PCP
PLANT & CELL PHYSIOLOGY

BnaGVD: A Genomic Variation Database of Rapeseed (*Brassica napus*)

Tao Yan[1], Yao Yao[2], Dezhi Wu [1] and Lixi Jiang [1,*]

[1]Provincial Key Laboratory of Crop Gene Resource, Zhejiang University, 866 Yu-Hang-Tang Road, Hangzhou 310058, China
[2]Biomarker Technologies Corporation, Beijing 101300, China
*Corresponding author: E-mail, jianglx@zju.edu.cn; Fax, +86-571-8898-2905.
(Received 18 June 2020; Accepted 14 December 2020)

Rapeseed (*Brassica napus* L.) is a typical polyploid crop and one of the most important oilseed crops worldwide. With the rapid progress on high-throughput sequencing technologies and the reduction in sequencing cost, large-scale genomic data of a specific crop have become available. However, raw sequence data are mostly deposited in the sequence read archive of the National Center of Biotechnology Information and the European Nucleotide Archive, which is freely accessible to all researchers. Extensive tools for practical purposes should be developed to efficiently utilize these large raw data. Here, we report a web-based rapeseed genomic variation database (BnaGVD, http://rapeseed.biocloud.net/home) from which genomic variations, such as single-nucleotide polymorphisms (SNPs) and insertions/deletions (InDels) across a worldwide collection of rapeseed accessions, can be referred. The current release of the BnaGVD contains 34,591,899 high-quality SNPs and 12,281,923 high-quality InDels and provides search tools to retrieve genomic variations and gene annotations across 1,007 accessions of worldwide rapeseed germplasm. We implement a variety of built-in tools (e.g. BnaGWAS, BnaPCA and BnaStructure) to help users perform in-depth analyses. We recommend this web resource for accelerating studies on the functional genomics and the screening of molecular markers for rapeseed breeding.

Keywords: BnaGVD • Database • Genomic variations • Rapeseed.

Introduction

Rapeseed (*Brassica napus* L.) is an economically important crop grown mainly as a source of edible oil and protein-rich livestock feed. *Brassica napus* is a recent allopolyploid species, formed during the past ~7,500 years ago by an interspecific hybridization between two diploid progenitors, *Brassica rapa* (AA) ($n = 10$) and *Brassica oleracea* (CC) ($n = 9$) (Chalhoub et al. 2014). The emergence of high-throughput sequencing technologies and the availability of *B. napus* reference genome (Chalhoub et al. 2014) have provided new insights into the genomic localization of agronomically important quantitative trait loci, the identification of genes underlying genetic diversity and the determination of the nature of its origins (Lu et al. 2019, Wu et al. 2019).

Over the past decades, large quantities of genotypic data have been exponentially generated through rapid advancements in sequencing technologies and decreasing costs. High-quality reference genomes have also become available, thereby facilitating the improvement of genomic variation data at unprecedented rates and promoting systematic studies on population genetics, including genome-wide association studies (GWASs) (Hayes 2013), evolutionary studies (Wang et al. 2018) and comparative genomics (An et al. 2019).

Single-nucleotide polymorphisms (SNPs) and insertions/deletions (InDels) have been widely used in breeding programs of rapeseed and other crops. Biology researchers especially crop breeders have been benefited from SNPs in identifying functional genes, e,g, the key genes related to the nitrogen use efficiency in rice (Tang et al. 2019), genes for fertility restoration in maize (Jaqueth et al. 2020) and the genes for elevating seed oil content in rapeseed (Karunarathna et al. 2020). Based on the identification of key functional genes, molecular breeding and genetic editing targeting these genes can be further carried out. Approaches, such as GWAS, unveil genetic variability in the forms of SNPs and other types, and these approaches are effective in associating important agronomic and quality traits with their causal loci in various field crops, such as rice (Yano et al. 2016), maize (Wang et al. 2020), soybean (Zhou et al. 2015), rapeseed (Lu et al. 2019, Wu et al. 2019) and cotton (Ma et al. 2018).

