D1432–D1441 *Nucleic Acids Research, 2022, Vol. 50, Database issue* *Published online 10 November 2021*
https://doi.org/10.1093/nar/gkab1057

BRAD V3.0: an upgraded Brassicaceae database

Haixu Chen[1], Tianpeng Wang[1], Xiaoning He[1], Xu Cai[1], Runmao Lin[1], Jianli Liang[1], Jian Wu[1], Graham King[2] and Xiaowu Wang[1,*]

[1]Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, No.12, Haidian District, Beijing 100081, China and [2]Southern Cross Plant Science, Southern Cross University, Lismore, New South Wales, Australia

Received September 13, 2021; Revised October 14, 2021; Editorial Decision October 15, 2021; Accepted October 19, 2021

ABSTRACT

The Brassicaceae Database (BRAD version 3.0, BRAD V3.0; http://brassicadb.cn) has evolved from the former *Brassica* Database (BRAD V2.0), and represents an important community portal hosting genome information for multiple *Brassica* and related Brassicaceae plant species. Since the last update in 2015, the complex genomes of numerous Brassicaceae species have been decoded, accompanied by many omics datasets. To provide an up-to-date service, we report here a major upgrade of the portal. The Model-View-ViewModel (MVVM) framework of BRAD has been re-engineered to enable easy and sustainable maintenance of the database. The collection of genomes has been increased to 26 species, along with optimization of the user interface. Features of the previous version have been retained, with additional new tools for exploring syntenic genes, gene expression and variation data. In the 'Syntenic Gene @ Subgenome' module, we added features to view the sequence alignment and phylogenetic relationships of syntenic genes. New modules include 'MicroSynteny' for viewing synteny of selected fragment pairs, and 'Polymorph' for retrieval of variation data. The updated BRAD provides a substantial expansion of genomic data and a comprehensive improvement of the service available to the Brassicaceae research community.

INTRODUCTION

BRAD is a genome database and analysis portal that has been available to the *Brassica* research community for over a decade since its first release in 2010 (1). It represents a major resource for understanding complex genome structures and their association with evolutionary and domestication history. The Brassicaceae is a large family of flowering plants with 4636 known species in 340 genera (2). It includes many crops of agricultural, ornamental, condimental, medicinal and scientific significance (3), as well as the model plant species *Arabidopsis thaliana* for which extensive gene functional information has been accumulated over the past 30 years. BRAD was the first comprehensive database hosting whole genome information for the first sequenced *Brassica* species, *B. rapa*, a mesohexaploid diploid resulting in a chromosomally segmented triplicated genome (4). The original key feature of BRAD was to associate the functional information for *A. thaliana* genes annotated in TAIR (5) with those identified in a *Brassica* crop through their syntenic relationships at the subgenome level. The subsequent sequencing of additional Brassicaceae genomes enabled detailed syntenic analysis to be extended to include 13 Brassicaceae species within BRAD v2.0 in 2015 (6).

Due to the rapid advances of sequencing technology in recent years, high-quality reference genome sequences of many Brassicaceae species have been either decoded or upgraded (7–35). This includes revisions to reference *B. rapa* (V3.0), *B. oleracea* (V2.0) and *B. nigra* (V2.0) as well as *de novo* assemblies for *Brassica carinata*, *Isatis indigotica* and *Raphanus sativus* and several other species. Several independent online genome databases and portals have been developed to host *Brassica* genome information. The www.brassica.info site is an information and reference portal for the Multinational Brassica Genome Project. In addition, the following provide specific browsers or analysis tools: the *Brassica napus* pan-genome information resource, BnPIR (36); SNP Database of *Brassica napus*, BnaSNPDB (37); The *Brassica oleracea* Genome Database, Bolbase (38); A Gene Expression Database for *Brassica* Crops, BrassicaEDB (39); The Genomic Variation Database of *B. napus*, BnaGVD (40). These databases have different capabilities, although for each their scope is limited either to a specific species, or a specific data type. BRAD remains the only genome database and analysis platform that can provide information linked across a number of Brassicaceae species with a wealth of gene and genomic features.

[*]To whom correspondence should be addressed. Tel: +86 10 82105971; Fax: +86 61 274123; Email: wangxiaowu@caas.cn
Present address: Xiaoning He, Institute of Plant Sciences Paris of Saclay (IPS2), UMR 9213/UMR1403, CNRS, INRA, Université Paris-Sud, Université d'Evry, Université Paris-Diderot, Sorbonne Paris-Cité, Bâtiment 630, 91405 Orsay, France.

The previous version of BRAD was built to provide a service for a fixed number of genomes. In consideration of the rapid increase of genomic data volume and updating speed, we rebuilt BRAD using the Model–View–ViewModel (MVVM) framework. The site has moved from the 'http://brassicadb.org' domain hosted on an institutional server to the 'http://brassicadb.cn' domain hosted in a cloud server.

MATERIALS AND METHODS

All data sources, data processing, web interface, and features of BRAD V3.0 are summarized in Figure 1. It consists of five parts: data collection, gene functional annotation, online analysis and visualization, database construction, web interface and toolkit development. The programs and packages used in the process are summarized in Supplementary Table S1.

Data sources

Based on data from previous version, when identifying genomic data to be included within BRAD, we wished to represent evolutionarily or economically important species of Brassicaceae that met the following criteria: being hosted in the version 2.0, of either having an upgraded assembly version, or that the assembly for a new Brassicaceae species was of high quality. Resequencing data for *B. rapa*, *B. oleracea* and *B. napus* were from published papers (41–43). All RNA-seq data were downloaded from NCBI SRA database (www.ncbi.nlm.nih.gov/sra) to represent as many organs or tissues as possible for each species, the metadata for which are summarised in Supplementary Table S2.

Syntenic Gene @ Subgenome

BLAST (44) was used by SynOrths (45) for the identification of syntenic genes. We reperformed syntenic gene analysis across all genomes in order to ensure consistency for synteny identification. In total, pairwise syntenic genes were identified from all protein-coding genes based on *A. thaliana* and *B. rapa* V1.5 using the default parameters of SynOrths. For species that have not undergone a *Brassica* whole-genome triplication (*Br-α* WGT) event, synteny was detected in relation to *A. thaliana*. For species with a *Br-α* WGT event (Figure 1A), three sets of subgenomes were allocated based on the *B. rapa* genome (Chiifu) V1.5, and then the genes in each subgenome were compared to the *A. thaliana* gene set. For *Camelina sativa*, we used the available subgenome information from supplementary data (20). For *Leavenworthia alabamica*, we treated it as a species that has not undergone WGT because its genome is of poor quality (Supplementary Figure S2). However, because *A. thaliana* is not the direct *Brassiceae* ancestor having a tPCK karyotype (46), species with a *Br-α* WGT event will lose syntenic gene pairs that otherwise are unique in the tPCK karyotype species. Therefore, the latter species were further compared with *B. rapa* genome V1.5, so that genes lacking synteny with *A. thaliana* may form syntenic gene pairs with *B. rapa* genome V1.5. Based on the synteny table for *A. thaliana* and *B. rapa* (46), we integrated these syntenic gene pairs into a combined synteny table (Supplementary Figure S1A).

Multiple sequence alignment and gene phylogeny

The available gene, CDS and protein sequences integrated in the BRAD-MySQL database are available for online analysis. After receiving user query parameters from the browser, the BRAD server queries the corresponding sequences in the database and writes it to a .fasta file. This file is then used to compute multiple sequence alignment and phylogenetic tree results using the default parameters of MUSCLE (47,48). Sequence and alignment results are packaged in JSON format and sent to the browser, with multiple sequence alignment rendered using the Vue plugin vue-svg-msa and phylogenetic tree results using vue-phylogram.

Polymorph

We added genomic variations for 1251 re-sequenced genotype accessions, including two sets for *B. rapa*, where one represented 524 accessions and used *B. rapa* (Chiifu) V3.0 as the reference genome (41), and the other was a subset of 199 from 524 accessions used (Chiifu) V1.5 as reference (42). Two *B. oleracea* data sets representing 119 diverse accessions used *B. oleracea* (JZS) V1.1 (42) and (JZS) V2.0 as the reference genomes. The *B. napus* data set represented 608 diverse accessions (43) and used the (Darmor-bzh) V5 genome as the reference. We further filtered the variation data with MAF $\geq$0.05 and missing rate $\leq$0.1, and built a phylogenetic tree using the default parameters of PHYLIP (49). This generated a tree file that was rendered with iTOL (50). The corresponding data could be retrieved from the database, with SVG images rendered on the web page.

Gene function and functional elements annotation

We used the gene models as described in the original publications for each of the genomes, but re-annotated all of them (Figure 1B). The protein sequences were further compared against the InterPro database (51) using the default parameters of InterProScan (52) to identify InterPro domains, Pfam domains (53) and gene ontology (54). The KEGG orthology (55) was annotated using the KAAS web server (56). All genes except those from *A. thaliana* have been compared with *A. thaliana* using BLASTP, where the parameters of BLASTP alignment to be satisfied included: identity >70%, coverage of query gene >75%, coverage of subject gene >75%. The gene with the highest identity was defined as the BLASTX (best hit) match to *A. thaliana*.

RNA-seq data analysis

HISAT2 (57) was used to map reads to the respective reference genome for each species. Samtools (58) was used to generate the sorting bam files. FeatureCounts (59) was used to calculate the number of reads. The TPM (transcripts per million) values and standard deviations were calculated using custom Python script. For all software mentioned are default parameters were used.

