The applications of new multi-locus GWAS methodologies in the genetic dissection of complex traits, volume II

Book

Published Version

Open Access (Editorial)

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.3389/fpls.2023.1340767

Publisher: Frontiers

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Check for updates

Editorial: The applications of new multi-locus GWAS methodologies in the genetic dissection of complex traits, volume II

Yuan-Ming Zhang[1]*, Zhenyu Jia[2] and Jim M. Dunwell[3]

[1]College of Plant Science and Technology, Huazhong Agricultural University, Wuhan, China, [2]Department of Botany and Plant Sciences, University of California, Riverside, Riverside, CA, United States, [3]School of Agriculture, Policy and Development, University of Reading, Reading, United Kingdom

Editorial on the Research Topic
The applications of new multi-locus GWAS methodologies in the genetic dissection of complex traits, volume II

Since the inception of multi-locus genome-wide association study (GWAS) methodologies (Segura et al., 2012; Liu et al., 2016; Wang et al., 2016), they have been widely applied to dissect the genetics of complex traits (Zhang et al., 2019). Recently, new methodologies such as 3VmrMLM (Li et al., 2022) have been established, resulting in numerous applications. Therefore, it is imperative to consolidate insights into the advantages and potential limitations of using these advanced multi-locus methods.

1 Multi-locus genome-wide association study methods

The evolution of GWAS methods can be divided into three phases: the initial phase of single-marker analysis (Risch and Merikangas, 1996), followed by the emergence of mixed-model-based methods (Zhang et al., 2005; Yu et al., 2006; Kang et al., 2008; Kang et al., 2010; Zhang et al., 2010; Segura et al., 2012; Zhou and Stephens, 2012; Liu et al., 2016), and more recently the integration of mixed models with machine learning methods (Wang et al., 2016; Wen et al., 2018; Li et al., 2022). Currently, rapid single-locus genome-wide scans and multi-locus two-step methods are widely used. However, advocates are leaning towards mixed model plus machine learning methods, such as 3VmrMLM (Li et al., 2022), as they comprehensively consider all effects while controlling for all polygenic backgrounds.

In most methods, the marker genotypes QQ, Qq and qq are typically coded as 2, 1 and 0 respectively, indicating their breeding values in a random mating population. In this context, the parameter to be estimated is the allele substitution effect (α), controlling for the

α-based polygenic background, as in FASTmrEMMA. Locus identification becomes difficult as α approaches zero. Meanwhile, the detection of dominant effects proves challenging due to the small coefficient resulting from similar frequencies of two alleles at a locus, and significant differences in the frequencies of two alleles are equivalent to the presence of a rare allele. When these genotypes are coded as 1, 0 and -1, the estimated effect becomes the additive effect, as seen in methods such as mrMLM. This method is particularly applicable when the majority of marker genotypes are homozygous, as observed in crops such as rice, wheat and soybean. However, the assumption of random mating often does not fit well. Therefore, these situations can lead to reduced power and contribute to missing heritability. To solve these challenges, a recommended approach is to include all effects in a mixed model while controlling for all polygenic backgrounds, as demonstrated in methods such as 3VmrMLM (Li et al., 2022).

When analyzing real data, the inflation factor or quantile-quantile plot serves as a common metric to assess method performance. This is important for single-marker genome-wide scanning methods, as opposed to the multi-locus two-step mrMLM and 3VmrMLM methods. The latter methods use a more relaxed *P*-value threshold during the genome-wide scan, aiming to select potentially associated markers rather than to identify significant loci. Given the complementary nature of these methods, it is often advisable to employ multiple methods when analyzing a single trait (Zhang et al., 2019). This increases the probability of identifying more significant/ suggested loci. A method is good for analyzing the trait when it mines the most known and candidate genes around these loci, in which the candidate genes should be supported by strong evidence. When presenting the results, emphasis can be placed on the highlighted loci containing known and candidate genes in the Manhattan plot.

2 The applications of new multi-locus GWAS methodologies in the genetic dissection of complex traits

Disease resistance is a key trait affecting crop yield. Three studies on this topic identified 54 resistance quantitative trait nucleotides (QTNs), while another study compared selection methods that simultaneously improve yield and resistance. Shu et al. used six multi-locus methods to identify 13 QTNs associated with maize resistance to southern corn rust (SCR). Validation included post-GWAS case-control sampling and allele/haplotype analysis. Candidate genes were mined using transcriptomic annotation analysis and confirmed using tissue-specific and stress-induced transcriptomic analysis. Allele/haplotype effects, resistance and susceptibility of each QTN and QTN-QTN combination in breeding were estimated. These authors advocated a diverse panel of well-designed breeding lines, rich in SNP markers, as a more effective approach for the discovery of small effect and broad SCR resistance loci. Channale et al. identified fourteen chickpea accessions resistant to *Pratylenchus thornei* and 24 resistance QTNs using six multi-locus methods, while six candidate genes were identified around these QTNs under biotic and abiotic stresses, although differential expression and functional analyses were not

