CellPress
PARTNER JOURNAL

Molecular Plant
Opinion

Genome-wide Association Studies in Rice: How to Solve the Low Power Problems?

In human genetics, genome-wide association study (GWAS) is a commonly used method for discovering genes and genetic variants contributing to human traits, usually through statistical examination of the associations between whole-genome sequence variants and one specific trait. The first effort in human GWAS, using genotyping data of 96 patients and 50 healthy controls, successfully identified one gene associated with age-related macular degeneration (Haines et al., 2005). Soon afterward, a large genetics project (The Wellcome Trust Case Control Consortium, 2007) that collected thousands of cases and controls in human populations uncovered many new genes underlying seven common diseases, convincingly demonstrating the power of GWAS in genetic mapping of human traits. Subsequently, the method of GWAS was introduced into plant genetics and improved the handling of various plant populations (Nordborg and Weigel, 2008; McMullen et al., 2009). During recent years, many large-scale GWAS have been carried out in plants, including those in *Arabidopsis*, rice, and maize (Atwell et al., 2010; Huang et al., 2010; Li et al., 2013). In fact, plants have unique advantages for GWAS (Xiao et al., 2017). For example, the mapping populations could be permanent: genotyped or sequenced only once but phenotyped many times for various research purposes.

As one of the model crops, here we used rice as an example to discuss current situations and future perspectives of GWAS in crops. Rice has a wide geographic distribution across the world, and there are a large number of accessions available in public germplasm seed banks. The diverse accessions show many phenotypic differences, and correspondingly contain numerous genomic variants. According to a rice pan-genome dataset generated from the genomes of 67 diverse rice accessions, there were totally 16.5 million single-nucleotide polymorphisms (SNPs), 5.5 million small insertions and deletions (indels), and 0.9 million structural variants (Zhao et al., 2018). For coding genes in rice, on average each gene contained 10 missense SNP sites and six polymorphic sites of relatively large effect, and extensive presence–absence variations were detected for thousands of rice coding genes. Facing a high level of genetic diversity in rice, GWAS is one of the best ways to associate the differences in genomic DNA in various accessions with variation in agronomic trait performances. In rice functional genomics studies, GWAS had helped researchers to identify several important genes underlying multiple complex traits, e.g., *OsSPL13* controlling grain size (Si et al., 2016), *GAD1*/*RAE2* controlling awn length (Yano et al., 2016), and *bZIP73* controlling cold tolerance (Liu et al., 2018). Moreover, GWAS information can be directly used in rice molecular breeding without considering whether the causative genes have been functionally characterized or not, because at most cases the associated SNPs identified from rice GWAS are tightly linked with the causative genes due to comprehensive genotyping data and extensive linkage disequilibrium at the local genomic regions.

Although genome sequencing of thousands of individuals or even larger populations in rice is now possible, there remain several challenges for rice GWAS. As proposed a decade ago (Nordborg and Weigel, 2008), conventional GWAS designs and methods often have low power to map multiple functional alleles within one gene, map rare alleles in the population, and resolve the population structure problems. All three problems exist in rice, which can be reflected from the population-scale genome data. Even without considering variants in the regulation regions, the coding variants have already created multiple alleles from missense SNPs to large-effect frame-shift mutations (see the examples for *waxy* gene controlling grain quality, and *Hd1* gene controlling heading date; Zhao et al., 2018). The rice pan-genome data also demonstrated that most naturally occurring variants are of low frequency in rice populations, and the low-frequency alleles scarcely appear in the peaks of GWAS Manhattan plots even with very large effect sizes. Furthermore, as a selfing species, population structure may be the biggest problem in rice GWAS, because there were many highly differentiated clades (e.g., *indica*, *temperate japonica*, *tropical japonica*, *aus*, *basmati*) in the phylogenetic trees of rice accessions for GWAS.