Genomic sequence retrieval and MicroSynteny

In the 'Genomic sequence retrieval', the Biopython (60) was used within a custom Python script to read the genome

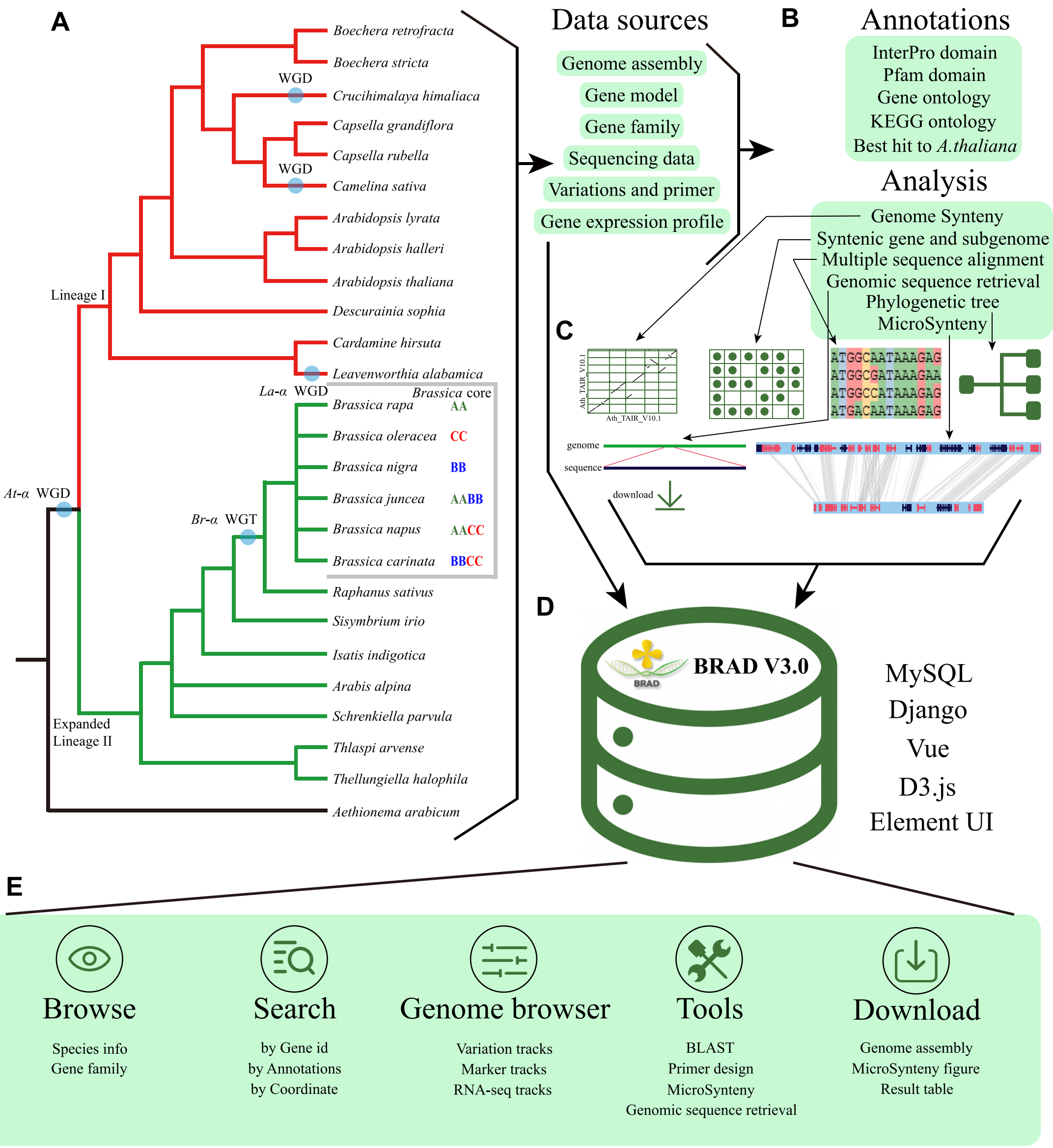


Figure 1. Overview of the Brassicaceae Database, BRAD. (**A**) Data sources included 25 species, 35 reference genomes in Brassicaceae and associated gene models, RNA-seq data. Lineage I and Expanded Lineage II were originally described by Mandakova et al (65–68). (**B**) Summary of data processing methods. (**C**) Display of web tools. (**D**) All data is stored in the MySQL relational database. The Django and Vue framework are used for interactive queries between the front and back ends. (**E**) An overview of the Web interface and the usage of the BRAD.

sequence and serve the genomic sequence to users. Based on 'Genomic sequence retrieval', we developed the 'MicroSynteny' module. Firstly, this obtains the sequences after receiving the query data transmitted by the web page. Secondly, the second sequence is used to build the BLAST library, and the first sequence acts as the query sequence. Finally, the collated results, including the gene and CDS information in the two fragments extracted from the database, are packaged in JSON format and sent to the browser and rendered by D3.js (https://d3js.org/). The methodology for this tool is described in detail in Supplementary Figure S1B.

Data integration

Syntenic genes, non-syntenic ortholog genes, transcriptome data, genomic synteny data and gene annotations were integrated in the BRAD-MySQL database. For variation data, we used the method of separating samples and data for storage and reading, and custom Python scripts for integration in the background. The relationships between data sets and main functions is shown in Supplementary Figure S1C.

Database construction and Web interface

All pre-processed data were integrated into the BRAD-MySQL database (Figure 1D). It contains a table of all gene model information, a table of all syntenic genes and a table of subgenome information, as well as a table including gene sequence, CDS sequence and protein sequence. The Django framework was used to query data from the database backend and build an API interface to send the data obtained from database to the browser. The Vue framework was used

Table 1. Genomic information included in the database

Abbreviation	Species	Genome assembly	Pub. date	Contig N50	Scaffold N50	Genome size (Mb)	Source
Araal_Paj_V4	*Arabis alpina*	version 4	02 Feb 2015	-	788 kb	375	(22)
Aetar_V1.0	*Aethionema arabicum*	version 1.0	30 Jun 2013	-	118 kb	191	(10)
Araha_Tad_V1.1	*Arabidopsis halleri*	version 1.1	27 Sep 2016	-	712 kb	124	(21)
Araly_MN47_V1.0	*Arabidopsis lyrata*	version 1.0	10 Apr 2011	-	-	200	(8)
Araly_MN47_V2.1	*Arabidopsis lyrata*	version 2.1	-	-	-	201	phytozome V12
Ath_TAIR_V10.1	*Arabidopsis thaliana*	version 10.1	14 Dec 2000	-	-	116	(7)
Braca_zd1_V1.0	*Brassica caritana*	Version 1.0	04 Feb 2021	1.44 Mb	60 Mb	1113	(11)
Braju_tum_V1.5	*Brassica juncea*	version 1.5	05 Sep 2016	61 kb	855 kb	784	(25)
Brana_ZS_V2.0	*Brassica napus*	version 2.0	28 Aug 2017	39.57 kb	602.22 kb	932	(28)
Brana_Dar_V5	*Brassica napus*	version 5	04 Sep 2014	-	-	850	(34)
Brana_Dar_V10	*Brassica napus*	version 10	15 Dec 2020	11.48 Mb	-	896	(33)
Brana_ZS_PB_V1.0	*Brassica napus*	version PB 1.0	19 Oct 2020	1.64 Mb	-	888	(27)
Brani_San_V1.1	*Brassica nigra*	version 1.1	30 Jun 2020	1.48 Mb	68.5 Mb	388	(29)
Brani_Ni100_V2	*Brassica nigra*	version 2.0	03 Feb 2020	17.1 Mb	-	491	(26)
Boere_V1.0	*Boechera retrofracta*	version 1.0	28 Mar 2018	-	2.3 Mb	217	(9)
Boest_V1.2	*Boechera stricta*	version 1.2	-	-	-	184	phytozome V12
Braol_JZS_V1.1	*Brassica oleracea*	version 1.1	23 May 2014	26 kb	1.45 Mb	374	(12)
Braol_JZS_V2.0	*Brassica oleracea*	version 2.0	09 Aug 2020	2.37 Mb	-	542	(30)
Braol_HDEM_V1.0	*Brassica oleracea*	version 1.0	02 Nov 2018	9.49 Mb	29.52 Mb	539	(16)
Brara_Chiifu_V1.5	*Brassica rapa*	version 1.5	28 Aug 2011	27 kb	1.97 Mb	279	(4)
Brara_Chiifu_V2.5	*Brassica rapa*	version 2.5	03 Apr 2017	-	3.38 Mb	375	(13)
Brara_Chiifu_V3.0	*Brassica rapa*	version 3.0	15 Aug 2018	1.45 Mb	4.44 Mb	341	(31)
Brara_Z1_V1	*Brassica rapa*	version 1.0	02 Nov 2018	5.52 Mb	15.39 Mb	390	(16)
Carhi_V1.0	*Cardamine hirsuta*	version 1.0	31 Oct 2016	-	509.5 kb	193	(15)
Capgr_V1.1	*Capsella grandiflora*	version 1.1	-	-	-	103	phytozome V12
Cruhi_V1.1	*Crucihimalaya himalaica*	version 1.1	20 Mar 2019	-	2.09 Mb	427	(23)
Capru_Mon_V1.1	*Capsella rubella*	version 1.1	09 Jun 2013	134.1 kb	15.1 Mb	135	(14)
Camsa_DH55_V1.0	*Camelina sativa*	version 1.0	23 Apr 2014	33.41 Kb	30.09 Mb	619	(20)
Desso_V1.1	*Descurainia sophioides*	version 1.1	-	-	-	124	phytozome V12
Isain_V1	*Isatis indigotica*	version 1.0	01 Feb 2020	1.18 Mb	36.17 Mb	294	(17)
Leaal_V1.0	*Leavenworthia alabamica*	version 1.0	30 June 2013	-	70 kb	167	(10)
Rapsa_Xiang_V1.0	*Raphanus sativus*	version 1.0	01 Nov 2015	39.62 kb	1.35 Mb	381	(18)
Sisir_V1.0	*Sisymbrium irio*	version 1.0	30 June 2013	-	135 kb	249	(10)
Schpa_V1.0	*Schrenkiella parvula*	version 1.0	07 Aug 2011	5.29 Mb	-	134	(24)
Thlar_MN106_V1.1	*Thlaspi arvense*	version 1.1	27 Jan 2015	140.82 kb	-	406	(19)
Theha_V1.0	*Thellungiella halophila*	version 1.0	21 Mar 2013	272 kb	8 Mb	235	(35)

to develop a user-friendly web interface, and to implement multiple custom dynamic charts using the Element UI library.

Toolkit development

Primer3 (61) was used to design primers for specific parameters. Synteny viewer was developed to visualize the synteny between different genomes. Web BLAST was driven by Sequenceserver (62). Genome sequences, gene models, markers, variation and omics data were displayed by JBrowse (63).