Large rapeseed genomic variation datasets have been generated through next-generation sequencing (NGS) technology (Wu et al. 2019). With the availability of large data, a major hurdle has emerged, i.e. how to effectively integrate and share them with the rapeseed research community to speed up rapeseed breeding. The identification of key SNPs and InDels from large NGS datasets is also laborious and requiring extensive computational resources; as such, current SNP and InDel datasets are not user friendly. Several genomic variation databases, including SorGSD for sorghum (Luo et al. 2016), RiceVarMap for rice (Zhao et al. 2015), PeachVar-DB for peach (Cirilli et al. 2018), CitGVD for citrus (Li et al. 2020) and BnaSNPDB for rapeseed (Yan et al. 2020), have been developed. Here, we build BnaGVD, a comprehensive database of rapeseed genomic

Plant Cell Physiol. 0(0): 1–6 doi:10.1093/pcp/pcaa169, Advance Access publication on 5 January 2021, available online at https://academic.oup.com/pcp

variations, to provide a publicly available and free data service for rapeseed research. BnaGVD contains 1,007 rapeseed germplasm accessions from 39 countries in the world and 5,559,254, 5,526,961, and 28,509,014 high-quality SNPs and 1,858,671, 1,919,465, and 8,503,787 high-quality InDels by mapping reads to Darmor-*bzh* (Chalhoub et al. 2014), Tapidor (Bayer et al. 2017) and NY7 (Zou et al. 2019) reference genomes. BnaGVD comprises functional gene annotations and several tools for in-depth analyses, namely, BnaGWAS for performing genome-wide associations mapping based on built-in phenotype data or user-uploaded data, BnaPCA for visualizing genetic diversity based on SNPs in given samples, BnaStructure for visualizing population structure based on SNPs in given samples and PCR primer design and GBrowse for examining genomic variations and genes. These resources and analysis tools will become freely accessible to the rapeseed research community. This web resource will also help researchers focusing on the population genetics, functional genomics and molecular breeding of rapeseed.

Results

Database content and access

All raw paired-end reads from previously published NGS data (Wu et al. 2019) were collected and processed with an in-house pipeline (**Fig. 1**) (see Materials and methods). The BnaGVD database contains four functional sections, namely, 'Browse', 'Search', 'Tools' and 'Download'. It also has a 'Help' section that contains detailed introductory information and user tutorials. These built-in functional modules not only work independently but also cooperate with one another in a single work flow.

Browse

The 'Browse' section in BnaGVD provides data browsing functions, including SNP/InDel browse through which genomic variations can be searched according to reference genomes, chromosomes and germplasms. Gene annotation is used to browse gene information and annotations, and gene family is utilized to browse gene cluster information based on the Pfam data (Finn et al. 2014). Users can browse and filter gene annotations based on three reference genomes, namely, Darmor-*bzh*, Tapidor and NY7, annotation databases or gene IDs (**Fig. 2a**).

The genes of the three reference genomes are annotated on the basis of different databases, such as Nr, gene ontology (GO), Kyoto encyclopedia of genes and genomes (KEGG), Swiss-Prot, InterPro and Pfam (**Fig. 2b**). A new page with gene information, gene structure, genomic sequence, coding sequence (CDS sequence), transcript sequence and peptide sequence is opened by clicking on the hyperlink associated with the gene ID (**Fig. 2c**, **Supplementary Fig. S1**).

Search

The 'Search' section provides a user-friendly web interface to extract SNP/InDel information by specifying the chromosomal start and end regions, gene IDs and SNP/InDel IDs (**Fig. 3a**). Three modules, namely, Multicriteria Search, Comparative Search and Gene Search, are under the 'Search' drop-down menus. The Multicriteria Search provides SNP/InDel information for a single variation (**Fig. 3b**), and Comparative Search generates SNP/InDel information among multiple varieties (**Fig. 3c**). Multicriteria Search displays information on SNP/InDel IDs, chromosomal positions, associated genes and upstream and downstream sequences of SNP/InDel IDs (**Fig. 3b**). Comparative Search directly displays SNP/InDel IDs, chromosomal positions and associated genes of all varieties (**Fig. 3c**). Both of them allow users to jump to the GBrowse interface to browse the detailed information of variations through a given hyperlink (**Supplementary Fig. S2a**). Gene Search module allows users to search for gene information based on different reference genomes, chromosomal positions and gene or protein IDs (**Supplementary Fig. S2b**).

