Plant Communications
Correspondence

CellPress
Partner Journal

BnaOmics: A comprehensive platform combining pan-genome and multi-omics data from *Brassica napus*

Dear Editor,

Brassica napus was originally formed ~7500 years ago by the interspecific hybridization of *B. rapa* and *B. oleracea* (Chalhoub et al., 2014). It accounts for approximately 13%–16% of global vegetable oil production and provides an excellent model for polyploid genomics and evolutionary research in plants. Currently, the Brassicaceae Database (BRAD V3.0; Chen et al., 2022), Genoscope (https://www.genoscope.cns.fr/brassicanapus/), and Ensembl Plants (https://plants.ensembl.org) are used for genomic research pertaining to *B. napus*. However, they only provide genome browsers for research into the *B. napus* cultivar Darmor-*bzh*. The *B. napus* pan-genome information resource (BnPIR) also provides eight *B. napus* reference genomes (Song et al., 2021), gene information, and resequencing data. The *B. napus* variation information resource (BnVIR) provides information on genetic variation, including single-nucleotide polymorphisms (SNPs), insertions or deletions (INDELs), and structural variations (SVs) (Yang et al., 2022).

In recent years, many whole-genome assemblies, re-sequencing data for thousands of germplasm accessions, and other omics data (such as transcriptomes and epigenomes) have been made available for *B. napus* (supplemental note). However, despite the abundance of omics data, there is no user-friendly platform that enables genomics researchers and breeders to search, visualize, and analyze omics information. Therefore, we aimed to construct a pan-genome of *B. napus* that integrates information on genetic diversity of large populations and enables visualization and analysis of multi-omics data in a pan-genomic context. To achieve this, we integrated several reference genomes and multi-omics data from a significant number of *B. napus* accessions, resulting in a comprehensive *B. napus* pan-genome (supplemental note) that is 1.99 Gb in size and contains 159 099 high-confidence genes (supplemental note). Based on the data described above, we developed a comprehensive database called BnaOmics (https://bnaomics.ocri-genomics.net/) that includes genome sequences, gene annotations, gene expression profiles in multiple tissues and populations, synteny blocks, epigenomes, and common bioinformatics tools for analysis and visualization (Figure 1). A comparison between BnaOmics and other published *B. napus* databases, including the *B. napus* multi-omics information resource (BnIR; Yang et al., 2023), reveals that BnaOmics integrates 26 *B. napus* reference genomes, re-sequencing data from 2885 accessions (2504 low depth and 381 high depth), 8 736 523 high-quality SVs, 112 886 high-quality presence/absence variants (PAVs), thousands of transcriptomes, and hundreds of epigenomes (supplemental note and supplemental Tables 2 and 3). In particular, multi-omics information is displayed in a pan-genome context, which allows users to search or analyze data based on gene IDs from different *B. napus* reference genomes.

The BnaOmics pan-genome contains more complete and unique genomic sequences than the HAZU pan-genome (Song et al., 2021) because it integrates a larger number of sequences and variations from more extensive germplasm resources (supplemental note and supplemental Figure 1). The data resources and tools provided by BnaOmics will benefit research into the biology and breeding of *B. napus*, as demonstrated by examples of its use (supplemental note).

AN OVERVIEW OF BnaOmics

By integrating different multi-omics data (genome, variome, transcriptome, and epigenome), we developed three main modules for BnaOmics: Reference Genomes, Data, and Searches. In brief, the Reference Genomes module provides basic information for the linear pan-genome and 26 *de novo* assembled genomes of *B. napus* (supplemental Table 1). The main web page for *B. napus* germplasm contains various sections of information and menus for accessing both the genome data and bioinformatics analysis tools (supplemental Figure 2), including an overview, descriptions, whole genomes, links, downloads, genes, germplasm, sequences, BLAST, JBrowse, a sequence retrieval tool, and a synteny viewer. The overview and description section provides basic information on *B. napus* germplasm, including names, chromosome numbers, genome assemblies, pictures, and germplasm descriptions. The genes, germplasm, sequences, BLAST, JBrowse, sequence retrieval, and synteny viewer menus provide links to the related web page tools. The Data module (Figure 1A and supplemental Figure 3) includes 26 reference genomes and a linear pan-genome of *B. napus*, variations (SNPs and SVs) from 2885 accessions (supplemental note), related information on the gene expression of 1234 *B. napus* samples, and 166 samples of DNA methylation, chromatin interaction, and histone modification. The Search module (supplemental Figure 4) provides a means of querying gene functional annotation, germplasm, traits, synteny blocks, SNPs, and SVs. All genes and transcripts from the 26 reference and pan-genomes have been integrated into BnaOmics. For each gene or mRNA, there is a detailed information page including the ID, location, gene structure, sequence, gene expression, and functional annotation sections (supplemental Figure 4B). The search function for DNA methylation, histone modification,

Published by the Plant Communications Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and CEMPS, CAS.