RESULTS

Overview of data

The BRAD covers a comprehensive representation of species types from the Brassicaceae family. V3.0 includes 36 genome assemblies of 26 species from 12 tribes (64) within the Brassicaceae (Figure 1A, Table 1, Supplementary Table S3). Compared with V2.0, 23 additional genome assemblies were included representing 13 species and 6 tribes. Thirteen genome assemblies were from Lineage I, 22 from Expanded Lineage II, and *Aethionema arabicum* represent an ancestral species (65–68). At the species level, 12 are from Lineage I and 13 from Expanded Lineage II. At the tribe level, five are from Lineage I and six from Expanded Lineage II. In addition, BRAD v3.0 also provides different versions of the genomes for the economically important *Brassica* crop species, which will assist in a wide range of studies.

The syntenic genes identified differed slightly from BRAD V2.0 (Supplementary Table S4). More than 1 million syntenic genes were identified from all protein-coding genes, among which 1 052 853 were syntenic with *A. thaliana* genes. Of all species except *A. thaliana*, 11 were derived from Lineage I, of which *Crucihimalaya himalaica*, *Camelina sativa* and *Leavenworthia alabamica* have been reported as having undergone WGT. Among these, we identified 27 225, 57 423 and 16 836 syntenic genes respectively. The low number of syntenic genes in *Leavenworthia alabamica* was largely due to the low quality of genome assembly. In the remaining eight species, we identified 17 591–21 132 syntenic genes. There are 13 species belong to Expanded Lineage II. For the four diploid *Brassiceae* species, we collected four assemblies of *B. rapa*, three of *B. oleracea*, two of *B. nigra* and one *R. sativus*. Among them, we identified 25 173–34 578 syntenic genes. For the three allotetraplloid *Brassiceae* species, we collected four assemblies of *B. napus*, one of *B. carinata* and one of *B. juncea*. Among them, we identified 49 442–66 872 syntenic genes. In the remaining six species that have not undergone WGT, we identified 12 968–19 016 syntenic genes. Besides, we identified 91 920 tandem arrays, 221 524 tandem duplication genes (Supplementary Table S5).

We also annotated a total of 1.83 million protein-coding genes of all genomes except for *A. thaliana*, with gene ontology terms (54) and KEGG orthology (55), and identify BLASTX (best hit) matches to *A. thaliana*. About 3.85 million functional elements were annotated, including InterPro domains (51), Pfam domains (53) (Supplementary Table S6). TPM values for 308 sets of RNA-seq data associated with 10 genomes, and variations of 1251 sets for re-sequencing data were integrated in BRAD. Through standardized calculation, the uniformity of data was greatly improved.

Organization of features

Pop-up retrieval system. Multiple web pages formed the user interface and may be navigated between each other through a pop-up window (Figure 2A). Any gene ID within the database may be clicked to trigger the pop-up window, and then pages selected within this to navigate to and retrieve the gene. A range of relevant information may thus conveniently be retrieved using the pop-up retrieval system.

The major features of the user interface were organized into three menus, *Browse*, *Search* and *Tools*. In BRAD V3.0 the original *Browse* sub-menu was retained, apart from adding 'Species info' for a page to show the species where corresponding genomes has been collated. Most of the new features and updated features are located within the *Search* and *Tools* menus. Under *Search*, the original 'Flanking region' and 'Synteny @ Genome' ('Syntenic Figure' in BRAD V2.0) capability were retained. Minor changes were made to improve data retrieval and presentation for the 'Annotations', 'Nonsyntenic Orth' and 'Gene sequence'. Substantial changes were made for 'Syntenic gene @ Subgenome' ('Syntenic genes' in BRAD V2.0). New features 'Transcriptome' and 'Polymorph' were also added in this menu. Under the *Tools*, the BLAST and genome browser were retained. Moreover, new tools were added, including 'Multiple sequence alignment (MSA)', 'Gene Phylogeny', 'Genomic sequence retrieval', 'MicroSynteny' and 'Primer Design'.

Major updates and new features in the menus of *Search* and *Tools* are highlighted below.

Search

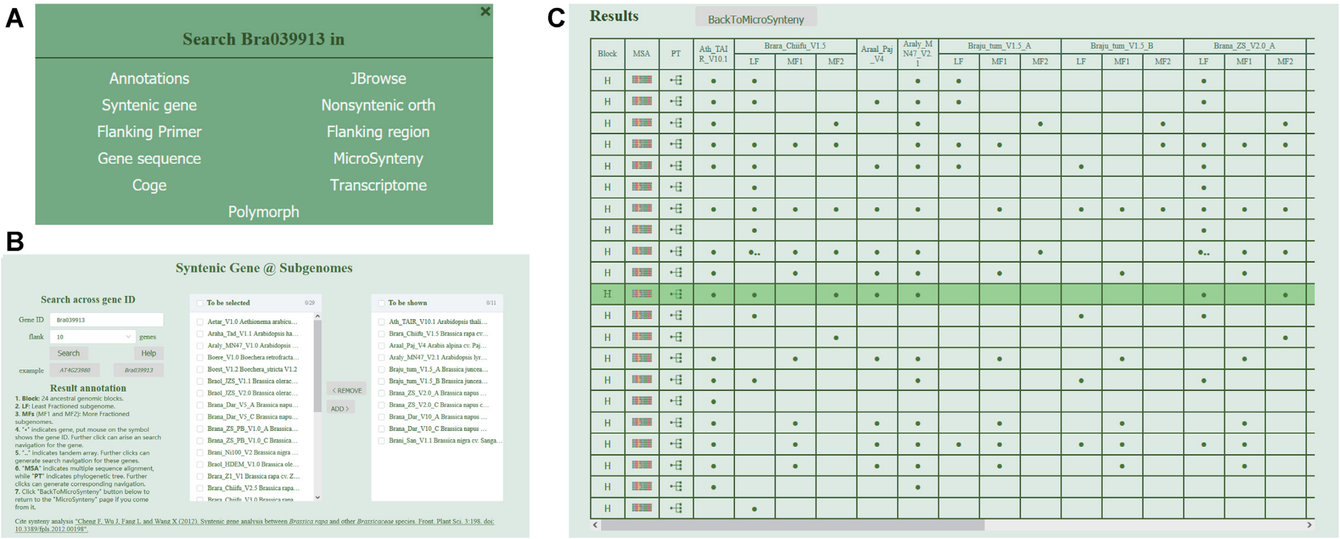
Syntenic Gene @ Subgenome. This is a redesigned syntenic gene retrieval system across multiple reference genomes. By adding or removing genomes and subgenomes one by one to the 'To be shown' selection, users are able to order these as required (Figure 2B). We retained the same table style to indicate which syntenic genes were retrieved. In addition, two buttons were added to the table to allow 'Multiple sequence alignment' and 'Gene Phylogeny' to be readily displayed for the retrieved genes (Figure 2C).

Polymorph. This is a new module developed to make use of the increasing availability of genomic variation data. Across the 5 data sets, there are a total of ~2.25M SNPs (from 199 samples against Chiifu V3.0) (42). For the *B. rapa* sets, there are 3.97M (524 on Chiifu V3.0) (41). For *B. oleracea* 3.85M (119 on JZS V1.1) (42), and 6.13M (119 on JZS V2.0). For *B. napus* 2.05M SNP (608 on Darmor-bzh V5) (43), respectively. Variation for each sample is displayed in the table and may be downloaded as an .xlsx file (Figure 3A). 'Polymorph' is not only a tool to search for SNP variations, but more importantly allows analysis of variation features among the genetic groups. We classified all genotype accessions into different groups based on either morphotypic features or whole genome phylogenetic branches. We added a function calculating P-value to help users better evaluate the role of variation among different groups. Users can optionally select query group and background group for query. The branches with different colour represent different groups, and the dots at the end of each branch represent its genotype (Figure 3C). The phylogenetic tree may be downloaded as an SVG file, which is very helpful for the users to see the distribution of the variation. Taking the variation site A09 25472702 in gene BraA09g032840.3C as an example, we selected Chinese cabbage as a query and all other groups as background. The P-value of 2.515576e–37 between query and background, indicated the variation site is strongly selected in Chinese cabbage (*B. rapa* subsp. *pekinensis*) (Figure 3B). Based on this module, we identified some unique variation sites in Chinese cabbage (Supplementary Table S7).

Other updates. The Swiss-Prot and TrEMBL annotations were replaced with Pfam domain annotations. The annotation ID in the result set links to GO, InterPro, KEGG and Pfam databases to query the corresponding detailed information. 'BLASTX to *A. thaliana*' links the target gene to the model plant *A. thaliana* (Figure 4A). There are four genomes (*A. thaliana* TAIR V10.1, *B. rapa* cv. Chiifu V1.5, *B. rapa* cv. Chiifu V3.0, *B. oleracea* cv. JZS V2.0) that could be used as the subject genome of BLASTP to retrieve nonsyntenic orthologous genes (Figure 4B), whereas BRAD V2.0 only had *Arabidopsis thaliana* V10.1. By entering the gene ID or coordinate and selecting the flank length, the genes existing in the region would be identified and shown in the table (Figure 4C). Location information and full-length sequence, CDS and protein sequence of the gene may be searched by gene ID (Figure 4D). The expression values of a gene in different tissues may be shown in images and tables (Figure 4E).

Tools

MicroSynteny. This feature complements the 'Syntenic Gene @ Subgenome', and visualizes additional syntenic details between two genomic genes or regions. Both the gene ID and the genome region can be used as input in this module, and homologous regions between two segments are indicated in (Figure 5A). In addition, corresponding links were made through the pop-up retrieval system in 'Syntenic Gene @ Subgenome', so that the regional synteny between syntenic gene fragments may be queried conveniently. The images are able to be saved in SVG, PNG and JPG formats by right-clicking, and also expanded horizontally by scrolling up and down. Each horizontal rectangle may be clicked to trigger the pop-up window. The homologous fragments between the two fragments can also be further clicked to see the result of sequence alignment.



