Plant Communications
Correspondence

CellPress
Partner Journal

MRBIGR: A versatile toolbox for genetic regulation inference from population-scale multi-omics data

Dear Editor,

Genome-wide association studies (GWAS) are pivotal genetic methodologies for identifying genomic regions associated with specific traits in crops. However, they often face challenges with complex traits that involve multiple quantitative trait loci and environmental factors (Liu and Yan, 2019; Tam et al., 2019). Despite recent advancements in GWAS tools such as GAPIT3 (Tang et al., 2016) and HAPPI GWAS (Slaten et al., 2020), which have facilitated the exploration of relationships between molecular and phenotypic traits, these tools still adhere to conventional GWAS designs and exhibit limited capabilities in establishing associations across multi-omics layers.

Mendelian randomization (MR) analysis, initially developed for causal inference in epidemiology, leverages genetic variation as instrumental variables to estimate the causal impact of environmental exposures on phenotypic outcomes (Akiyama, 2021). MR has been successfully applied in both human and plant sciences to unravel complex regulatory networks (Zhu et al., 2016; Liu et al., 2020), overcoming some limitations of traditional GWAS methods. Integrating MR with multi-omics association analysis could provide new insights into identifying key genes affecting traits and dissecting complex genetic regulatory networks in crops. However, existing tools lack a unified framework that combines multi-omics association analysis with MR-based causal inference, which limits the comprehensive exploration of complex datasets.

To address this gap, we introduce MRBIGR (MR-based Inference of Genetic Regulation), a versatile toolbox designed for genetic regulation inference in population-scale multi-omics data. MRBIGR features a user-friendly graphical user interface (GUI) that integrates key functionalities throughout the entire GWAS analysis and MR strategy (Figure 1; Supplemental Table 1; supplemental notes). This package provides a comprehensive solution from raw data processing to causal interpretation and is freely accessible at https://gitee.com/crazyhsu/MRBIGR.

To demonstrate the application of MRBIGR in population-scale multi-omics data analysis, we utilized a maize dataset from the Maizego database. This dataset includes genotypic data for 527 maize inbred lines and transcriptomic and metabolomic data from kernels of 368 inbred lines sampled 15 days after pollination, alongside phenotypic data for nine kernel-related agronomic traits (Supplemental Table 2; supplemental materials).

In processing the genotypic data, we employed the '*geno*' module of MRBIGR to filter out SNPs with a missing rate greater than 10% and a minor-allele frequency below 5%, retaining 2 040 959 SNPs. To enhance computational efficiency and minimize redundancy, we further compressed the SNP density using a linkage-disequilibrium-based strategy ($r^2 \geq 0.7$, window size of 125 kb), resulting in the selection of 381 013 tag SNPs. Comparisons of phylogenetic trees and principal-component analysis between the filtered and tag SNPs indicated that using tag SNPs significantly reduced computation time while preserving the consistency of population structure (Supplemental Figure 1). This demonstrates the efficacy of MRBIGR's SNP pruning for population genetic studies.

Of the metabolites analyzed, only 102 were annotated (Supplemental Table 3). We performed clustering analysis and principal-component analyses on their abundances, revealing that temperate maize lines contain a significantly higher flavonoid content compared to tropical maize lines (Supplemental Figure 2). We then used the '*gwas*' module to analyze the abundance of these 102 metabolites (supplemental methods), the expression levels of all genes, and nine agronomic traits. Our results revealed significant association for 14 flavonoid metabolites within a region around 48 Mb on chromosome 1 (Supplemental Figures 3A and 3B). Interestingly, this region includes the *Pericarp Color1* (*P1*) gene, a known regulator of the flavonoid metabolic pathway (Liu et al., 2017). The GWAS signal for *P1* co-localizes with those of the flavonoid metabolites, suggesting that *P1* plays a crucial role in regulating flavonoid content (Supplemental Figure 3B). However, in the GWAS analysis of agronomic traits, only kernel number per row displayed a significant association region (Supplemental Figure 3C), which did not co-localize with any GWAS signals of annotated metabolites, suggesting that GWAS analysis of agronomic traits alone may not effectively reveal the underlying regulatory networks.

