



IntAssoPlot: An R Package for Integrated Visualization of Genome-Wide Association Study Results With Gene Structure and Linkage Disequilibrium Matrix

Fengyu He^{1,2,3†}, Shuangcheng Ding^{1,2,3†}, Hongwei Wang^{1,2,3*} and Feng Qin^{4*}

¹ Engineering Research Center of Ecology and Agricultural Use of Wetland, Ministry of Education, Agricultural College, Yangtze University, Jingzhou, China, ² Hubei Key Laboratory of Waterlogging Disaster and Agricultural Use of Wetland, Agricultural College, Yangtze University, Jingzhou, China, ³ Hubei Collaborative Innovation Center for Grain Industry, Agricultural College, Yangtze University, Jingzhou, China, ⁴ College of Biological Sciences, China Agricultural University, Beijing, China

OPEN ACCESS

Edited by:

Nunzio D'Agostino,
Università degli Studi di Napoli
Federico II, Italy

Reviewed by:

Ramiro Magno,
University of Algarve, Portugal
Pariya Behrouzi,
Wageningen University & Research,
Netherlands
Julio Isidro Sanchez,
University College Dublin, Ireland

*Correspondence:

Hongwei Wang
wanghw@yangtzeu.edu.cn;
whweve@163.com
Feng Qin
qinfeng@cau.edu.cn

† These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Bioinformatics and Computational
Biology,
a section of the journal
Frontiers in Genetics

Received: 02 January 2020

Accepted: 04 March 2020

Published: 20 March 2020

Citation:

He F, Ding S, Wang H and Qin F
(2020) IntAssoPlot: An R Package
for Integrated Visualization
of Genome-Wide Association Study
Results With Gene Structure
and Linkage Disequilibrium Matrix.
Front. Genet. 11:260.
doi: 10.3389/fgene.2020.00260

Genome-wide association study (GWAS), exploring the historical and evolutionary recombinations at the population level, is a major method adopted to identify quantitative trait loci (QTL) for complex traits. However, to summarize GWAS results, gene structure, and linkage disequilibrium (LD) in a single view, multiple tools are required. It is tedious to generate these three results and manually put them together; moreover, it may eventually lead to inaccuracies. On the other hand, genotype markers are usually detected by DNA- and/or RNA-Seq. For GWAS analysis based on RNA-Seq, markers from DNA-Seq provide more genetic information when displaying LD. The currently released software package does not have this function for an integrated analysis of LD, using genotypic markers different from that in association analysis. Here, we present an R package, IntAssoPlot, which provides an integrated visual display of GWAS results, along with LD and gene structure information, in a publication-ready format. The main panel of an IntAssoPlot plot has a connecting line linking the genome-wide association *P*-values on the $-\log_{10}$ scale with the gene structure and LD matrix. Importantly, IntAssoPlot is designed to plot GWAS results with LD calculated from genotypes different from those in GWAS analysis. IntAssoPlot provides a powerful visualization tool to gain an integrated insight into GWAS results. The functions provided by IntAssoPlot increase the efficiency by revealing GWAS results in a publication-ready format. Inspection of the output image can provide important biological information, including the loci that passed the genome-wide significance threshold, genes located at or near the significant loci, and the extent of LD within the selected region.

Keywords: genome-wide association study, gene structure, LD, visualization, linking line, integration

INTRODUCTION

During the past few years, a sufficient number of molecular markers and availability of fast and accurate variance component estimation methods have made GWAS an ideal tool to detect genetic relationships with complex traits (Mackay, 2001; Yu and Buckler, 2006; Wang and Qin, 2017). However, it is difficult to summarize GWAS results, genomic context, and LD in a single

and publication-ready view. Generally, researchers use Haploview (Barrett et al., 2005) and plink (Purcell et al., 2007) for LD analysis and visualization, CSDS (Hu et al., 2015) and Phytozome (Goodstein et al., 2012) for gene structure visualization, and TASSEL (Bradbury et al., 2007) and GAPIT (Lipka et al., 2012) for visualizing associations. Subsequently, the images generated are compiled together by manual copy and paste steps.