Search Bra039913 in

Annotations | JBrowse | Syntenic gene | Nonsyntenic orth | Flanking Primer | Flanking region | Gene sequence | MicroSynteny | Coge | Transcriptome | Polymorph

Syntenic Gene @ Subgenomes

Search across gene ID

Gene ID Bra039913

flank 10 genes

Search | Help

example AT4G23980 | Bra039913

Result annotation

1. **Block:** 24 ancestral genomic blocks.
2. **LF:** Least Fractioned subgenome.
3. **MFs** (MF1 and MF2): More Fractioned subgenomes.
4. "•" indicates gene, put mouse on the symbol shows the gene ID. Further click can arise an search navigation for the gene.
5. ".." indicates tandem array. Further clicks can generate search navigation for these genes.
6. "**MSA**" indicates multiple sequence alignment, while "**PT**" indicates phylogenetic tree. Further clicks can generate corresponding navigation.
7. Click "BackToMicroSynteny" button below to return to the "MicroSynteny" page if you come from it.

☐ To be selected 0/29

☐ Aetar_V1.0 Aethionema arabicu…
☐ Araha_Tad_V1.1 Arabidopsis ha…
☐ Araly_MN47_V1.0 Arabidopsis …
☐ Boere_V1.0 Boechera retrofracta…
☐ Boest_V1.2 Boechera_stricta V1.2
☐ Braol_JZS_V1.1 Brassica olerac…
☐ Braol_JZS_V2.0 Brassica olerac…
☐ Brana_Dar_V5_A Brassica napu…
☐ Brana_Dar_V5_C Brassica napu…
☐ Brana_ZS_PB_V1.0_A Brassica…
☐ Brana_ZS_PB_V1.0_C Brassica…
☐ Brani_Ni100_V2 Brassica nigra …
☐ Braol_HDEM_V1.0 Brassica ole…
☐ Brara_Z1_V1 Brassica rapa cv. Z…
☐ Brara_Chiifu_V2.5 Brassica rapa…
☐ Brara_Chiifu_V3.0 Brassica rapa

< REMOVE | ADD >

☐ To be shown 0/11

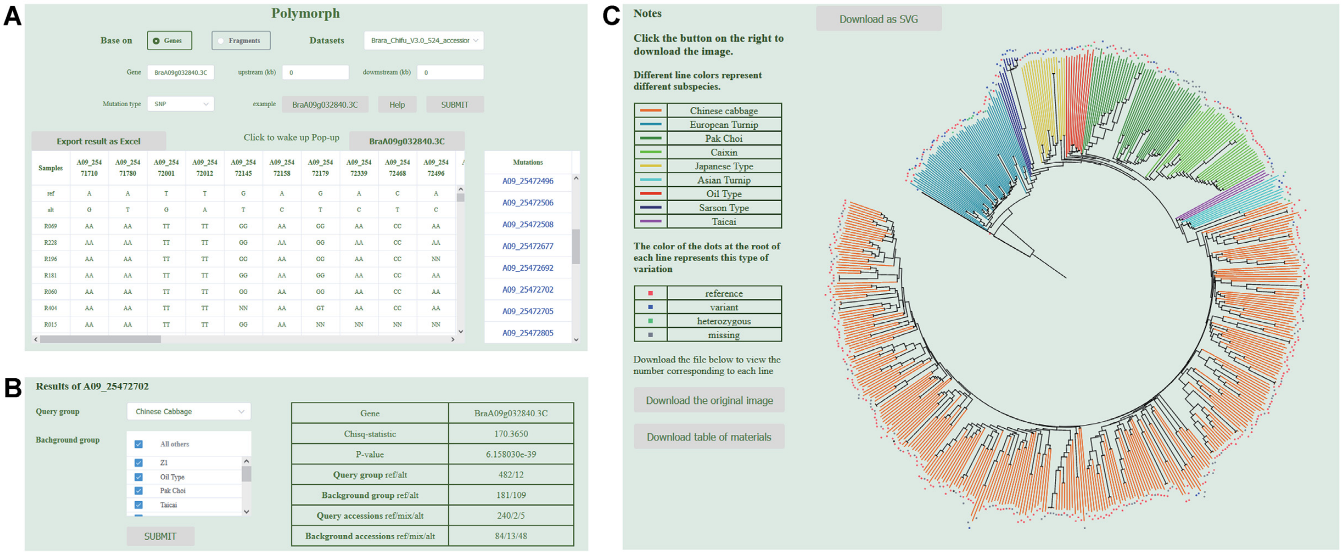
☐ Ath_TAIR_V10.1 Arabidopsis thali…
☐ Brara_Chiifu_V1.5 Brassica rapa cv…
☐ Araal_Paj_V4 Arabis alpina cv. Paj…
☐ Araly_MN47_V2.1 Arabidopsis lyr…
☐ Braju_tum_V1.5_A Brassica juncea…
☐ Braju_tum_V1.5_B Brassica juncea…
☐ Brana_ZS_V2.0_A Brassica napus …
☐ Brana_ZS_V2.0_C Brassica napus c…
☐ Brana_Dar_V10_A Brassica napus …
☐ Brana_Dar_V10_C Brassica napus …
☐ Brani_San_V1.1 Brassica nigra cv. Sanga…

Cite synteny analysis "Cheng F, Wu J, Fang L and Wang X (2012). Syntenic gene analysis between *Brassica rapa* and other *Brassicaceae* species. Front. Plant Sci. 3:198. doi: 10.3389/fpls.2012.00198".

Results | BackToMicroSynteny

Block	MSA	PT	Ath_TAIR_V10.1	Brara_Chiifu_V1.5			Araal_Paj_V4	Araly_MN47_V2.1	Braju_tum_V1.5_A			Braju_tum_V1.5_B			Brana_ZS_V2.0_A		
				LF	MF1	MF2			LF	MF1	MF2	LF	MF1	MF2	LF	MF1	MF2
H	▦	⊣	●	●				●	●						●		
H	▦	⊣	●	●			●	●	●						●		
H	▦	⊣	●			●		●			●			●			●
H	▦	⊣	●	●	●	●		●	●	●				●	●	●	●
H	▦	⊣	●	●			●	●	●			●			●		
H	▦	⊣		●											●		
H	▦	⊣	●	●	●	●	●	●		●		●	●	●	●	●	●
H	▦	⊣		●											●		
H	▦	⊣	●	●..	●	●	●	●			●				●..	●	●
H	▦	⊣	●		●		●	●		●			●			●	
H	▦	⊣	●	●		●	●	●							●		●
H	▦	⊣		●								●			●		
H	▦	⊣				●											●
H	▦	⊣	●		●		●	●		●			●			●	
H	▦	⊣	●	●				●				●			●		
H	▦	⊣	●														
H	▦	⊣	●		●		●	●		●			●			●	
H	▦	⊣	●		●		●	●	●	●		●	●		●	●	
H	▦	⊣	●		●		●	●		●			●			●	
H	▦	⊣	●					●									
H	▦	⊣		●													

Figure 2. Search of Syntenic Gene @ Subgenome. (**A**) Pop-up retrieval system. (**B**) Search Syntenic genes in multiple genomes and subgenomes. (**C**) Result of syntenic genes.



Polymorph

Base on ● Genes ○ Fragments Datasets Brara_Chifu_V3.0_524_accessior

Gene BraA09g032840.3C upstream (kb) 0 dowmstream (kb) 0

Mutation type SNP example BraA09g032840.3C | Help | SUBMIT

Export result as Excel | Click to wake up Pop-up | BraA09g032840.3C

Samples	A09_25471710	A09_25471780	A09_25472001	A09_25472012	A09_25472145	A09_25472158	A09_25472179	A09_25472339	A09_25472468	A09_25472496
ref	A	A	T	T	G	A	G	A	C	A
alt	G	T	G	A	T	C	T	C	T	C
R069	AA	AA	TT	TT	GG	AA	GG	AA	CC	AA
R228	AA	AA	TT	TT	GG	AA	GG	AA	CC	AA
R196	AA	AA	TT	TT	GG	AA	GG	AA	CC	NN
R181	AA	AA	TT	TT	GG	AA	GG	AA	CC	AA
R060	AA	AA	TT	TT	GG	AA	GG	AA	CC	AA
R404	AA	AA	TT	TT	NN	AA	GT	AA	CC	AA
R015	AA	AA	TT	TT	GG	AA	NN	NN	NN	NN

Mutations
A09_25472496
A09_25472506
A09_25472508
A09_25472677
A09_25472692
A09_25472702
A09_25472705
A09_25472805

Results of A09_25472702

Query group Chinese Cabbage

Bachground group

☑ All others
☑ Z1
☑ Oil Type
☑ Pak Choi
☑ Taicai

SUBMIT

Gene	BraA09g032840.3C
Chisq-statistic	170.3650
P-value	6.158030e-39
Query group ref/alt	482/12
Background group ref/alt	181/109
Query accessions ref/mix/alt	240/2/5
Background accessions ref/mix/alt	84/13/48

Notes

Click the button on the right to download the image.

Different line colors represent different subspecies.

🟧	Chinese cabbage
🟦	European Turnip
🟩	Pak Choi
🟩	Caixin
🟨	Japanese Type
🟦	Asian Turnip
🟥	Oil Type
🟦	Sarson Type
🟪	Taicai

The color of the dots at the root of each line represents this type of variation

🟥	reference
🟦	variant
🟩	heterozygous
■	missing

Download the file below to view the number corresponding to each line

Download as SVG | Download the original image | Download table of materials

Figure 3. Variation data retrieval. (**A**) Variation retrieval and dynamic table display. (**B**) Variation distribution of different groups and the result of chi-square test. (**C**) Visualization of the distribution of variation in different groups.