To determine whether MR can reveal the causal relationships between genes and metabolites, we used the '*mr*' module to analyze the *P1* gene and 102 annotated metabolites (supplemental methods). The analysis revealed that all statistically significant metabolites were flavonoids (Supplemental Figure 4A), which exhibited strong MR effects (Supplemental Figure 4B). Further analysis of the regulatory relationships between *P1* and other genes, using previously obtained GWAS results, identified 9520 significantly associated loci (SALs) linked to 8929 genes, including *P1*. A subsequent pairwise reciprocal MR analysis between *P1* and each of the 8928 genes (supplemental methods) identified 16 genes significantly associated with *P1* (Supplemental Table 4; Supplemental Figure 4C). Gene Ontology enrichment analysis of these 16 genes using the '*go*' module revealed significant enrichment in pathways related to flavonoid and anthocyanin metabolism (Supplemental Figure 5A). Building on the pairwise MR results among *P1*, the 16 genes significantly

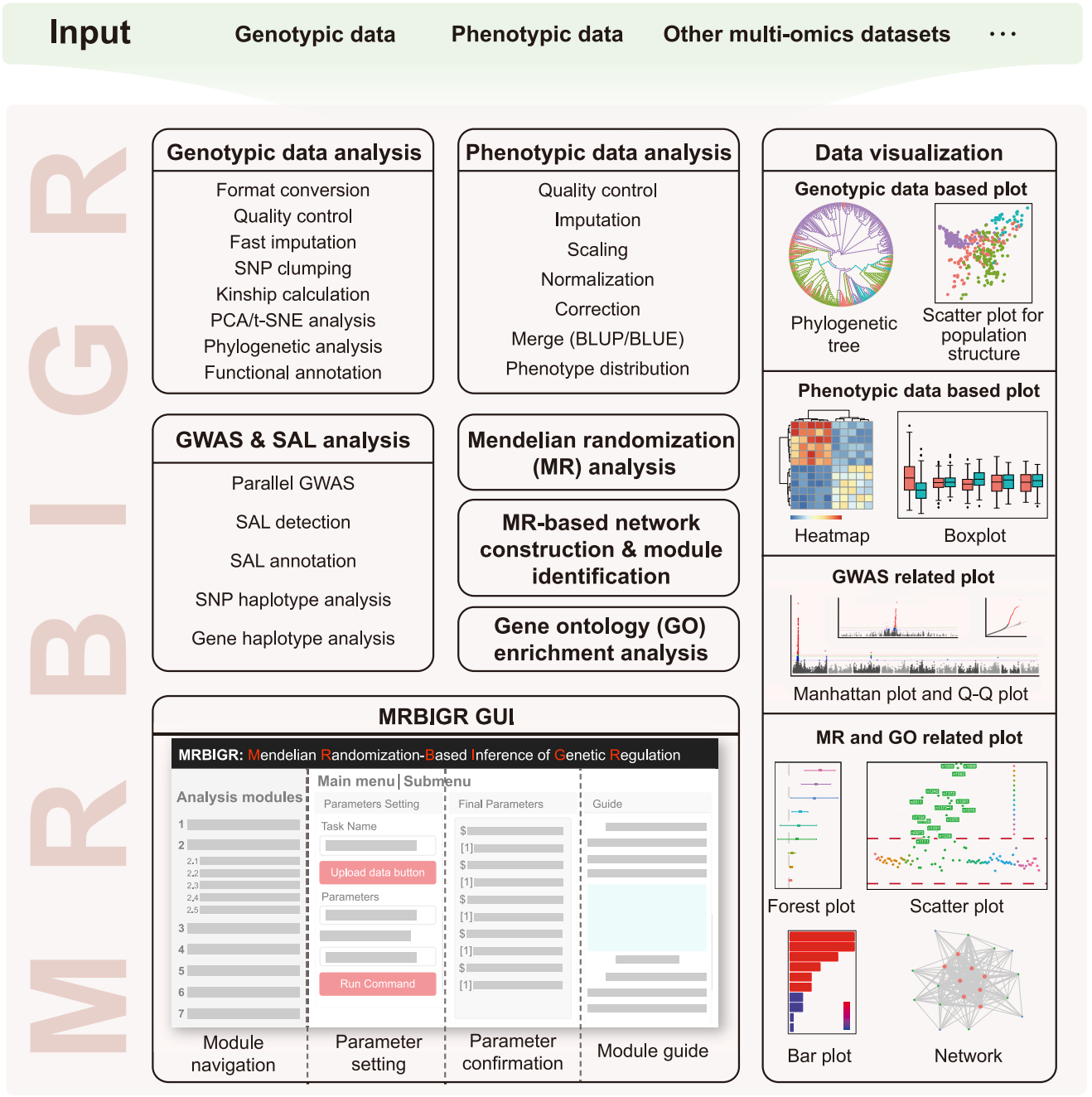

Figure 1. Functional modules of MRBIGR.
MRBIGR features seven modules for genotypic data analysis, phenotypic data analysis, GWAS and SAL analysis, MR analysis, MR-based network analysis, Gene Ontology (GO) enrichment analysis, and data visualization. The software's graphical user interface (GUI) simplifies data analysis by enabling users to intuitively select modules and efficiently set parameters.

associated with *P1*, and genes in the flavonoid metabolism pathway (totaling 62 genes), we constructed a causal network comprising 21 flavonoid-related genes using the '*net*' module, identifying *P1* as a hub gene (Supplemental Figure 5B). Additionally, to evaluate the ability of MR analysis in dissecting the contributions of genes and metabolites to agronomic traits, we performed MR analysis using the quantitative profiles of *P1* and flavonoid metabolites as exposure variables for the nine agronomic traits (supplemental methods). The results indicated that *P1* and flavonoid metabolites had significant positive effects on yield traits, such as kernel number per row and 100-grain weight (100GW), with 100GW showing statistical significance (Supplemental Figure 6). This suggests that *P1*-regulated flavonoid biosynthesis may be involved in grain yield, potentially influencing 100GW.

To investigate the impact of *P1* on flavonoid accumulation during domestication, we employed the '*plot*' module to analyze