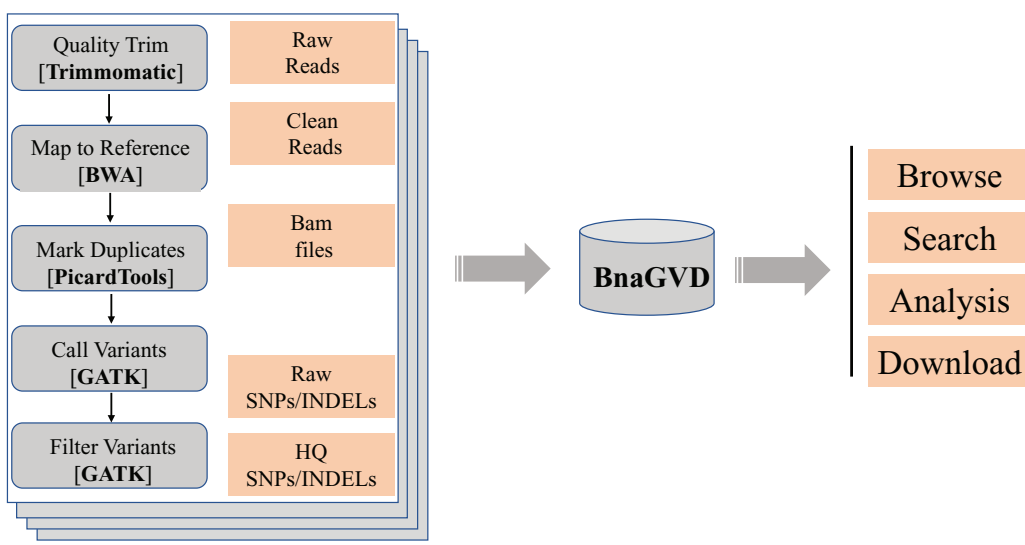

Fig. 1 Pipeline used in the processing of rapeseed NGS data to construct BnaGVD.