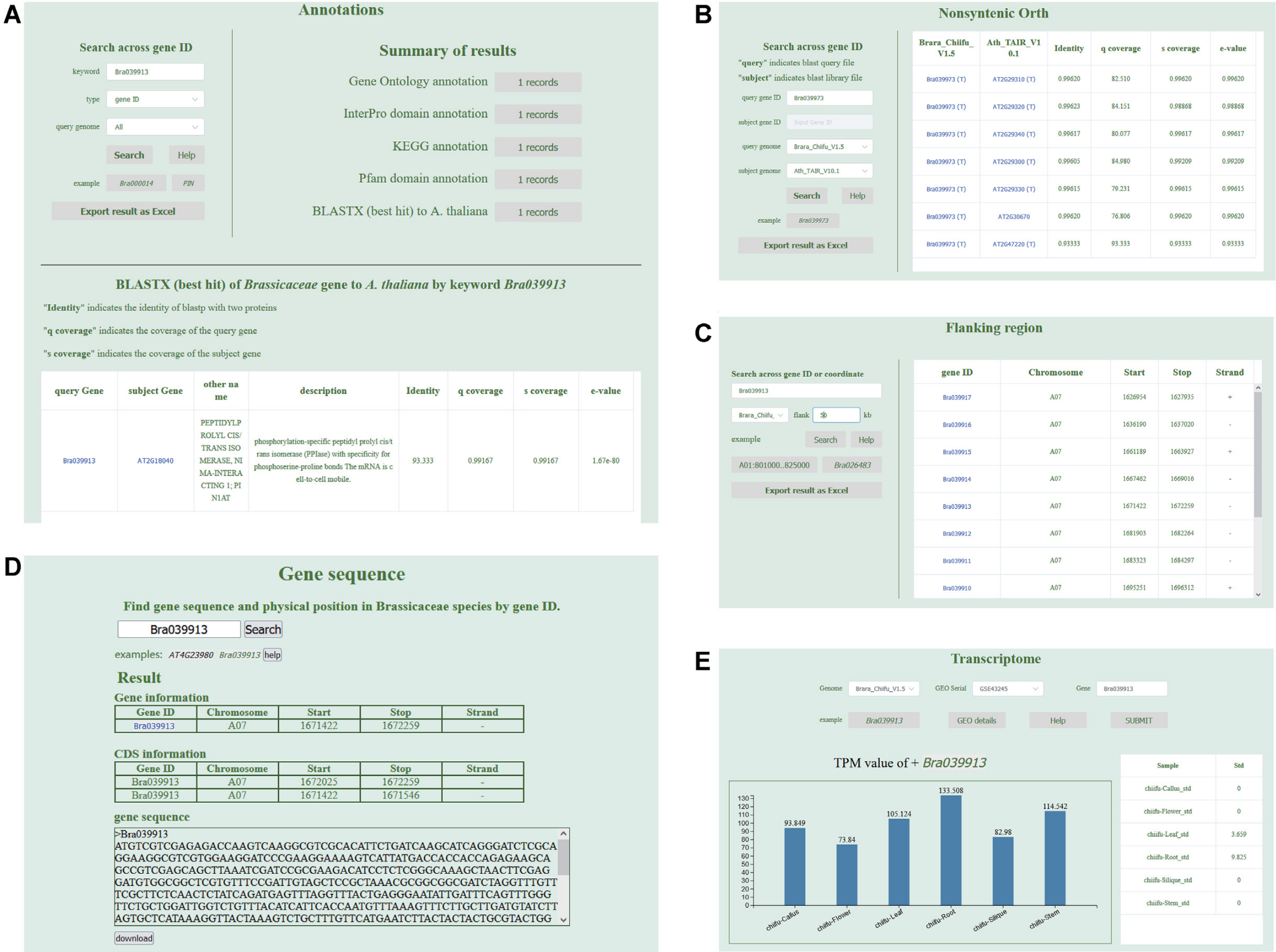
Primer design. This module was provided to design primers within selected genes (Figure 5B). When the user navigates from the pop-up window to this feature, the gene is presented as the default input, and a 150 bp flanking region is used to design the primer that would amplify the full length of the gene. All parameters are set with hyperlinks to the Primer3 manual (http://primer3.org/manual.html). In addition, a sequence can be input for primer design by selecting the button 'Fragment'.

Multiple sequence alignment (MSA) and Gene Phylogeny (PT). These are two new features provided to display the evolutionary relationship between protein and CDS sequences, which will help users further interpret the target gene. The second and third columns of syntenic gene results represent MSA and PT of all genes in each row (Figure 5C, D). Users are also free to add or subtract genes for queries.

Genomic sequence retrieval. The corresponding genome sequence may be obtained by submitting the information of reference genome, chromosome, start and stop position and strand, and downloaded as a .fasta format file.

BLAST. We integrated all the collected genome sequences, CDS sequences and protein sequences into Web BLAST so that users may perform BLAST retrieval of target sequence. The output may be downloaded in three different formats, HTML, TSV and XML.

Genome browser. We have implemented a genome browser using JBrowse together with plugins (63) to visualize



Search across gene ID

keyword: Bra039913

type: gene ID

query genome: All

Summary of results

Gene Ontology annotation	1 records
InterPro domain annotation	1 records
KEGG annotation	1 records
Pfam domain annotation	1 records
BLASTX (best hit) to A. thaliana	1 records

BLASTX (best hit) of *Brassicaceae* gene to *A. thaliana* by keyword *Bra039913*

"**Identity**" indicates the identity of blastp with two proteins

"**q coverage**" indicates the coverage of the query gene

"**s coverage**" indicates the coverage of the subject gene

query Gene	subject Gene	other name	description	Identity	q coverage	s coverage	e-value
Bra039913	AT2G18040	PEPTIDYLPROLYL CIS/TRANS ISOMERASE, NIMA-INTERACTING 1; PIN1AT	phosphorylation-specific peptidyl prolyl cis/trans isomerase (PPIase) with specificity for phosphoserine-proline bonds The mRNA is cell-to-cell mobile.	93.333	0.99167	0.99167	1.67e-80

Search across gene ID

"**query**" indicates blast query file

"**subject**" indicates blast library file

query gene ID: Bra039973

subject gene ID: Input Gene ID

query genome: Brara_Chiifu_V1.5

subject genome: Ath_TAIR_V10.1

Brara_Chiifu_V1.5	Ath_TAIR_V10.1	Identity	q coverage	s coverage	e-value
Bra039973 (T)	AT2G29310 (T)	0.99620	82.510	0.99620	0.99620
Bra039973 (T)	AT2G29320 (T)	0.99623	84.151	0.98868	0.98868
Bra039973 (T)	AT2G29340 (T)	0.99617	80.077	0.99617	0.99617
Bra039973 (T)	AT2G29300 (T)	0.99605	84.980	0.99209	0.99209
Bra039973 (T)	AT2G29330 (T)	0.99615	79.231	0.99615	0.99615
Bra039973 (T)	AT2G30670	0.99620	76.806	0.99620	0.99620
Bra039973 (T)	AT2G47220 (T)	0.93333	93.333	0.93333	0.93333

Search across gene ID or coordinate

Bra039913

Brara_Chiifu_ flank 50 kb

gene ID	Chromosome	Start	Stop	Strand
Bra039917	A07	1626954	1627935	+
Bra039916	A07	1636190	1637020	-
Bra039915	A07	1661189	1663927	+
Bra039914	A07	1667462	1669016	-
Bra039913	A07	1671422	1672259	-
Bra039912	A07	1681903	1682264	-
Bra039911	A07	1683323	1684297	-
Bra039910	A07	1695251	1696312	+

Find gene sequence and physical position in Brassicaceae species by gene ID.

Bra039913

examples: *AT4G23980* *Bra039913*

Result

Gene information

Gene ID	Chromosome	Start	Stop	Strand
Bra039913	A07	1671422	1672259	-

CDS information

Gene ID	Chromosome	Start	Stop	Strand
Bra039913	A07	1672025	1672259	-
Bra039913	A07	1671422	1671546	-

gene sequence

```
>Bra039913
ATGTCGTCGAGAGACCAAGTCAAGGCGTCGCACATTCTGATCAAGCATCAGGGATCTCGCA
GGAAGGCGTCGTGGAAGGATCCCGAAGGAAAAGTCATTATGACCACCACCAGAGAAGCA
GCCGTCGAGCAGCTTAAATCGATCCGCGAAGACATCCTCTCGGGCAAAGCTAACTTCGAG
GATGTGGCGGCTCGTGTTTCCGATTGTAGCTCCGCTAAACGCGGCGGCGATCTAGGTTTGTT
TCGCTTCTCAACTCTATCAGATGAGTTTAGGTTTACTGAGGGAATATTGATTTCAGTTTGGG
TTCTGCTGGATTGGTCTGTTTACATCATTCACCAATGTTTAAAGTTTCTTGCTTGATGTATCTT
AGTGCTCATAAAGGTTACTAAAGTCTGCTTTGTTCATGAATCTTACTACTACTGCGTACTGG
```

Genome: Brara_Chiifu_V1.5 GEO Serial: GSE43245 Gene: Bra039913

Sample	Std
chiifu-Callus_std	0
chiifu-Flower_std	0
chiifu-Leaf_std	3.659
chiifu-Root_std	9.825
chiifu-Silique_std	0
chiifu-Stem_std	0

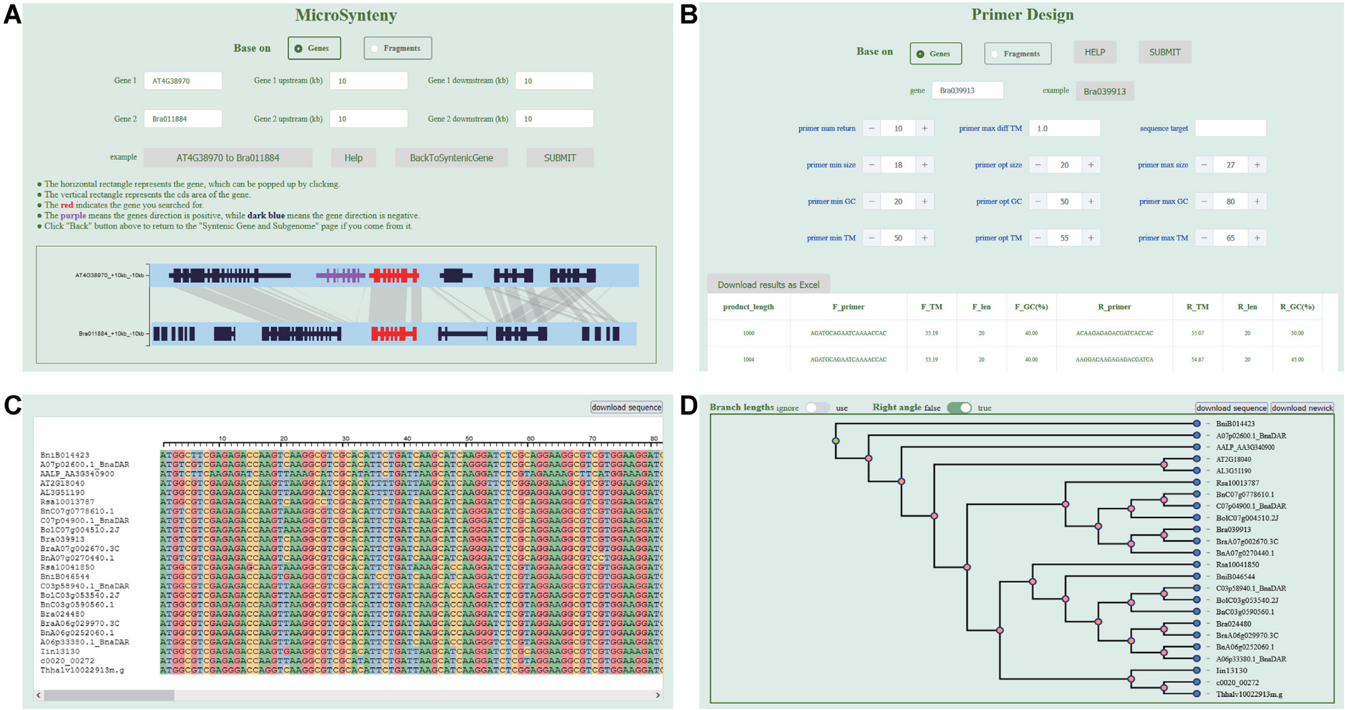
Figure 4. Several modules of BRAD. (**A**) Gene annotation information retrieval. (**B**) Non-syntenic ortholog genes search module. (**C**) Table showing the genes in flanking region. (**D**) Summary of gene sequence information. (**E**) Bar plot and table showing the gene expression level in TPM.