nucleotide diversity within a 10-kb window upstream of the *P1* gene. This analysis revealed significantly lower diversity in tropical maize compared to temperate maize (Supplemental Figure 7A). Further analysis of how the genotype at the leading SNP for *P1* affects its expression showed that individuals with the AA genotype consistently exhibit significantly higher *P1* expression levels in both temperate (TEMP) and tropical (TST) maize lines (Supplemental Figure 6B). However, among tropical maize accessions, significantly fewer lines carried the minor allele (AA) compared to the major allele (GG) (Supplemental Figure 7C), indicating that the AA genotype may have undergone selection during domestication, leading to a reduction in its frequency (Supplemental Figure 7A). Notably, several target genes of *P1*, such as *C2*, *CGT1*, *UGT1*, *PR1*, and *FNS1* (Supplemental Table 4), exhibited similar expression patterns to *P1* (Supplemental Figure 7D), further supporting the hypothesis that the minor allele AA is critical for the regulation of flavonoid metabolism.

In summary, we developed MRBIGR, a versatile toolbox that bridges population-scale multi-omics association analysis with genetic regulation inference. A key advantage of MRBIGR is its integration of multiple efficient tools within a user-friendly graphical user interface (GUI), streamlining the analysis of genotypic, phenotypic, and GWAS data (Supplemental Figure 8; Supplemental Tables 1 and 5; supplemental notes). MRBIGR also incorporates MR methods to enable directional causal inference from multi-omics data, providing more precise insights into regulatory networks across different omics layers compared to traditional approaches such as mutual relationship or co-expression analyses (supplemental notes). Using MRBIGR, we analyzed a maize multi-omics dataset and delineated the causal role of the pericarp color-regulating gene *P1* in the flavonoid metabolic network. Additionally, we applied MRBIGR to a rice multi-omics dataset (Gong et al., 2013) and demonstrated that certain lysophosphatidyl cholines negatively influence MR effects on 1000-grain weight (Supplemental Figure 9; supplemental notes), which supports findings that lysophosphatidyl cholines may negatively affect grain weight by modulating starch synthesis pathways (Liu et al., 2013). These findings highlight the multi-species applicability and functional robustness of MRBIGR, as well as the unique advantages of MR for causal inference.

ALGORITHM AND MANUAL AVAILABILITY

For detailed insights into the algorithmic principles and functional implementation of MRBIGR, please refer to the supplemental algorithms. Reproducible scripts for case studies and the user manual for MRBIGR are available at https://gitee.com/crazyhsu/MRBIGR and https://mrbigr.github.io/, respectively.