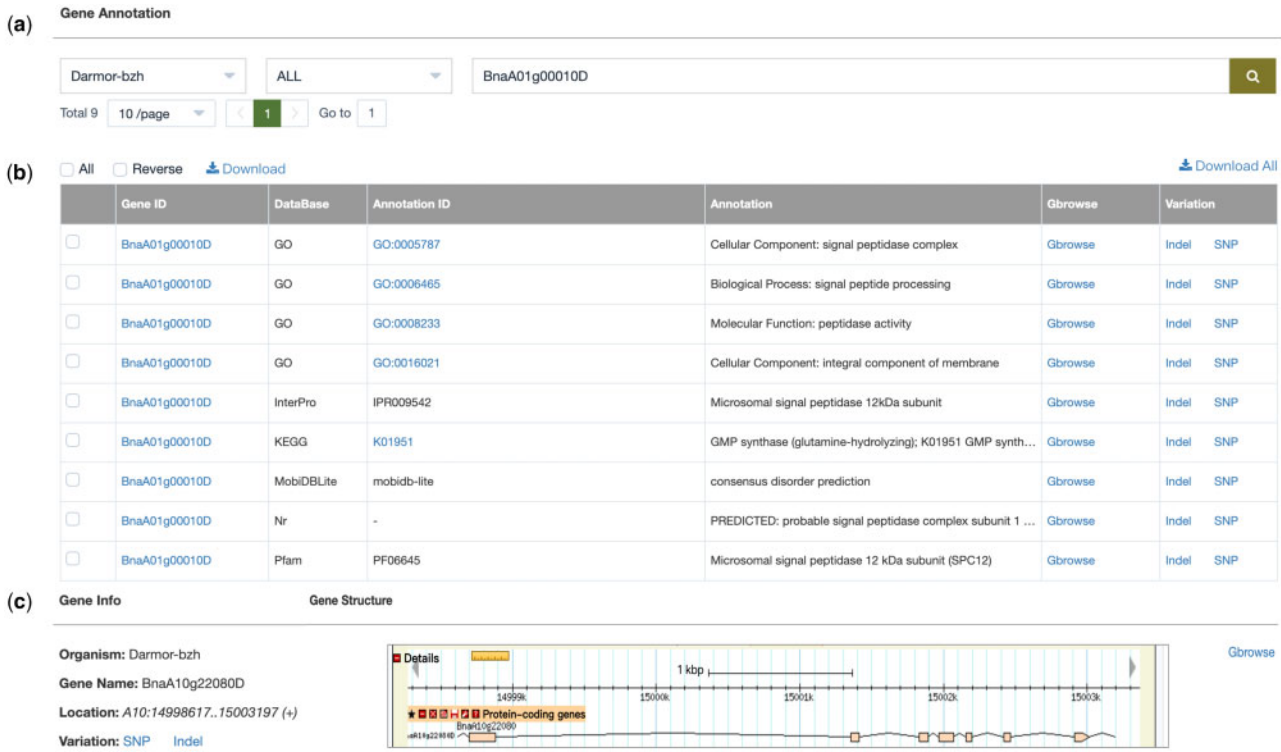

Fig. 2 Interface to search and filter gene annotations. (a) Gene annotation filtering options by reference genome, annotation database or gene ID. (b) Gene annotations retrieved according to different databases. (c) Interface showing the gene structure.

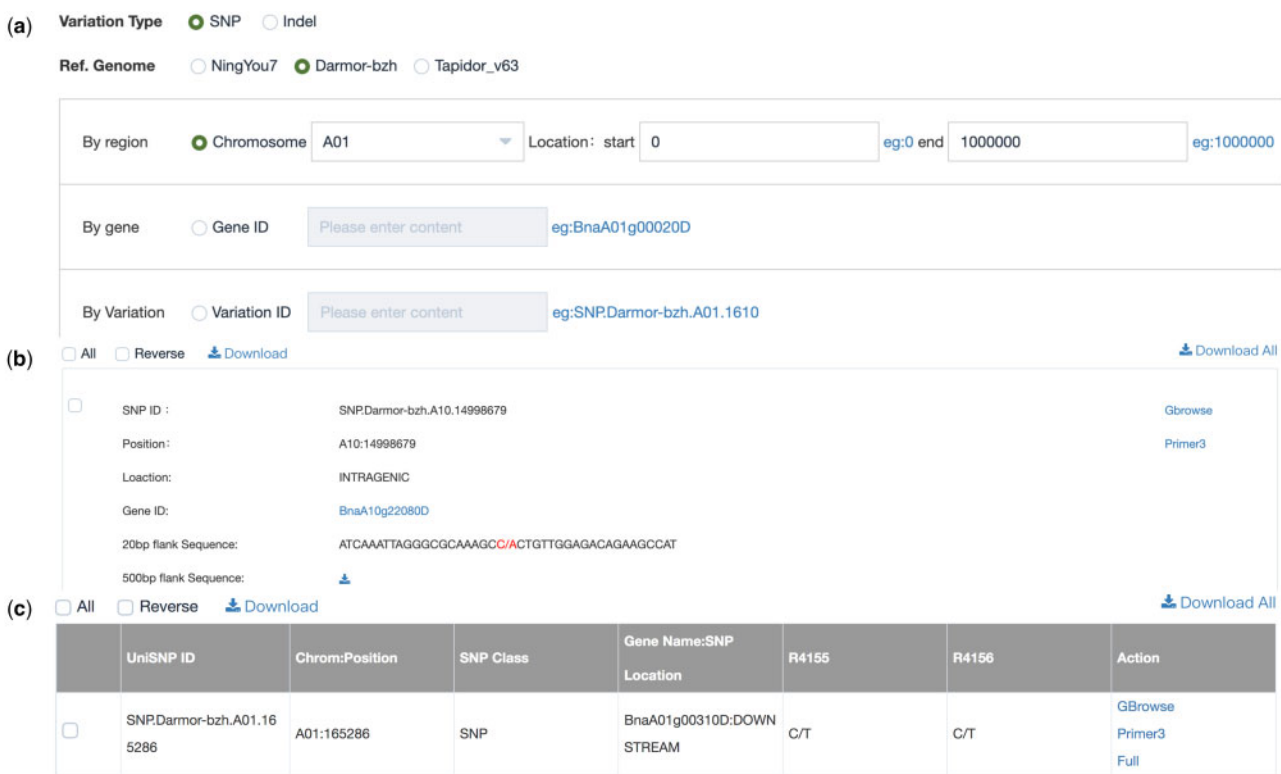

Fig. 3 Search functions implemented in BnaGVD. (a) Genomic variations associated with different chromosomal locations or genes. (b) Typical search result via Multicriteria Search. (c) Comparative search, which provides information about SNP/InDels among multiple varieties.

