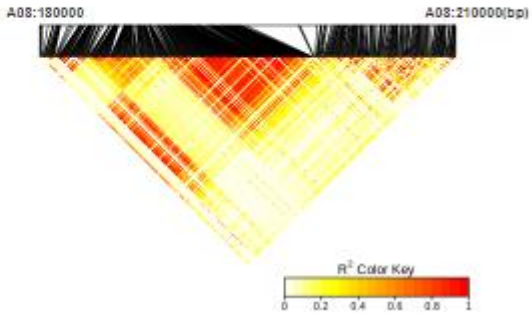
Reference genome should be chosen, and genomic region should be entered.



6.3.2 Output

Users can download the table by clicking the button in the upper right corner in the result page.





7 Download

Accession information, variation file of ZS11 and Darmor-*bzh*, phenotype data and database tutorial are available for download.

+ Accession information

- Accession information.xlsx

+ Variation information(Genome: ZS11)

- bna2311_ZS11_SNP.A01.vcf.gz
- bna2311_ZS11_SNP.A02.vcf.gz
- bna2311_ZS11_SNP.A03.vcf.gz
- bna2311_ZS11_SNP.A04.vcf.gz
- bna2311_ZS11_SNP.A05.vcf.gz
- bna2311_ZS11_SNP.A06.vcf.gz
- bna2311_ZS11_SNP.A07.vcf.gz
- bna2311_ZS11_SNP.A08.vcf.gz
- bna2311_ZS11_SNP.A09.vcf.gz
- bna2311_ZS11_SNP.A10.vcf.gz
- bna2311_ZS11_SNP.C01.vcf.gz
- bna2311_ZS11_SNP.C02.vcf.gz
- bna2311_ZS11_SNP.C03.vcf.gz
- bna2311_ZS11_SNP.C04.vcf.gz
- bna2311_ZS11_SNP.C05.vcf.gz
- bna2311_ZS11_SNP.C06.vcf.gz
- bna2311_ZS11_SNP.C07.vcf.gz
- bna2311_ZS11_SNP.C08.vcf.gz
- bna2311_ZS11_SNP.C09.vcf.gz
- bna2311_ZS11_SNP.scaffold.vcf.gz

+ Variation information(Genome: Darmor)

+ Phenotype information

- Flowering_time_data_from_Wu_2019.xlsx
- Flowering_time_data_from_Song_2020.xlsx

+ Transcription information

- Gene.expression.From.Tang_2021

+ Tutorial

- Tutorial.pdf

7 Help

Database tutorial, methods of data analysis, pipeline of variation identification and contact are available in Help.