genome sequences, gene models, markers, variations and omics data. JBrowse is more suited to visualizing multiple omics data than the GBrowse 2.0 (69) used by BRAD V2.0. Users may conveniently and efficiently access a region or gene of interest either by entering location or by selecting the 'JBrowse' option in the pop-up window. The tracks of reference sequence and gene models are embedded within this module.

DISCUSSION

We updated BRAD by rebuilding the whole system with the MVVM framework and including new genomic data. In the updated BRAD, the synteny analysis system incorporates tools to analyse synteny at different levels, including genome, genomic region, gene and sequence levels. The provision of a wide range of synteny analysis tools will not only help genome comparison, but also help in unravelling gene function and in molecular marker design. We provided 'Syntenic Gene @ Subgenome' as a central module, so that users are able to make queries by navigating through the pop-up window to 'MicroSynteny', 'Multiple sequence alignment' and 'Gene Phylogeny' conveniently. 'MicroSynteny' allows users to view detailed homology alignments between genomic segments where syntenic genes are located. The 'Multiple sequence alignment' and 'Gene Phylogeny' tools help users better explore and understand the evolutionary relationships between syntenic genes. Moreover, within the 'Syntenic Gene @ Subgenome' we added two windows for selection and removal of genomes. This optimization provides considerable flexibility for organizing result output.

'Polymorph' is designed not only for showing and retrieving genome variation data, but also mining these data. For generating the genome variation data, we allowed genotype accessions to be organised into groups based on phylogenetic tree and morphotypes. For any query group, we provided a test of P-value for the distribution of the SNP in the query and the selected background. Polymorph represents a powerful tool for users to relate variation data to important developmental or agronomic traits.



MicroSynteny

Base on ● Genes ○ Fragments

Gene 1: AT4G38970 | Gene 1 upstream (kb): 10 | Gene 1 dowmstream (kb): 10

Gene 2: Bra011884 | Gene 2 upstream (kb): 10 | Gene 2 dowmstream (kb): 10

example: AT4G38970 to Bra011884 | Help | BackToSyntenicGene | SUBMIT

- The horizontal rectangle represents the gene, which can be popped up by clicking.
- The vertical rectangle represents the cds area of the gene.
- The **red** indicates the gene you searched for.
- The **purple** means the genes direction is positive, while **dark blue** means the gene direction is negative.
- Click "Back" button above to return to the "Syntenic Gene and Subgenome" page if you come from it.

Primer Design

Base on ● Genes ○ Fragments | HELP | SUBMIT

gene: Bra039913 | example: Bra039913

primer num return: − 10 + | primer max diff TM: 1.0 | sequence target:

primer min size: − 18 + | primer opt size: − 20 + | primer max size: − 27 +

primer min GC: − 20 + | primer opt GC: − 50 + | primer max GC: − 80 +

primer min TM: − 50 + | primer opt TM: − 55 + | primer max TM: − 65 +

Download results as Excel

product_length	F_primer	F_TM	F_len	F_GC(%)	R_primer	R_TM	R_len	R_GC(%)
1000	AGATGCAGAATCAAAACCAC	55.19	20	40.00	ACAAGAGAGACGATCACCAC	55.07	20	50.00
1004	AGATGCAGAATCAAAACCAC	55.19	20	40.00	AAGGACAAGAGAGACGATCA	54.87	20	45.00

download sequence

```
                             10        20        30        40        50        60        70        80
BniB014423           ATGGCTTCGAGAGACCAAGTCAAGGCGTCGCACATTCTGATCAAGCATCAAGGATCTCGCAGGAAGGCGTCGTGGAAGGATC
A07p02600.1_BnaDAR   ATGTCGTCGAGAGACCAAGTCAAGGCGTCGCACATTCTGATCAAGCATCAGGGATCTCGCAGGAAGGCGTCGTGGAAGGATC
AALP_AA3G340900      ATGTCTTCAAGAGATCAAGTTAAAGCATCGCATATTCTGATTAAGCATCAAGGATCTCGTAGAAAAGCTTCATGGAAAGATC
AT2G18040            ATGGCGTCGAGAGACCAAGTTAAGGCATCGCACATTTTGATTAAGCATCAAGGTTCTCGGAGGAAAGCGTCGTGGAAGGATC
AL3G51190            ATGGCGTCGAGAGACCAAGTTAAGGCATCGCACATTTTGATTAAGCATCAAGGATCTCGGAGGAAGGCGTCGTGGAAGGATC
Rsa10013787          ATGGCGTCGAGAGACCAAGTCAAGGCCTCGCACATTCTGATCAAGCATCAAGGATCTCGCAGGAAGGCGTCGTGGAAGGATC
BnC07g0778610.1      ATGTCGTCGAGAGACCAAGTAAAGGCGTCGCACATTCTGATCAAGCATCAGGGATCTCGCAGGAAGGCGTCGTGGAAGGATC
C07p04900.1_BnaDAR   ATGTCGTCGAGAGACCAAGTAAAGGCGTCGCACATTCTGATCAAGCATCAGGGATCTCGCAGGAAGGCGTCGTGGAAGGATC
BolC07g004510.2J     ATGTCGTCGAGAGACCAAGTAAAGGCGTCGCACATTCTGATCAAGCATCAGGGATCTCGCAGGAAGGCGTCGTGGAAGGATC
Bra039913            ATGTCGTCGAGAGACCAAGTCAAGGCGTCGCACATTCTGATCAAGCATCAGGGATCTCGCAGGAAGGCGTCGTGGAAGGATC
BraA07g002670.3C     ATGTCGTCGAGAGACCAAGTCAAGGCGTCGCACATTCTGATCAAGCATCAGGGATCTCGCAGGAAGGCGTCGTGGAAGGATC
BnA07g0270440.1      ATGTCGTCGAGAGACCAAGTCAAGGCGTCGCACATTCTGATCAAGCATCAGGGATCTCGCAGGAAGGCGTCCTGGAAGGATC
Rsa10041850          ATGGCGTCGAGAGAGCAAGTAAAGGCGTCGCACATTCTGATAAAGCACCAAGGATCTCGTAGGAAGGCGTCGTGGAAGGATC
BniB046544           ATGGCGTCGAGAGACCAAGTGAAGGCGTCGCACATCCTGATCAAGCATCAAGGATCTCGTAGGAAGGCGTCGTGGAAGGATC
C03p58940.1_BnaDAR   ATGGCGTCGAGAGACCAAGTTAAGGCGTCGCACATTCTGATCAAGCACCAAGGATCTCGTAGGAAGGCGTCGTGGAAGGATC
BolC03g053540.2J     ATGGCGTCGAGAGACCAAGTTAAGGCGTCGCACATTCTGATCAAGCACCAAGGATCTCGTAGGAAGGCGTCGTGGAAGGATC
BnC03g0590560.1      ATGGCGTCGAGAGACCAAGTTAAGGCGTCGCACATTCTGATCAAGCACCAAGGATCTCGTAGGAAGGCGTCGTGGAAGGATC
Bra024480            ATGGCGTCGAGAGACCAAGTTAAGGCGTCGCACATTCTGATCAAGCACCAAGGATCTCGTAGGAAGGCGTCGTGGAAGGATC
BraA06g029970.3C     ATGGCGTCGAGAGACCAAGTTAAGGCGTCGCACATTCTGATCAAGCACCAAGGATCTCGTAGGAAGGCGTCGTGGAAGGATC
BnA06g0252060.1      ATGGCGTCGAGAGACCAAGTTAAGGCGTCGCACATTCTGATCAAGCACCAAGGGTCTCGTAGGAAGGCGTCGTGGAAGGATC
A06p33380.1_BnaDAR   ATGGCGTCGAGAGACCAAGTTAAGGCGTCGCACATTCTGATCAAGCACCAAGGGTCTCGTAGGAAGGCGTCGTGGAAGGATC
Iin13130             ATGGCGTCGAGAGACCAAGTGAAGGCGTCGCACATTCTGATTAAGCATCAAGGATCTCGCAGGAAGGCGTCGTGGAAAGATC
c0020_00272          ATGGCGTCGAGAGACCAAGTTAAGGCGTCGCATATTCTGATTAAGCATCAAGGATCTCGTAGGAAGGCGTCGTGGAAGGATC
Thhalv10022913m.g    ATGGCGTCGAGGGACCAGGTCAAGGCGTCGCACATTCTGATTAAGCATCAAGGATCTCGGAGGAAGGCGTCGTGGAAGGATC
```

Branch lengths ignore ○ use | **Right angle** false ● true | download sequence | download newick

Figure 5. Some tools of BRAD. (**A**) Graphically show synteny between genomic region pairs. (**B**) Primers were designed to amplify the entire gene. (**C**) Multiple sequence alignment of syntenic genes. (**D**) Phylogenetic tree of syntenic genes.