To efficiently visualize GWAS results, packages such as LocusZoom, cgmisc, Ldlink, and Assocplots have been developed (Pruim et al., 2010; Kierczak et al., 2015; Machiela and Chanock, 2015; Khramtsova and Stranger, 2017). LocusZoom is a web-, Java- and R-based tool, Ldlink is a web-based tool, and cgmisc and Assocplots are stand-alone tools based on (Ihaka and Gentleman, 1996) and Python (Sanner, 1999), respectively. Ideally, the tools for visualizing GWAS results should represent information detailing (1) the loci passing the genome-wide significance threshold, (2) the genes present at or near the significant loci, and (3) the linkage disequilibrium (LD) structure of the significant loci. In order to display the information listed above, the function `plot.manhattan.genes`, in the package `cgmisc`, arranges the relative plots together in a single page. Even though the LocusZoom and `cgmisc` can display regional GWAS information, such as the association of signal relative to genomic position and LD (LD between the most significant associated loci with the rest), no connecting line linking the significant loci, gene structure, and LD matrix is shown. Furthermore, single nucleotide polymorphism (SNP) genotypes detected by RNA-Seq are usually located in the coding and transcriptional regulatory regions [16], which cannot fully represent the genomic variation. Therefore, for the GWAS results calculated from SNPs detected by RNA-Seq, it is necessary, though difficult, to display GWAS results with LD calculated from genome-wide SNPs by re-sequencing.

We, therefore, developed an R package called `IntAssoPlot`, which would not only display the relative significance loci, gene structure, and linkage disequilibrium matrix derived from marker dataset (which is same or different from that for GWAS), but would also draw a connecting line to link them in one panel without the need for further manual arrangement. Additionally, `IntAssoPlot` is designed to show GWAS results at a single gene level, displaying not only information related to significant loci and linkage disequilibrium structure, but also that related to the detailed genetic structures of the significant loci.

MATERIALS AND METHODS

Implementation

`IntAssoPlot`, built on `ggplot2` (Wickham, 2016) R package, imports R packages including (Zheng et al., 2012), `gdsfmt` (Zheng et al., 2012, 2017), `ggrepel`¹, and `reshape2` (Wickham, 2007). Its basic functionality includes plotting GWAS results at the level of a single gene (`IntGenicPlot`) and a single chromosomal region (`IntRegionalPlot`). Owing to the characteristics of the R

programming language (Ihaka and Gentleman, 1996), all these tools are open-source, and the package is easy to use and platform-free for bioinformaticians.

Plot Methods

Modules provided by `IntAssoPlot` are listed and compared with other packages in **Tables 1, 2**, respectively. The output of `IntAssoPlot` contains three layers, integrated to display scatter plot of *P*-values for marker-trait associations, LD heat map of lead SNP (the most significant associated SNP or selected SNP) with the rest, LD heat map matrix, highlighted selected marker, and linking line (**Table 1** and **Supplementary Figure 1**).

Features of `IntAssoPlot` are compared with those of released software in **Table 2**. As shown in **Table 2**, `IntAssoPlot` could integrated display of marker-trait associations, gene structure, and LD matrix. Importantly, `IntAssoPlot` could facilitate genotype datasets different from those in GWAS to generate an LD matrix. Moreover, the color, shape, and size of approximately all of lead SNPs, highlighted markers, and LD heat map could be modified in `IntAssoPlot`, making the output more versatile and informative.

There are four inputs required for `IntGenicPlot` in `IntAssoPlot`, namely transcript, genome annotation file, GWAS result, and genotype file. For `IntRegionalPlot`, the chromosome, left border (start position), right border (end position), genome annotation file, GWAS result, and genotype file are required. The main panel of `IntAssoPlot` contains three layers and connecting lines. The upper layer shows *P*-values for association on the $-\log_{10}$ scale in the vertical axis. The middle layer presents the chromosomal position along the horizontal axis with the gene structure while the lower layer shows the LD matrix.

RESULTS AND DISCUSSION

IntAssoPlot Workflow

R (The Comprehensive R Archive Network²) and RStudio (an integrated development environment for R³) should be downloaded and installed on the computer system. Currently, `IntAssoPlot` is hosted on GitHub. To download and build `IntAssoPlot` from GitHub, an R package devtools is required. Importantly, `IntAssoPlot` also imports some other R packages. When launching R, the depended packages, devtools, and `IntAssoPlot` can be installed using the following commands in R environment:

```
> install.packages(c("devtools", "ggplot2", "ggrepel", "reshape2",
"BiocManager"))
> BiocManager::install(c("SNPRelate", "gdsfmt"))
> devtools::install_github("whweve/IntAssoPlot", build =
TRUE, build_opts = c("--no-resave-data", "--no-manual"))
```

To generate images by `IntAssoPlot`, there are two steps: (1) data input, (2) generate the image. Basically, `IntAssoPlot` needs three types of data, including association mapping results, gene

¹<https://CRAN.R-project.org/package=ggrepel>

²<https://cran.r-project.org/>

³<https://www.rstudio.com/>

TABLE 1 | Summary of functions implemented in IntAssoPlot.