FUNDING

This study was supported by the Biological Breeding-National Science and Technology Major Project (2023ZD0407501 to X.W.), the Pinduoduo-China Agricultural University Research Fund (PC2024A01003 to X.W.), the National Key Research and Development Program of China (2023YFF1000100 to J.Y.), the National Natural Science Foundation of China (32341036 to J.Y.), the Hebei Provincial Science and Technology Plan Project Modern Breeding Industry Science and Technology Innovation Special Project (21326316D and 21326302D to J.Z.), and the S&T Program of Hebei (215A7612D to J.Z.).

ACKNOWLEDGMENTS

We acknowledge Prof. Xuehai Hu and Prof. Weibo Xie from Huazhong Agricultural University for providing the rice multi-omics data. No conflict of interest is declared.

AUTHOR CONTRIBUTIONS

X.W. and J.Y. conceived the study. J.Y. and J.L. supervised the research. F.X., Q.C., S.L., and J.Y. developed the software, with contributions from S.J., J.Z., and X.M. F.X. and S.L. performed the data analyses. F.X. and Q.C. wrote the manuscript. J.Y., J.L., and X.W. revised the manuscript. All authors reviewed and approved the final manuscript.

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Plant Communications Online*.

Received: May 9, 2024
Revised: July 24, 2024
Accepted: November 22, 2024
Published: November 23, 2024

Feng Xu[1,4], Qian Cheng[1,4], Songyu Liu[1], Shuqin Jiang[2], Jianan Zhang[2], Xiaowei Mao[3], Xiangfeng Wang[1], Jinsheng Lai[1,] and Jun Yan[1,*]*

[1]State Key Laboratory of Maize Bio-breeding, National Maize Improvement Center, Frontiers Science Center for Molecular Design Breeding, China Agricultural University, Beijing 100094, China
[2]Molbreeding Biotechnology Co., Ltd., Shijiazhuang, Hebei Province 051430, China
[3]Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China
[4]These authors contributed equally to this article.
***Correspondence: Jun Yan (yanjun@cau.edu.cn), Jinsheng Lai (jlai@cau.edu.cn)**
https://doi.org/10.1016/j.xplc.2024.101197

REFERENCES

Akiyama, M. (2021). Multi-omics study for interpretation of genome-wide association study. J. Hum. Genet. **66**:3–10.

Gong, L., Chen, W., Gao, Y., Liu, X., Zhang, H., Xu, C., Yu, S., Zhang, Q., and Luo, J. (2013). Genetic analysis of the metabolome exemplified using a rice population. Proc. Natl. Acad. Sci. USA **110**:20320–20325.

Liu, H., Luo, X., Niu, L., Xiao, Y., Chen, L., Liu, J., Wang, X., Jin, M., Li, W., Zhang, Q., et al. (2017). Distant eQTLs and non-coding sequences play critical roles in regulating gene expression and quantitative trait variation in maize. Mol. Plant **10**:414–426.

Liu, H.J., and Yan, J. (2019). Crop genome-wide association study: a harvest of biological relevance. Plant J. **97**:8–18.

Liu, L., Waters, D.L.E., Rose, T.J., Bao, J., and King, G.J. (2013). Phospholipids in rice: Significance in grain quality and health benefits: A review. Food Chem. **139**:1133–1145.

Liu, S., Li, C., Wang, H., Wang, S., Yang, S., Liu, X., Yan, J., Li, B., Beatty, M., Zastrow-Hayes, G., et al. (2020). Mapping regulatory variants controlling gene expression in drought response and tolerance in maize. Genome Biol. **21**:163.

Slaten, M.L., Chan, Y.O., Shrestha, V., Lipka, A.E., and Angelovici, R. (2020). HAPPI GWAS: holistic analysis with pre-and post-integration GWAS. Bioinformatics **36**:4655–4657.

Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., and Meyre, D. (2019). Benefits and limitations of genome-wide association studies. Nat. Rev. Genet. **20**:467–484.

Tang, Y., Liu, X., Wang, J., Li, M., Wang, Q., Tian, F., Su, Z., Pan, Y., Liu, D., Lipka, A.E., et al. (2016). GAPIT version 2: an enhanced integrated tool for genomic association and prediction. Plant Genome **9**. https://doi.org/10.3835/plantgenome2015.11.0120.

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