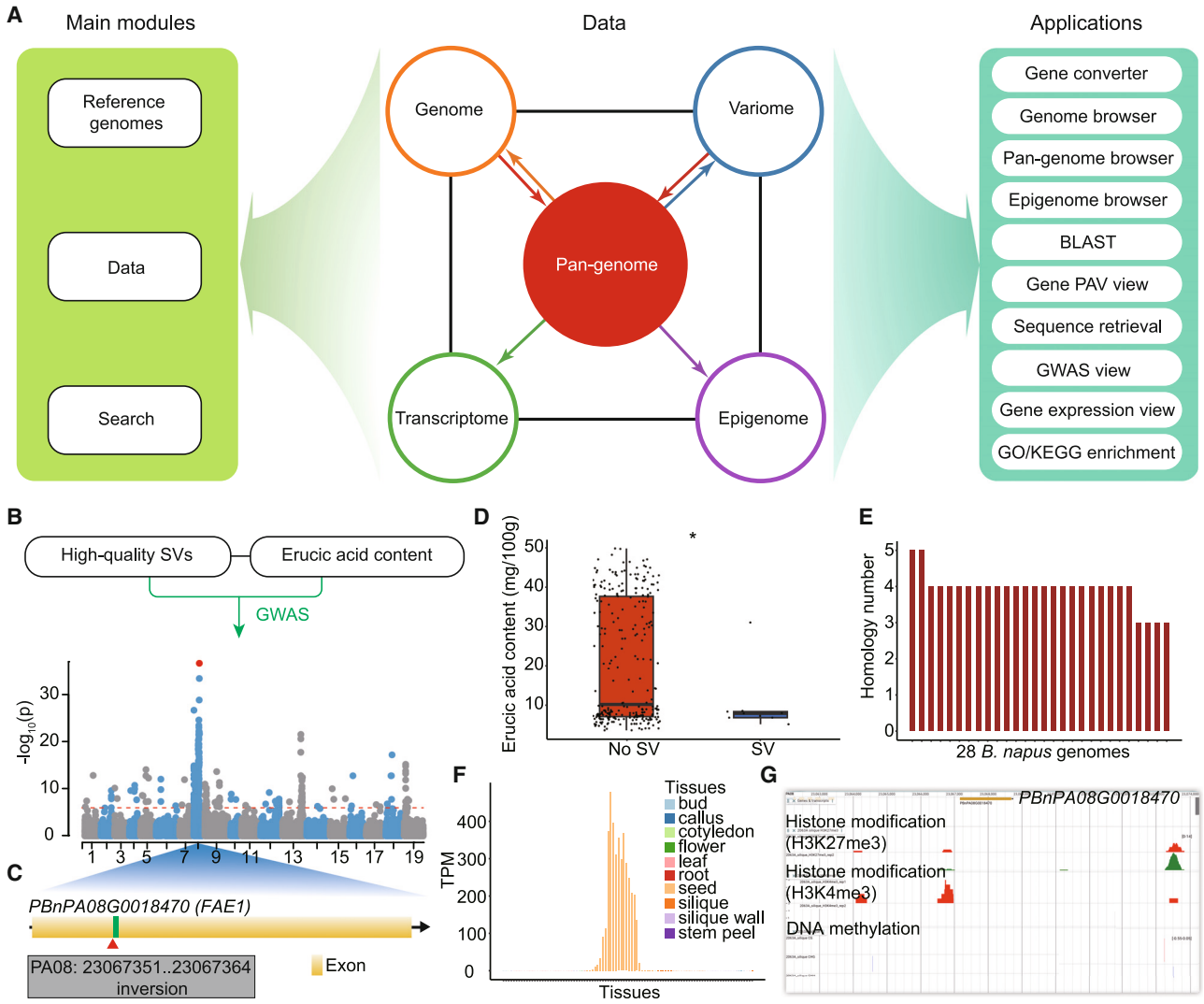

Figure 1. Overview of BnaOmics and its use in data mining

(A) Framework of BnaOmics including data, main modules, and applications. In the *Data* section, the pan-genome was constructed based on multiple *Brassica napus* reference genomes and the variome. Information pertaining to the genome, variome, transcriptome, and epigenome was visualized and analyzed in a pan-genome context.