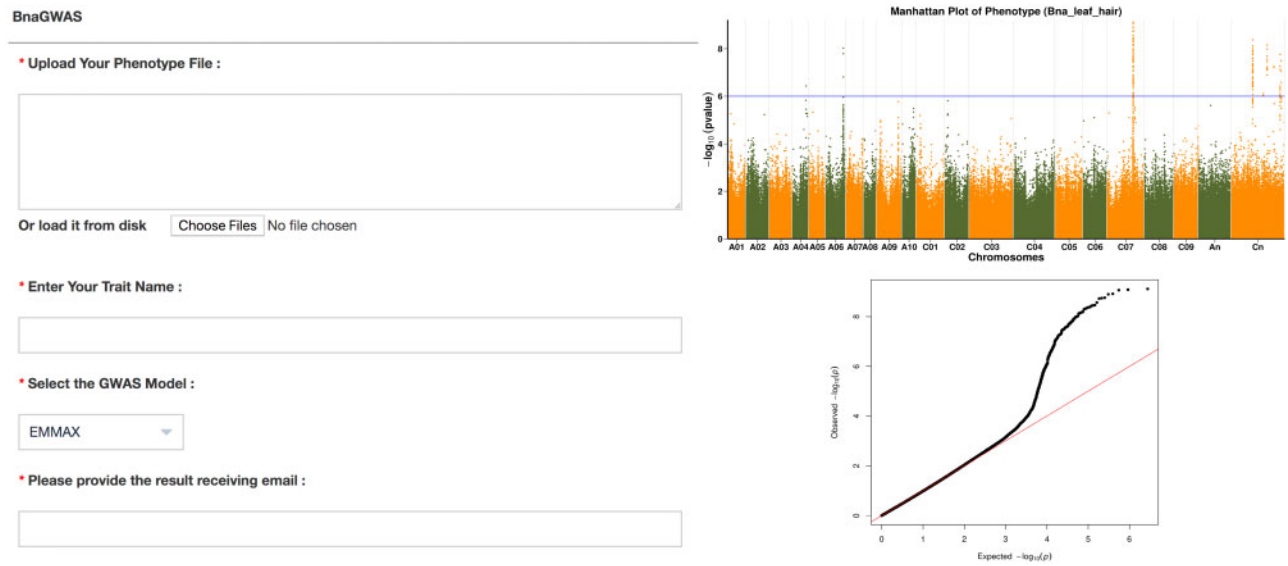

Fig. 4 BnaGWAS tool, which performs association mapping and result visualization in the form of Manhattan and QQ plots.

Tools

BnaGVD is designed for storing, displaying, and extracting information on genomic variation. A core germplasm collection (Xuan et al. 2020) with 300 accessions is established on the basis of whole rapeseed germplasms to use large genetic resources efficiently, and this collection is shared with several research units involved in rapeseed breeding in China. In a previous publication, we have showed that the core germplasm collection can represent the majority of the genetic variations in whole germplasm accessions (Xuan et al. 2020). Three tools, namely, BnaGWAS, BnaPCA and BnaStructure, are developed and integrated within BnaGVD to maximize the use of genetic information. BnaGWAS can be applied to perform GWAS and visualize the results in the form of Manhattan and quantile–quantile (QQ) plots (**Fig. 4**, **Supplementary material**). It is also built on the basis of the EMMAX model (Kang et al. 2010). Significant SNPs, including annotations of associated genes, can be extracted from association mapping results based on different databases (e.g. Nr, SWISS-PROT, GO, KEGG and eggNOG), which can be conducted in a single work flow. BnaPCA can be utilized to visualize genetic diversity based on SNPs in given samples provided by users (**Supplementary Fig. S3**). BnaStructure is designed to show the population structure based on SNPs in given samples (**Supplementary Fig. S4**). The results from the three tools are sent to users via e-mail. In addition to these three tools, GBrowse, Primer Design and BLAST are established to search for orthologs or paralogs of a sequence input by users. The BLAST platform integrates multiple programs, such as blastn, blastp, blastx, tblastn and tblastx, to retrieve hits of CDS and peptide sequences from three reference genomes. These tools will be helpful for mining genetic variations and genes related to important agronomic traits of rapeseed. Thus, it will facilitate the screening of molecular markers for accelerated rapeseed breeding.