Therefore, when considering the difficulties of conventional rice GWAS, next-generation permanent populations and next-generation analysis methods will be much needed in rice GWAS. There are at least two choices for next-generation permanent populations in rice: nested association mapping (NAM) populations and multi-parent advanced generation intercross (MAGIC) populations. The NAM experimental design, in which each diverse parental line is crossed with one common reference parent to generate multiple sets of recombinant inbred line populations, has been proved to be very powerful in maize (McMullen et al., 2009). The MAGIC population usually chooses eight or more founder lines to generate F_1 or F_2 lines, followed by multiple intercrossing processes among recombinant lines and multiple generations of selfing. Because the NAM and MAGIC designs can reshuffle the allelic combinations and change the population structures, it could be expected that genetic dissections from these new collaborative populations will discover more genes underlying complex traits in rice. Moreover, it is also important and useful to perform comparative GWAS (e.g., examining the genetic basis of one trait detected in multiple species and searching for the homologous genes) and use metabolites as intermediate traits

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

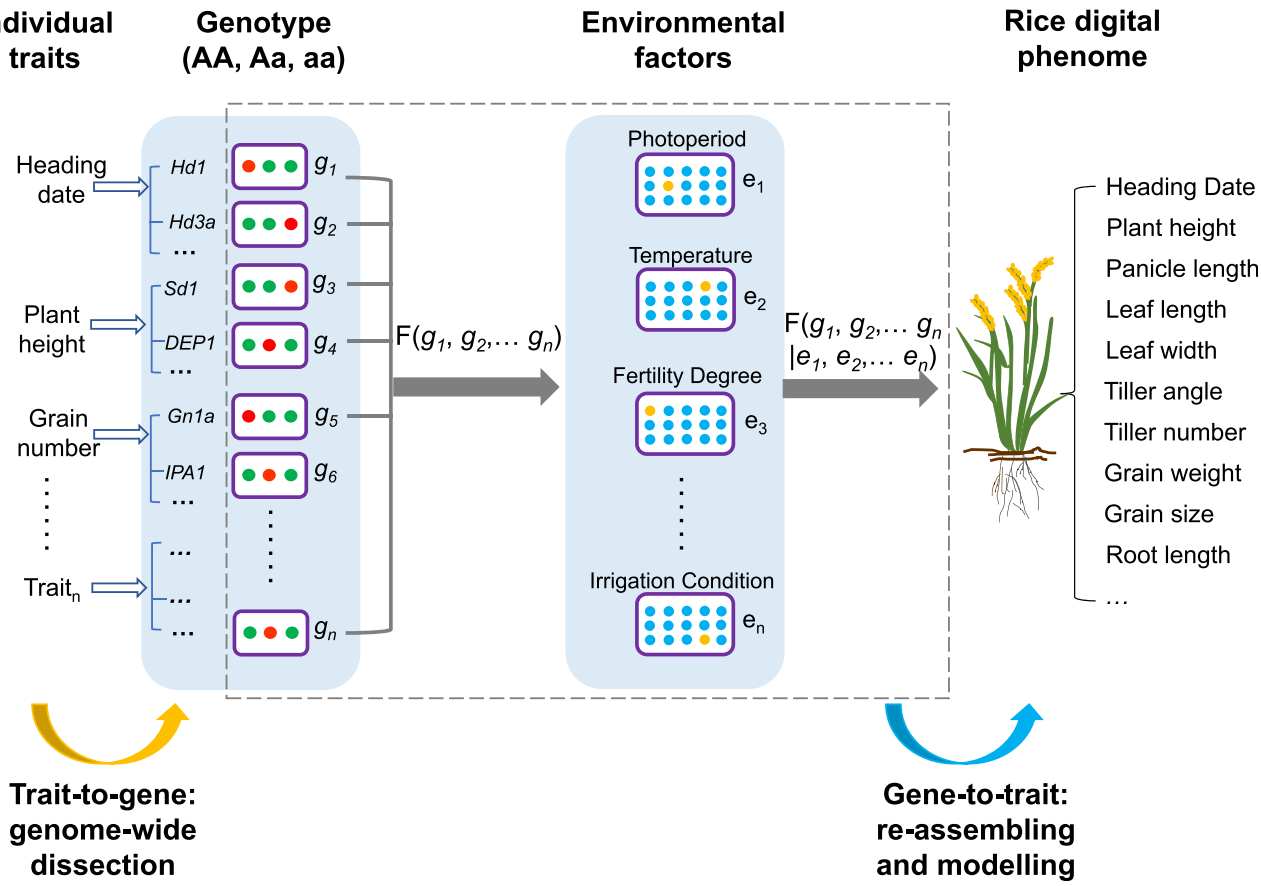

Figure 1. Genetics Studies of Complex Traits in Rice from the Trait-to-Gene Stage to the Gene-to-Trait Stage.
For the genotypes, it is assumed that each causative gene only contained two alleles (A and a) and three genotypes (AA, Aa, and aa), where the genotypes of one hypothetical line of hybrid rice are colored red. For each environmental factor, it is assumed that there are only 15 grades, where the growing environment of the hypothetical line is colored orange.

in dissecting more complex phenotypic traits, as proposed and proved by the studies on rice and maize (Chen et al., 2016). In breeding practices, some of the NAM and MAGIC lines with the breakdown of linkage drags may be useful straightforwardly. Moreover, the value of some exotic lines that have been overlooked currently may deserve new emphasis, and the inclusion of exotic lines in the founder parents of the NAM or MAGIC populations has the potential to make some breakthroughs in future breeding.

The developments of technical methods in molecular biology (e.g., high-throughput sequencing and CRISPR/Cas9), and many studies on genetic mappings and validations, have greatly accelerated rice genetic research (Li et al., 2018). Most quantitative trait loci (QTL) for grain yield, grain quality, and disease resistance in rice, especially those with large effects, have been or will be functionally characterized in the next few years, which, however, does not mean the end of quantitative genetics in rice but