(B) SV-GWAS analysis of erucic acid content. The probability threshold was set to $P = 1e{-}6$.

(C) Gene structure and variation information pertaining to *PBnPA08G0018470* (*FAE1*).

(D) Comparison of erucic acid content between two groups with different SV genotypes, which was conducted using Welch's *t*-test with a significance level of 0.01.

(E) The number of genes homologous to *PBnPA08G0018470* (*FAE1*) in 28 genomes, including the BnaOmics pan-genome, Westar, Darmor_v10, Darmor_v5, Express617, FAFU_ZS11, Gangan, No2127, Quinta, Shengli, HZAU_Tapidor, HZAU_Westar, Zheyou7, HZAU_ZS11, HTR001, Kale, Laurentian_ONT, Laurentian, Mendel, Ningyou7, Tapidor_v63, Westar_ONT, ZS11_ONT, HZAU pan-genome, ZS11_2016, ZS4-2, ZS9, and ZY821.

(F) Expression information pertaining to *PBnPA08G0018470* (*FAE1*) in 10 tissues.

(G) Histone modification (H3K27me3 and H3K4me3) and DNA methylation (CG, CHG, and CHH) information from the gene body and upstream and downstream of *PBnPA08G0018470* (*FAE1*).

and chromatin interaction provides information about CG, CHG, and CHH methylation levels, histone modification regions, chromatin interaction frequency, the A/B compartment (open/closed chromatin compartment), and the topologically associating domain region (supplemental Figure 4I–4K). Furthermore, information about the germplasm, traits, synteny blocks, and SNPs/SVs can be accessed conveniently (supplemental Figure 4C–4H). For easy use of the search function, gene IDs from different reference genomes are allowed as inputs, and a table of one-to-one correspondences between gene names from different reference genomes is provided (supplemental Figure 4G).

APPLICATION TOOLKITS

In addition to the three main modules, we designed several common, easy-to-use toolkits, including *Gene converter*, *Genome browser*, *Pan-genome browser*, *Epigenome browser*, *BLAST*,

Gene PAV view, *Sequence retrieval*, *GWAS view*, *Gene expression view*, and *GO/KEGG enrichment*. *Gene converter* is used to easily search for homologous genes among the different genome assemblies of *B. napus* (supplemental Figure 5K). The *Genome browser* and *Pan-genome browser* are used to view gene models, SNPs, INDELS, PAVs, and SVs of 26 reference genomes and the linear pan-genome (supplemental Figure 5F and 5G). The *Epigenome browser* is used to view chromatin interactions, DNA methylation, and histone modification information pertaining to the gene body and intergenic regions (supplemental Figure 5H). *BLAST* searches can be performed against 26 *B. napus* reference genomes and the linear pan-genome, CDS sequences, and protein sequences to identify homologous sequences (supplemental Figure 5B–5E). The *Gene PAV view* was designed to view gene PAV information among 381 high-depth resequencing accessions (supplemental Figure 5P). *Sequence retrieval* can be performed to obtain sequences for specific genomic regions, genes, mRNAs, and/or proteins from 26 *B. napus* genomes and the linear pan-genome (supplemental Figure 5L). *GWAS view* enables users to view SNP, PAV, and SV-GWAS results for 22 agronomic traits (supplemental Figure 5M). *Gene expression view* provides gene expression information for 10 different tissues at the population level (supplemental Figure 5N). *GO/KEGG enrichment* was developed to perform GO or KEGG analysis on a gene list from 26 *B. napus* genomes and the linear pan-genome (supplemental Figure 5Q).