Download

All built-in data, including genotypic data, NGS sequencing data, reference data, and phenotypic data, are provided in the BnaGVD platform for users to download. All the raw sequences of the rapeseed accessions used in this study have been deposited in the public databases of the National Center of Biotechnology Information (NCBI) under SRP155312 (**https://www.ncbi.nlm.nih.gov/sra/SRP155312**) and the National Genomics Database Center under CRA001854 (**https://bigd.big.ac.cn/gsa/browse/CRA001854**). The results from 'Browse' and /Search' tools can be downloaded from the result pages by clicking on the link named 'Download' (**Figs. 2b, 3b**, c).

Help

A 'Help' section, including FAQs, is provided to assist the users of BnaGVD. An analysis pipeline step-by-step user guide is also available to ensure that first-time users can easily access BnaGVD.

Discussion and Conclusion

Biological research has largely benefited from a variety of databases with different applications. In a previous database (BnSNPDB), we provided tools to draw linkage disequilibrium heat map (LDheatmap), SNP distribution map and phylogenetic tree by retrieving SNPs of 1,007 rapeseed accessions. In this article, we developed new online tools for GWAS, principal component analysis (PCA) and population structure analysis in the same rapeseed population. In addition to SNPs, we provided method to retrieve InDels for genetic polymorphism analysis. This database can serve as an important online tool for genomic variation analyses complementary to what we previously provided (Yan et al. 2020). The rapeseed genomic variations stored in the BnaGVD are from the largest worldwide collection of rapeseed accessions and are the most

comprehensive. The database integrated 1,007 worldwide rapeseed germplasm accessions from 39 countries and 34,591,899 high-quality SNPs and 12,281,923 high-quality InDels. BnaGVD contributes to future rapeseed studies, including population genomic, mining of key genes related to important agronomic traits and whole-genome molecular marker-assisted breeding.

Previously published databases, such as SorGSD for sorghum (Luo et al. 2016), mainly focused on storing genomic variations (SNPs/InDels). Several *Brassica* crops related database, such as CropSNPdb (Scheben et al. 2019), have been established to store and retrieve SNPs. The CropSNPdb is a database of SNP array data for Brassica crops and hexaploid bread wheat, which is essential for the molecular breeding of rapeseed. However, the CropSNPdb just provides a simple query interface and the SNP array data were produced by *Brassica* 60K arrays and not by the NGS technology and just contains 459 rapeseed accessions. By comparison, the BnaGVD was designed with built-in tools and pipelines, such as BnaGWAS for GWAS, BnaPCA for genetic diversity visualization based on SNPs in given samples provided by users and BnaStructure for population structure visualization. As such, BnaGVD could be utilized to extract and browse information about genomic variation and perform GWAS, gene annotation and genomic variation visualization. We also constructed a core germplasm collection containing 300 accessions, which represented the genetic diversity of 1,007 rapeseed genome, for follow-up rapeseed research (Xuan et al. 2020).

We also cooperated with the majority of the research community to share the core germplasm collection to greatly improve the mining of key genes related to important agronomic traits and facilitate the screening of molecular markers. Thus, we could accelerate rapeseed breeding.

With the rapidly decreasing sequencing cost, more genomes will be sequenced and assembled, and genomic variations are no longer limited to SNPs/InDels. In the next version, a large number of rapeseed individuals, reference genomes, phenotypes and genetic structure variations, including copy number variations, structural variations and transposons, will be integrated in BnaGVD.

In summary, we constructed a rapeseed genomic variation database (BnaGVD) as a platform for bioinformatic analysis, including examination of genomic variation, gene annotation and visualization of SNPs and GWAS results. We also cooperated with domestic and international research communities to share the core germplasm collection and promote the screening of molecular markers for the genetic improvement of rapeseed.