performed. Subsequently, Nandudu et al. performed univariate and multivariate GWAS for cassava brown streak disease (CBSD) severity using GEMMA. Univariate GWAS identified five QTNs and multivariate GWAS identified 17 QTNs. Gene ontology analysis mined trait-related candidate genes. In addition, Mediterranean corn borer resistance studies have shown a strong negative correlation between yield and resistance. To determine the effectiveness of genomic selection over phenotypic selection in improving both traits, Gesteiro et al. compared different selection programmes. Genomic selection proved to be the most successful method for improving yield, although phenotypic or genotypic selection for yield may be more effective for improving both traits simultaneously.

To address the problem of over-application of nitrogen fertilizer and to improve nitrogen use efficiency (NUE), Liao et al. performed GWAS for eleven traits in 419 rice landraces, using 208,993 SNPs and the MLM, mrMLM and 3VmrMLM methods. This investigation led to the identification of key QTNs associated with NUE. Eight known genes and 75 candidate genes were identified around these QTNs, and seven candidate genes were further confirmed by RT-qPCR, including *LOC_Os10g33210* and *LOC_Os05g51690*. The results provide valuable genetic resources for molecular breeding of rice cultivars with improved NUE.

3 Future perspectives

Large-scale gene mining is a crucial aspect of future research efforts. On the one hand, the loci identified offer cost-effective opportunities for genomic selection in crops. On the other hand, the wealth of candidate genes around these loci can be explored using multi-omics analysis. In the multi-omics era, a wide range of data, databases, platforms and techniques are becoming available. The integration of genetic loci with multi-omics information is emerging as an inevitable trend that will shape the future of research and exploration in this field.

Author contributions

Y-MZ: Writing – original draft, Writing – review & editing. ZJ: Writing – review & editing. JD: Writing – review & editing.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

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.* 42348–354. doi: 10.1038/ng.548

Kang, H. M., Zaitlen, N. A., Wade, C. M., Kirby, A., Heckerman, D., Daly, M. J., et al. (2008). Efficient control of population structure in model organism association mapping. *Genetics* 178, 1709–1723. doi: 10.1534/genetics.107.080101

Li, M., Zhang, Y. W., Zhang, Z. C., Xiang, Y., Liu, M. H., Zhou, Y. H., et al. (2022). A compressed variance component mixed model for detecting QTNs and QTN-by-environment and QTN-by-QTN interactions in genome-wide association studies. *Mol. Plant* 15, 630–650. doi: 10.1016/j.molp.2022.02.012

Liu, X., Huang, M., Fan, B., Buckler, E. S., and Zhang, Z. (2016). Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genet.* 12, e1005767. doi: 10.1371/journal.pgen.1005767

Risch, N., and Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science* 273, 1516–1517. doi: 10.1126/science.273.5281.1516

Segura, V., Vilhjálmsson, B. J., Platt, A., Korte, A., Seren, Ü., Long, Q., et al. (2012). An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat. Genet.* 44825–830. doi: 10.1038/ng.2314

Wang, S. B., Feng, J. Y., Ren, W. L., Huang, B., Zhou, L., Wen, Y. J., et al. (2016). Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. *Sci. Rep.* 6, 19444. doi: 10.1038/srep19444

Wen, Y. J., Zhang, H., Ni, Y. L., Huang, B., Zhang, J., Feng, J. Y., et al. (2018). Methodological implementation of mixed linear models in multi-locus genome-wide association studies. *Brief. Bioinform.* 19, 700–712. doi: 10.1093/bib/bbw145

Yu, J., Pressoir, G., Briggs, W. H., Bi, I. V., Yamasaki, M., Doebley, J. F., et al. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38, 203–208. doi: 10.1038/ng1702

Zhang, Y. M., Jia, Z., and Dunwell, J. M. (2019). Editorial: the applications of new multi-locus GWAS methodologies in the genetic dissection of complex traits. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00100

Zhang, Y. M., Mao, Y., Xie, C., Smith, H., Luo, L., and Xu, S. (2005). Mapping quantitative trait loci using naturally occurring genetic variance among commercial inbred lines of maize (*Zea mays* L.). *Genetics* 169, 2267–2275. doi: 10.1534/genetics.104.033217

Zhang, Z., Ersoz, E., Lai, C. Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., et al. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* 42, 355–360. doi: 10.1038/ng.546

Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* 44, 821–824. doi: 10.1038/ng.2310