DATA MINING USING BnaOmics

Because BnaOmics integrates a wide variety of *B. napus* multi-omics data, it is useful for deep-mining data for fundamental research and molecular breeding. For example, a candidate gene that regulates erucic acid content, *PBnPA08G0018470* (*FAE1*), was identified based on SV-GWAS (Figure 1B and 1C). In BnaOmics, the significant signals around this gene were displayed with the *GWAS view* tool, and a comparison of phenotypes between the two SV genotypes is clearly shown (Figure 1D). By searching for information related to gene *PBnPA08G0018470* using various modules and tools, users can obtain functional annotation and homology data from 28 *B. napus* genomes, expression patterns from 10 tissues, and epigenome information (Figure 1D–1G).

In summary, BnaOmics provides a central, comprehensive platform for genomics, transcriptomics, epigenomics, and variations in *B. napus*. It contains genome sequences, mRNAs, proteins, and comprehensive gene functional annotations, as well as genome synteny blocks, germplasm, traits, gene expression profiles, gene PAVs, and epigenomics of *B. napus*. BnaOmics will be continuously updated with new genome, RNA sequencing, epigenome, and genetic datasets relating to *B. napus*. We will also develop novel data mining and analysis tools that are user friendly. BnaOmics is a comprehensive resource for basic research and breeding in *B. napus*. It enables researchers to search for and visualize results in a pan-genome context and provides a good example for research into other species.

FUNDING

This research was funded by the Young Top-notch Talent Cultivation Program of Hubei Province for C.T., the National Natural Science Foundation of China (U20A2034, 32170366, 32070217, and 31770250), the National Key Research and Development Program of China (2021YFD1600500), the Central Public-interest Scientific Institution Basal Research Fund (2021-2060302-061-027 and 2021-2060302-061-029), the Agricultural Science and Technology Innovation Program of the Chinese Academy of Agricultural Sciences (CAAS-ASTIP-2013-OCRI), and the China Agricultural Research System of MOF and MARA (CARS-12).

DATA AVAILABILITY

Sources of all datasets are described in the supplemental information. All datasets are available at https://bnaomics.ocri-genomics.net/download.

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Plant Communications Online*.

AUTHOR CONTRIBUTIONS

C.T. and S.L. conceived the project. M.H. collected publicly available reference genomes of *B. napus*. S.Y., Y.Z., M.T., L.L., and X.-H.C. collected publicly available re-sequencing and RNA sequencing data. X.-B.C. carried out the analysis, constructed the database, and wrote the manuscript. C.T. supervised the manuscript.

ACKNOWLEDGMENTS

We thank Isobel Parkin and Gary Peng from Saskatoon Research Centre of Agriculture and Agri-Food Canada for meaningful discussion and constructive comments. The numerical calculations in this study have been performed on the supercomputing system in the Supercomputing Center of Oil Crops Research Institute, Chinese Academy of Agricultural Sciences. No conflict of interest declared.

Received: June 21, 2022
Revised: April 11, 2023
Accepted: April 23, 2023
Published: April 25, 2023

Xiaobo Cui[1], Ming Hu[1], Shengli Yao[2], Yuanyuan Zhang[1], Minqiang Tang[3], Lijiang Liu[1], Xiaohui Cheng[1], Chaobo Tong[1,] and Shengyi Liu[1,*]*

[1]Key Laboratory of Biology and Genetic Improvement of Oil Crops, The Ministry of Agriculture and Rural Affairs, Oil Crops Research Institute, Chinese Academy of Agricultural Sciences, Wuhan 430062, China
[2]School of Life Science and Technology, Wuhan Polytechnic University, Wuhan 430023, China
[3]Key Laboratory of Genetics and Germplasm Innovation of Tropical Special Forest Trees and Ornamental Plants (Ministry of Education), College of Forestry, Hainan University, Haikou 570228, China
***Correspondence: Chaobo Tong (tongchaobo@126.com), Shengyi Liu (liusy@oilcrops.cn)**
https://doi.org/10.1016/j.xplc.2023.100609

REFERENCES

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

Chen, H., Wang, T., He, X., Cai, X., Lin, R., Liang, J., Wu, J., King, G., and Wang, X. (2022). Brad V3.0: an upgraded Brassicaceae database. Nucleic Acids Res. **50**:D1432–D1441.

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

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

Yang, Z., Wang, S., Wei, L., Huang, Y., Liu, D., Jia, Y., Luo, C., Lin, Y., Liang, C., Hu, Y., et al. (2023). BnIR: a multi-omics database with various tools for *Brassica napus* research and breeding. Mol. Plant **16**:775–789.