Multiple new types of Brassicaceae genomic data, such as Hi-C, ChIP-seq, ATAC-seq, DAP-seq, methylation, single cell transcriptome etc., have been released over the past five years. However, a comprehensive database to integrate multiomics data for this important taxon has been lacking. The *Arabidopsis* eFP browser is an excellent tool for displaying gene expression profile data (70). With the accumulation of well-organized transcriptome experiments, especially in the extensively investigated *Brassica* crops, building an eFP browser in BRAD will increase the value of the database. Pan-genomes have been constructed for all *Brassica* species (71) and *Raphunas sativus* (72). Several genomes of *B. oleracea* have also been reported. Designing tools to analyse and visualize the pan-genome data is obviously an urgent need for the *Brassica* research community, and BRAD will continue to advance this effort.

DATA AVAILABILITY

The BRAD can be accessed through the web server at http://brassicadb.cn. Transcriptome data processing and some of the backend code BRAD uses is available on GitHub repository at https://github.com/dahaigui/BRAD-scrpit.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The calculation of this work is done on the computing server of Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences.

FUNDING

National Programon Key Research Project [2021YFF1000047]; National Natural Science Foundation of China [31630068]; National Program on Key Research Project [2016YFD0100307]; Agricultural Science and Technology Innovation Program (ASTIP); Key Laboratory of Biology and Genetic Improvement of Horticultural Crops, Ministry of Agriculture, P.R. China. Funding for open access charge: Research projects funded to Xiaowu Wang.
Conflict of interest statement. None declared.

REFERENCES

1. Cheng,F., Liu,S., Wu,J., Fang,L., Sun,S., Liu,B., Li,P., Hua,W. and Wang,X. (2011) BRAD, the genetics and genomics database for *Brassica* plants. *BMC Plant Biol.*, **11**, 136.
2. Francis,A., Lujan-Toro,B.E., Warwick,S.I., Macklin,J.A. and Martin,S.L. (2021) Update on the Brassicaceae species checklist. *Biodivers Data J.*, **9**, e58773.
3. Avato,P. and Argentieri,M.P. (2015) Brassicaceae: a rich source of health improving phytochemicals. *Phytochem Rev.*, **14**, 1019–1033.
4. Wang,X., Wang,H., Wang,J., Sun,R., Wu,J., Liu,S., Bai,Y., Mun,J.H., Bancroft,I., Cheng,F. *et al.* (2011) The genome of the mesopolyploid crop species *Brassica* rapa. *Nat Genet.*, **43**, 1035–1039.
5. Huala,E., Dickerman,A.W., Garcia-Hernandez,M., Weems,D., Reiser,L., LaFond,F., Hanley,D., Kiphart,D., Zhuang,M., Huang,W. *et al.* (2001) The *Arabidopsis* Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. *Nucleic Acids Res.*, **29**, 102–105.
6. Wang,X., Wu,J., Liang,J., Cheng,F. and Wang,X. (2015) *Brassica* database (BRAD) version 2.0: integrating and mining Brassicaceae species genomic resources. *Database (Oxford)*., **2015**, bav093.
7. Arabidopsis Genome,I. (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*., **408**, 796–815.

8. Hu,T.T., Pattyn,P., Bakker,E.G., Cao,J., Cheng,J.F., Clark,R.M., Fahlgren,N., Fawcett,J.A., Grimwood,J., Gundlach,H. *et al.* (2011) The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nat Genet*., **43**, 476–481.
9. Kliver,S., Rayko,M., Komissarov,A., Bakin,E., Zhernakova,D., Prasad,K., Rushworth,C., Baskar,R., Smetanin,D., Schmutz,J. *et al.* (2018) Assembly of the *Boechera retrofracta* genome and evolutionary analysis of apomixis-associated genes. *Genes (Basel)*., **9**, 4.
10. Haudry,A., Platts,A.E., Vello,E., Hoen,D.R., Leclercq,M., Williamson,R.J., Forczek,E., Joly-Lopez,Z., Steffen,J.G., Hazzouri,K.M. *et al.* (2013) An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions. *Nat. Genet*., **45**, 891–898.
11. Song,X., Wei,Y., Xiao,D., Gone,K., Sun,P., Ren,Y., Yuan,J., Wu,T., Yang,Q., Li,X. *et al.* (2021) Brassica carinata genome characterization clarifies U's triangle model of the evolution and polyploidy in *Brassica*. *Plant Physiol*., **186**, 388–406.
12. Liu,S., Liu,Y., Yang,X., Tong,C., Edwards,D., Parkin,I.A., Zhao,M., Ma,J., Yu,J., Huang,S. *et al.* (2014) The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. *Nat. Commun*., **5**, 3930.
13. Cai,C., Wang,X., Liu,B., Wu,J., Liang,J., Cui,Y., Cheng,F. and Wang,X. (2017) *Brassica rapa* Genome 2.0: a reference upgrade through sequence re-assembly and gene re-annotation. *Mol. Plant*., **10**, 649–651.
14. Slotte,T., Hazzouri,K.M., Agren,J.A., Koenig,D., Maumus,F., Guo,Y.L., Steige,K., Platts,A.E., Escobar,J.S., Newman,L.K. *et al.* (2013) The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nat. Genet*., **45**, 831–835.
15. Gan,X., Hay,A., Kwantes,M., Haberer,G., Hallab,A., Ioio,R.D., Hofhuis,H., Pieper,B., Cartolano,M., Neumann,U. *et al.* (2016) The *Cardamine hirsuta* genome offers insight into the evolution of morphological diversity. *Nat. Plants*., **2**, 16167.
16. Belser,C., Istace,B., Denis,E., Dubarry,M., Baurens,F.C., Falentin,C., Genete,M., Berrabah,W., Chevre,A.M., Delourme,R. *et al.* (2018) Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. *Nat. Plants*., **4**, 879–887.
17. Kang,M., Wu,H., Yang,Q., Huang,L., Hu,Q., Ma,T., Li,Z. and Liu,J. (2020) A chromosome-scale genome assembly of *Isatis indigotica*, an important medicinal plant used in traditional Chinese medicine: An *Isatis* genome. *Hortic. Res*., **7**, 18.
18. Zhang,X., Yue,Z., Mei,S., Qiu,Y., Yang,X., Chen,X., Cheng,F., Wu,Z., Sun,Y., Jing,Y. *et al.* (2015) A *de novo* genome of a Chinese radish cultivar. *Hortic. Plant J*., **03**, 10.
19. Dorn,K.M., Fankhauser,J.D., Wyse,D.L. and Marks,M.D. (2015) A draft genome of field pennycress (*Thlaspi arvense*) provides tools for the domestication of a new winter biofuel crop. *DNA Res*., **22**, 121–131.
20. Kagale,S., Koh,C., Nixon,J., Bollina,V., Clarke,W.E., Tuteja,R., Spillane,C., Robinson,S.J., Links,M.G., Clarke,C. *et al.* (2014) The emerging biofuel crop *Camelina sativa* retains a highly undifferentiated hexaploid genome structure. *Nat. Commun*., **5**, 3706.
21. Briskine,R.V., Paape,T., Shimizu-Inatsugi,R., Nishiyama,T., Akama,S., Sese,J. and Shimizu,K.K. (2017) Genome assembly and annotation of *Arabidopsis halleri*, a model for heavy metal hyperaccumulation and evolutionary ecology. *Mol. Ecol. Resour*., **17**, 1025–1036.
22. Willing,E.M., Rawat,V., Mandakova,T., Maumus,F., James,G.V., Nordstrom,K.J., Becker,C., Warthmann,N., Chica,C., Szarzynska,B. *et al.* (2015) Genome expansion of *Arabis alpina* linked with retrotransposition and reduced symmetric DNA methylation. *Nat Plants*., **1**, 14023.
23. Zhang,T., Qiao,Q., Novikova,P.Y., Wang,Q., Yue,J., Guan,Y., Ming,S., Liu,T., De,J., Liu,Y. *et al.* (2019) Genome of *Crucihimalaya himalaica*, a close relative of Arabidopsis, shows ecological adaptation to high altitude. *Proc. Natl. Acad. Sci. U.S.A*., **116**, 7137–7146.
24. Dassanayake,M., Oh,D.H., Haas,J.S., Hernandez,A., Hong,H., Ali,S., Yun,D.J., Bressan,R.A., Zhu,J.K., Bohnert,H.J. *et al.* (2011) The genome of the extremophile crucifer *Thellungiella parvula*. *Nat. Genet*., **43**, 913–918.
25. Yang,J., Liu,D., Wang,X., Ji,C., Cheng,F., Liu,B., Hu,Z., Chen,S., Pental,D., Ju,Y. *et al.* (2016) The genome sequence of allopolyploid *Brassica juncea* and analysis of differential homoeolog gene expression influencing selection. *Nat. Genet*., **48**, 1225–1232.
26. Perumal,S., Koh,C.S., Jin,L., Buchwaldt,M., Higgins,E., Zheng,C., Sankoff,D., Robinson,S.J., Kagale,S., Navabi,Z.-K. *et al.* (2020) High contiguity long read assembly of *Brassica nigra* allows localization of active centromeres and provides insights into the ancestral *Brassica* genome. bioRxiv doi: https://doi.org/10.1101/2020.02.03.932665, 04 February 2020, preprint: not peer reviewed.
27. Chen,X., Tong,C., Zhang,X., Song,A., Hu,M., Dong,W., Chen,F., Wang,Y., Tu,J., Liu,S. *et al.* (2021) A high-quality *Brassica napus* genome reveals expansion of transposable elements, subgenome evolution and disease resistance. *Plant Biotechnol. J*., **19**, 615–630.
28. Sun,F., Fan,G., Hu,Q., Zhou,Y., Guan,M., Tong,C., Li,J., Du,D., Qi,C., Jiang,L. *et al.* (2017) The high-quality genome of *Brassica napus* cultivar 'ZS11' reveals the introgression history in semi-winter morphotype. *Plant J*., **92**, 452–468.
29. Paritosh,K., Pradhan,A.K. and Pental,D. (2020) A highly contiguous genome assembly of *Brassica nigra* (BB) and revised nomenclature for the pseudochromosomes. *BMCGenomics*, **21**, 887.
30. Cai,X., Wu,J., Liang,J., Lin,R., Zhang,K., Cheng,F. and Wang,X. (2020) Improved *Brassica oleracea* JZS assembly reveals significant changing of LTR-RT dynamics in different morphotypes. *Theor. Appl. Genet*., **133**, 3187–3199.
31. Zhang,L., Cai,X., Wu,J., Liu,M., Grob,S., Cheng,F., Liang,J., Cai,C., Liu,Z., Liu,B. *et al.* (2018) Improved *Brassica rapa* reference genome by single-molecule sequencing and chromosome conformation capture technologies. *Hortic. Res*., **5**, 50.
32. Wu,H.J., Zhang,Z., Wang,J.Y., Oh,D.H., Dassanayake,M., Liu,B., Huang,Q., Sun,H.X., Xia,R., Wu,Y. *et al.* (2012) Insights into salt tolerance from the genome of *Thellungiella salsuginea*. *Proc. Natl. Acad. Sci. U.S.A.,* **109**, 12219–12224.
33. Rousseau-Gueutin,M., Belser,C., Silva,C.D., Richard,G., Istace,B., Cruaud,C., Falentin,C., Boideau,F., Boutte,J., Delourme,R. *et al.* (2020) Long-reads assembly of the *Brassica napus* reference genome, Darmor-bzh. *GigaScience*, **9**, giaa137.
34. 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.
35. Yang,R., Jarvis,D.E., Chen,H., Beilstein,M.A., Grimwood,J., Jenkins,J., Shu,S., Prochnik,S., Xin,M., Ma,C. *et al.* (2013) The reference genome of the halophytic plant *Eutrema salsugineum*. *Front. Plant Sci*., **4**, 46.
36. Song,J., Liu,D., Xie,W., Yang,Z., Guo,L., Liu,K., Yang,Q. and Chen,L. (2021) BnPIR: *Brassica napus* pan-genome information resource for 1689 accessions. *Plant Biotechnol. J*., **19**, 412–414.
37. Yan,T., Wang,Q., Maodzeka,A., Wu,D. and Jiang,L. (2020) BnaSNPDB: an interactive web portal for the efficient retrieval and analysis of SNPs among 1,007 rapeseed accessions. *Comput. Struct. Biotechnol. J*., **18**, 2766–2773.
38. Yu,J., Zhao,M., Wang,X., Tong,C., Huang,S., Tehrim,S., Liu,Y., Hua,W. and Liu,S. (2013) Bolbase: a comprehensive genomics database for *Brassica oleracea*. *BMC Genomics*., **14**, 664.
39. Chao,H., Li,T., Luo,C., Huang,H., Ruan,Y., Li,X., Niu,Y., Fan,Y., Sun,W., Zhang,K. *et al.* (2020) BrassicaEDB: a gene expression database for *Brassica* crops. *Int. J. Mol. Sci*., **21**, 16.
40. Yan,T., Yao,Y., Wu,D. and Jiang,L. (2021) BnaGVD: A genomic variation database of rapeseed (*Brassica napus*). *Plant Cell Physiol*., **62**, 2.
41. Cai,X., Chang,L., Zhang,T., Chen,H., Zhang,L., Lin,R., Liang,J., Wu,J., Freeling,M. and Wang,X. (2021) Impacts of allopolyploidization and structural variation on intraspecific diversification in *Brassica rapa*. *Genome Biol*., **22**, 166.
42. Cheng,F., Sun,R., Hou,X., Zheng,H., Zhang,F., Zhang,Y., Liu,B., Liang,J., Zhuang,M., Liu,Y. *et al.* (2016) Subgenome parallel selection is associated with morphotype diversification and convergent crop domestication in *Brassica rapa* and *Brassica oleracea*. *Nat Genet*, **48**., 1218–1224.
43. 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.
44. Altschul,S.F., Madden,T.L., Schäffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*., **25**, 3389–3402.