provides enormous opportunities (e.g., to *de novo* create ideal alleles of the functionally characterized genes, and to freely design the breeding roadmaps). As for the accompanying challenges, we need to take into account that, although the biological functions of these causative genes are increasingly clear, many concerns about quantitative genetics still await in-depth studies, including the following issues: (1) the precise effects of each of multiple alleles; (2) their genetic effects in heterozygous states; (3) the epistatic interactions between QTL; and (4) the interactions between QTL and environmental factors (e.g., long-day or short-day photoperiod). To be specific, even with all causative genes well characterized, performances of one specific trait cannot be simply predicted for each genotypic combination of causative genes, because the effects of each genotypic combination is not simply the sum of the effects of all of the QTLs (Forsberg et al., 2017), and the heterozygous genotypes rarely act in completely additive mode or completely dominance mode (see the review by Liu and Yan [2018]). This constitutes the importance of the studies of issues (2) and (3). Furthermore, the real situation may be more complicated when there are multiple environmental factors in experimental fields (issue (4)) and there are often more than two alleles in diverse germplasm populations (issue (1)). Next-generation permanent populations, high-profile omics data, functional data, and next-generation analysis approaches (e.g., inspiration from machine learning algorithms) are all needed to generate mathematical models for these issues, which probably will help solve the problems involving missing heritability (the gene loci identified through GWAS and other genetic studies can only explain part of the phenotypic variance) and low phenotype prediction (the predictions based on the genotypes at the identified gene loci are still not very precise).

We believe that quantitative genetics studies on complex traits in rice, and also in some other crops, will step into a new era, from the trait-to-gene stage to the gene-to-trait stage (Figure 1). During the trait-to-gene stage, the primary focus is the genetic dissections of complex traits and the identification of all causative genes one by one. At the gene-to-trait stage when most causative genes have been functionally validated, we might be able to reassemble these gene/QTL elements, plus the known environmental factors, through methods such as mathematical modeling, along with deeper understanding of the molecular mechanisms underlying gene regulatory networks. At that time we can ascertain how and why each rice genome is going to perform from a systems perspective. The rice community may need another decade to fill the gaps between numerous causative genes and the phenotypic performances of the whole rice system.