Materials and Methods

Data sources and processing

All raw paired-end reads from previously published NGS data (Wu et al. 2019) were collected and processed with an in-house pipeline (**Fig. 1**). For data processing, all raw paired-end reads were trimmed with Trimmomatic (v0.3.0, parameters: LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:50) (Bolger et al. 2014) for quality control and then mapped to the reference *B. napus* genome by using the MEM algorithm of the Burrows–Wheeler Aligner program (Li and Durbin 2010). Mapping results were sorted, and duplicate reads were marked using SAMtools (Li et al. 2009) and Picard (**https://broadinstitute.github.io/picard/**; v1.9.4) with default parameters. SNPs and InDels were identified and filtered with GATK (the detailed pipeline is available at **https://github.com/YTLogos/myscript/blob/master/Variants_calling/variants_calling_pipeline.sh**) (McKenna et al. 2010). SNPs were annotated by utilizing the SnpEFF software (Cingolani et al. 2012). The genes from the reference genomes Darmor-*bzh* (Chalhoub et al. 2014), Tapidor (Bayer et al. 2017) and NY7 (Zou et al. 2019) were annotated on the basis of NCBI's nonredundant protein database (Nr) (Pruitt et al. 2004), GO (The Gene Ontology Consortium, 2017), KEGG (Aoki-Kinoshita and Kanehisa 2007), SWISS-PROT protein sequence database (Bairoch and Apweiler 2000), InterPro database (Hunter et al. 2009), Pfam database (Finn et al. 2014) and eggNOG database (Huerta-Cepas et al. 2019) and then integrated into BnaGVD. With 34,591,899 SNPs, 12,281,923 InDels and their corresponding gene annotations, a web interface was designed to search for, browse, analyze, and download data.

Database construction

The BnaGVD database was built on the basis of a J2EE framework with MySQL (**http://www.mysql.org/**; a free and popular relational database management system) as the database engine. Web user interfaces were developed with JSP (JavaServer Pages; a technology facilitating rapid development of dynamic web pages based on the Java programming language), HTML5 and CSS3. BnaGVD-integrated tools, such as BnaGWAS, BnaPCA and BnaStructure, were compiled with R (v3.6.0) and Perl (v3.5.6) and operated in a Linux environment. GBrowse (Stein et al. 2002) (**http://gbrowser.sourceforge.net**) was adopted for chromosome-based genomic variation and gene visualization. Primer Design (v4.1.0) was implemented to retrieve primers for *B. napus* SNPs, InDels and genes. BLAST (v2.2.31) was used to search for orthologs or paralogs of input sequences in BnaGVD.

Supplementary Data

Supplementary data are available at PCP online.

Funding

Zhejiang Provincial Key Research Project (2021C02057) and Jiangsu Collaborative Innovation Centre for Modern Crop Production.

Disclosures

The authors have no conflicts of interest to declare.

References

An, H., Qi, X., Gaynor, M.L., Hao, Y., Gebken, S.C., Mabry, M.E., et al. (2019) Transcriptome and organellar sequencing highlights the complex origin and diversification of allotetraploid Brassica napus. *Nat. Commun.* 10: 2878.

Aoki-Kinoshita, K.F. and Kanehisa, M. (2007) Gene annotation and pathway mapping in KEGG. *Methods Mol. Biol.* 396: 71–91.

Bairoch, A. and Apweiler, R. (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* 28: 45–48.

Bayer, P.E., Hurgobin, B., Golicz, A.A., Chan, C.-K.K., Yuan, Y., Lee, H., et al. (2017) Assembly and comparison of two closely related Brassica napus genomes. *Plant Biotechnol. J.* 15: 1602–1610.

Bolger, A.M., Lohse, M. and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120.

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I.A.P., Tang, H., Wang, X., et al. (2014) Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. *Science* 345: 950–953.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., et al. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, *SnpEff. Fly* 6: 80–92.

Cirilli, M., Flati, T., Gioiosa, S., Tagliaferri, I., Ciacciulli, A., Gao, Z., et al. (2018) PeachVar-DB: a curated collection of genetic variations for the interactive analysis of peach genome data. *Plant Cell Physiol*. 59: e2–e2.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., et al. (2014) Pfam: the protein families database. *Nucl. Acids Res.* 42: D222–D230.

Hayes, B. (2013) Overview of statistical methods for genome-wide association studies (GWAS). *Methods Mol. Biol.* 1019: 149–169.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., et al. (2019) eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res*. 47: D309–D314.