45. Cheng,F., Wu,J., Fang,L. and Wang,X. (2012) Syntenic gene analysis between *Brassica rapa* and other Brassicaceae species. *Front. Plant Sci.*, **3**, 198.
46. Cheng,F., Mandakova,T., Wu,J., Xie,Q., Lysak,M.A. and Wang,X. (2013) Deciphering the diploid ancestral genome of the Mesohexaploid *Brassica rapa*. *Plant Cell.*, **25**, 1541–1554.
47. Edgar,R.C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics.*, **5**, 113.
48. Edgar,R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, **32**, 1792–1797.
49. Eguchi,Y. (2011) PHYLIP-GUI-Tool (PHYGUI): adapting the functions of the graphical user interface for the PHYLIP package. *J. Biomed. Sci. Eng.*, **4**, 90–93.
50. Letunic,I. and Bork,P. (2021) Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.*, **49**, W293–W296.
51. Mitchell,A.L., Attwood,T.K., Babbitt,P.C., Blum,M., Bork,P., Bridge,A., Brown,S.D., Chang,H.Y., El-Gebali,S., Fraser,M.I. *et al.* (2019) InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.*, **47**, D351–D360.
52. Jones,P., Binns,D., Chang,H.Y., Fraser,M., Li,W., McAnulla,C., McWilliam,H., Maslen,J., Mitchell,A., Nuka,G. *et al.* (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics.*, **30**, 1236–1240.
53. Mistry,J., Chuguransky,S., Williams,L., Qureshi,M., Salazar,A.G., Sonnhammer,E.L.L., Tosatto,S.E.C., Paladin,L., Raj,S., Richardson,L.J. *et al.* (2021) Pfam: The protein families database in 2021. *Nucleic Acids Res.*, **49**, D412–D419.
54. The Gene Ontology Consortium. (2019) The Gene Ontology Resource: 20 years and still Going strong. *Nucleic Acids Res.*, **47**, D330–D338.
55. Kanehisa,M., Sato,Y., Kawashima,M., Furumichi,M. and Tanabe,M. (2016) KEGG as a reference for gene and protein annotation. *Nucleic Acids Res.*, **44**, D457–D462.
56. Moriya,Y., Itoh,M., Okuda,S., Yoshizawa,A.C. and Kanehisa,M. (2007) KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.*, **35**, W182–W185.
57. Kim,D., Langmead,B. and Salzberg,S.L. (2015) HISAT: a fast spliced aligner with low memory requirements. *Nat Methods.*, **12**, 357–360.
58. Danecek,P., Bonfield,J.K., Liddle,J., Marshall,J., Ohan,V., Pollard,M.O., Whitwham,A., Keane,T., McCarthy,S.A., Davies,R.M. *et al.* (2021) Twelve years of SAMtools and BCFtools. *Gigascience.*, **10**, 2.
59. Liao,Y., Smyth,G.K. and Shi,W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics.*, **30**, 923–930.
60. Cock,P.J., Antao,T., Chang,J.T., Chapman,B.A., Cox,C.J., Dalke,A., Friedberg,I., Hamelryck,T., Kauff,F., Wilczynski,B. *et al.* (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics.*, **25**, 1422–1423.
61. Untergasser,A., Cutcutache,I., Koressaar,T., Ye,J., Faircloth,B.C., Remm,M. and Rozen,S.G. (2012) Primer3–new capabilities and interfaces. *Nucleic Acids Res.*, **40**, e115.
62. Priyam,A., Woodcroft,B.J., Rai,V., Moghul,I., Munagala,A., Ter,F., Chowdhary,H., Pieniak,I., Maynard,L.J., Gibbins,M.A. *et al.* (2019) Sequenceserver: a modern graphical user interface for custom BLAST databases. *Mol Biol Evol.*, **36**, 2922–2924.
63. Skinner,M.E., Uzilov,A.V., Stein,L.D., Mungall,C.J. and Holmes,I.H. (2009) JBrowse: a next-generation genome browser. *Genome Res.*, **19**, 1630–1638.
64. Al-Shehbaz,A.I. (2012) A generic and tribal synopsis of the Brassicaceae (Cruciferae). *TAXON.*, **61**, 931–954.
65. Cheng,F., Wu,J., Cai,X., Liang,J., Freeling,M. and Wang,X. (2018) Gene retention, fractionation and subgenome differences in polyploid plants. *Nat Plants.*, **4**, 258–268.
66. Mandakova,T., Hlouskova,P., German,D.A. and Lysak,M.A. (2017) Monophyletic origin and evolution of the largest crucifer genomes. *Plant Physiol.*, **174**, 2062–2071.
67. Nikolov,L.A., Shushkov,P., Nevado,B., Gan,X., Al-Shehbaz,I.A., Filatov,D., Bailey,C.D. and Tsiantis,M. (2019) Resolving the backbone of the Brassicaceae phylogeny for investigating trait diversity. *New Phytol.*, **222**, 1638–1651.
68. Bailey,C.D., Koch,M.A., Mayer,M., Mummenhoff,K., O'Kane,S.L. Jr, Warwick,S.I., Windham,M.D. and Al-Shehbaz,I.A. (2006) Toward a global phylogeny of the Brassicaceae. *Mol Biol Evol.*, **23**, 2142–2160.
69. Donlin,M.J. (2009) Using the Generic Genome Browser (GBrowse). *Curr. Protoc. Bioinformatics.*, **9**, Unit 9.12.
70. Winter,D., Vinegar,B., Nahal,H., Ammar,R., Wilson,G.V. and Provart,N.J. (2007) An 'Electronic Fluorescent Pictograph' browser for exploring and analyzing large-scale biological data sets. *PLoS One.*, **2**, e718.
71. He,Z., Ji,R., Havlickova,L., Wang,L., Li,Y., Lee,H.T., Song,J., Koh,C., Yang,J., Zhang,M. *et al.* (2021) Genome structural evolution in Brassica crops. *Nat Plants.*, **7**, 757–765.
72. Zhang,X., Liu,T., Wang,J., Wang,P., Qiu,Y., Zhao,W., Pang,S., Li,X., Wang,H., Song,J. *et al.* (2021) Pan-genome of Raphanus highlights genetic variation and introgression among domesticated, wild and weedy radishes. *Mol. Plant.*, **2021**, https://doi.org/10.1016/j.molp.2021.08.005.