Received: November 19, 2018
Revised: November 19, 2018
Accepted: November 20, 2018
Published: December 10, 2018

*Xiaoyi Zhou and Xuehui Huang**

College of Life Sciences, Shanghai Normal University, Shanghai 200031, China
***Correspondence: Xuehui Huang (xhhuang@shnu.edu.cn)**
https://doi.org/10.1016/j.molp.2018.11.010

REFERENCES

Atwell, S., Huang, Y.S., Vilhjálmsson, B.J., Willems, G., Horton, M., Li, Y., Meng, D., Platt, A., Tarone, A.M., Hu, T.T., et al. (2010). Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. Nature **465**:627–631.

Chen, W., Wang, W., Peng, M., Gong, L., Gao, Y., Wan, J., Wang, S., Shi, L., Zhou, B., Li, Z., et al. (2016). Comparative and parallel genome-wide association studies for metabolic and agronomic traits in cereals. Nat. Commun. **7**:12767.

Forsberg, S.K., Bloom, J.S., Sadhu, M.J., Kruglyak, L., and Carlborg, Ö. (2017). Accounting for genetic interactions improves modeling of individual quantitative trait phenotypes in yeast. Nat. Genet. **49**:497–503.

Haines, J.L., Hauser, M.A., Schmidt, S., Scott, W.K., Olson, L.M., Gallins, P., Spencer, K.L., Kwan, S.Y., Noureddine, M., Gilbert, J.R., et al. (2005). Complement factor H variant increases the risk of age-related macular degeneration. Science **308**:419–421.

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., Li, C., Zhu, C., Lu, T., Zhang, Z., et al. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. Nat. Genet. **42**:961–967.

Li, H., Peng, Z., Yang, X., Wang, W., Fu, J., Wang, J., Han, Y., Chai, Y., Guo, T., Yang, N., et al. (2013). Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. Nat. Genet. **45**:43–50.

Liu, H., and Yan, J. (2018). Crop genome-wide association study: A harvest of biological relevance. Plant J. https://doi.org/10.1111/tpj.14139.

Li, Y., Xiao, J., Chen, L., Huang, X., Cheng, Z., Han, B., Zhang, Q., and Wu, C. (2018). Rice functional genomics research: past decade and future. Mol. Plant **11**:359–380.

Liu, C., Ou, S., Mao, B., Tang, J., Wang, W., Wang, H., Cao, S., Schläppi, M.R., Zhao, B., Xiao, G., et al. (2018). Early selection of bZIP73 facilitated adaptation of *japonica* rice to cold climates. Nat. Commun. **9**:3302.

McMullen, M.D., Kresovich, S., Villeda, H.S., Bradbury, P., Li, H., Sun, Q., Flint-Garcia, S., Thornsberry, J., Acharya, C., Bottoms, C., et al. (2009). Genetic properties of the maize nested association mapping population. Science **325**:737–740.

Nordborg, M., and Weigel, D. (2008). Next-generation genetics in plants. Nature **456**:720–723.

Si, L., Chen, J., Huang, X., Gong, H., Luo, J., Hou, Q., Zhou, T., Lu, T., Zhu, J., Shangguan, Y., et al. (2016). OsSPL13 controls grain size in cultivated rice. Nat. Genet. **48**:447–456.

The Wellcome Trust Case Control Consortium. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature **447**:661–678.

Xiao, Y., Liu, H., Wu, L., Warburton, M., and Yan, J. (2017). Genome-wide association studies in maize: praise and stargaze. Mol. Plant **10**:359–374.

Yano, K., Yamamoto, E., Aya, K., Takeuchi, H., Lo, P.C., Hu, L., Yamasaki, M., Yoshida, S., Kitano, H., Hirano, K., 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, Q., Feng, Q., Lu, H., Li, Y., Wang, A., Tian, Q., Zhan, Q., Lu, Y., Zhang, L., Huang, T., et al. (2018). Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. Nat. Genet. **50**:278–284.