Hunter, S., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., et al. (2009) InterPro: the integrative protein signature database. *Nucleic Acids Res*. 37: D211–D215.

Jaqueth, J.S., Hou, Z., Zheng, P., Ren, R., Nagel, B.A., Cutter, G., et al. (2020) Fertility restoration of maize CMS-C altered by a single amino acid substitution within the Rf4 bHLH transcription factor. *Plant J.* 101: 101–111.

Kang, H.M., Sul, J.H., Service, S.K., Zaitlen, N.A., Kong, S.-Y., Freimer, N.B., et al. (2010) Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* 42: 348–354.

Karunarathna, N.L., Wang, H., Harloff, H.-J., Jiang, L. and Jung, C. (2020) Elevating seed oil content in a polyploid crop by induced mutations in SEED FATTY ACID REDUCER genes. *Plant Biotechnol. J.* 18: 2251–2266.

Li, H. and Durbin, R. (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26: 589–595.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al.; 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25: 2078–2079.

Li, Q., Qi, J., Qin, X., Dou, W., Lei, T., Hu, A., et al. (2020) CitGVD: a comprehensive database of citrus genomic variations. *Hortic. Res.* 7: 12.

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

Luo, H., Zhao, W., Wang, Y., Xia, Y., Wu, X., Zhang, L., et al. (2016) SorGSD: a sorghum genome SNP database. *Biotechnol. Biofuels* 9: 6.

Ma, Z., He, S., Wang, X., Sun, J., Zhang, Y., Zhang, G., et al. (2018) Resequencing a core collection of upland cotton identifies genomic variation and loci influencing fiber quality and yield. *Nat. Genet.* 50: 803–813.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20: 1297–1303.

Pruitt, K.D., Tatusova, T. and Maglott, D.R. (2004) NCBI reference sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res*. 33: D501–504.

Scheben, A., Verpaalen, B., Lawley, C.T., Chan, C.-K.K., Bayer, P.E., Batley, J., et al. (2019) CropSNPdb: a database of SNP array data for Brassica crops and hexaploid bread wheat. *Plant J.* 98: 142–152.

Stein, L.D., Mungall, C., Shu, S., Caudy, M., Mangone, M., Day, A., et al. (2002) The generic genome browser: a building block for a model organism system database. *Genome Res*. 12: 1599–1610.

Tang, W., Ye, J., Yao, X., Zhao, P., Xuan, W., Tian, Y., et al. (2019) Genome-wide associated study identifies NAC42-activated nitrate transporter conferring high nitrogen use efficiency in rice. *Nat. Commun.* 10: 5279.

Wang, B., Lin, Z., Li, X., Zhao, Y., Zhao, B., Wu, G., et al. (2020) Genome-wide selection and genetic improvement during modern maize breeding. *Nat. Genet.* 52: 565–567.

Wang, L., He, F., Huang, Y., He, J., Yang, S., Zeng, J., et al. (2018) Genome of wild mandarin and domestication history of mandarin. *Mol. Plant* 11: 1024–1037.

Wu, D., Liang, Z., Yan, T., Xu, Y., Xuan, L., Tang, J., 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.

Xuan, L., Yan, T., Lu, L., Zhao, X., Wu, D., Hua, S., et al. (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.

The Gene Ontology Consortium. (2017) Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res*. 45: D331–D338.

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.

Yano, K., Yamamoto, E., Aya, K., Takeuchi, H., Lo, P., Hu, L., et al. (2016) Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. *Nat. Genet.* 48: 927–934.

Zhao, H., Yao, W., Ouyang, Y., Yang, W., Wang, G., Lian, X., et al. (2015) RiceVarMap: a comprehensive database of rice genomic variations. *Nucleic Acids Res*. 43: D1018–D1022.

Zhou, Z., Jiang, Y., Wang, Z., Gou, Z., Lyu, J., Li, W., et al. (2015) Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat. Biotechnol.* 33: 408–414.

Zou, J., Mao, L., Qiu, J., Wang, M., Jia, L., Wu, D., et al. (2019) Genome-wide selection footprints and deleterious variations in young Asian allotetraploid rapeseed. *Plant Biotechnol. J.* 17: 1998–2010.