Function	Layer	Shared modules	Differed modules
IntGenicPlot	Upper layer: a scatter plot displaying $-\log_{10}(P)$ for marker-trait associations. Middle layer: Gene annotation. Bottom layer: LD matrix	Scatter plot of $-\log_{10}(P)$ and marker site, LD heatmap of lead SNP (the most significant associated SNP or selected SNP) with the rest, LD matrix heatmap, highlight selected marker, linking line	Gene structure
IntRegionalPlot			Arrowed line reflecting gene length and strand

TABLE 2 | Comparison of features provided in GWAS visualization tools.

Features in a single view	LocusZoom	cgmisc	Ldlink	Assocplots	IntAssoPlot
Scatter plot of marker-trait association	✓	✓	✓	✓	✓
Lead SNP LD heatmap	✓		✓		✓
LD matrix heatmap			✓		✓
Linking line from association to LD matrix					✓
LD matrix from another genotype dataset					✓
Gene structure	✓				✓
Integrated display of association, gene structure, and LD matrix					✓

annotation file, and genotype file, of which we have included detailed examples of these data in IntAssoPlot. When reading data into the R environment, R function `read.table` can be used. Below, we provide a detailed case study to illustrate the usage of IntAssoPlot.

Case Study

We present the features of IntAssoPlot and provide an example of the resulting plot (**Figure 1**) using previously published data on the genetic basis of maize drought tolerance in young maize seedlings (Wang et al., 2016), and we have included commands used to generate **Figure 1** in **Table 3**.

Chromosomal regions can be selected in the following two ways: (1) the chromosome start and stop positions (**Figures 1A–D**) or (2) a transcript name (**Figures 1E,F**), with or without the specified marker. If no marker is specified, the most strongly associated marker will be treated as the lead SNP, while a specified marker would be treated as the lead SNP, if provided. Available modules for displaying LD (r^2 or D') include lead SNP LD (leadsnpLD in IntAssoPlot) and triangle LD (triangleLD in IntAssoPlot). While the lead SNP LD function calculates LD of the most significant or user-provided loci with the others, the triangle LD function calculates the matrix of each pairwise LD from the specified region. All LD values are colored according to the specified color scale, which can be further modified, allowing the package to fit into broader applications (**Figures 1A,B**). Currently, researchers have released multiple sets of genotype data. In our package, genotype files for displaying lead SNP LD and association results were assumed to originate from the same set of genotype files. However, in order to take advantage of the massive genotype information, markers for displaying triangle LD could be derived from another set of genotype data (**Figure 1C**). The output image, generated by IntRegionalPlot, which is a function in IntAssoPlot, could be zoomed-in or -out

(**Figure 1D**) by setting the length of left or/and right border. Meanwhile, IntGenicPlot allows us to show the association and LD related to flanking sequences, the length of which can be modified by the user (**Figure 1F**), thus revealing the effect of the promoter on the association results.

Notably, according to the input, linking lines and association signals can be selected and highlighted (**Figures 1G,H**), respectively. In IntGenicPlot and IntRegionalPlot, when setting a threshold for the genome-wide association significance level, a connecting line linking the significant association signals (removing completely marker), gene structure, and LD matrix will be drawn by default (**Figures 1A–G**). If linking association signals are provided, linking lines can be reset accordingly (**Figure 1H**). Meanwhile, association signals in the upper layer could be highlighted by resetting the shape, color, and size (**Figure 1G**), thus making the output more useful.

Generally, generating an image by the function IntGenicPlot requires very limited time. However, for integrated visualization of GWAS results on a regional scale, the time is limited by the number of markers and the samples required to display the LD matrix. In the current example (**Figure 1C**), we have selected 2355 SNPs located over a region of 400 kbp in 104 samples (Chia et al., 2012) to display the LD matrix. Thus, 2,771,835 LD dots need to be displayed. When tested on a desktop computer with an Intel® Core™ i7-6700K processor and Windows 10 operation system, it was found to take less than 30 s to generate the integrated image, hence demonstrating the efficiency and performance of IntAssoPlot. However, <3 min will be consumed to print the image in the R environment. In **Figures 1A,B**, working with 368 samples and 270 SNPs, <30 s was required to generate and print the output image. Therefore, future studies should aim to reduce the time required to display the large LD matrix. It was reported that, based on undirected graphical models, netgwas can reconstruct LD networks adjusting for markers

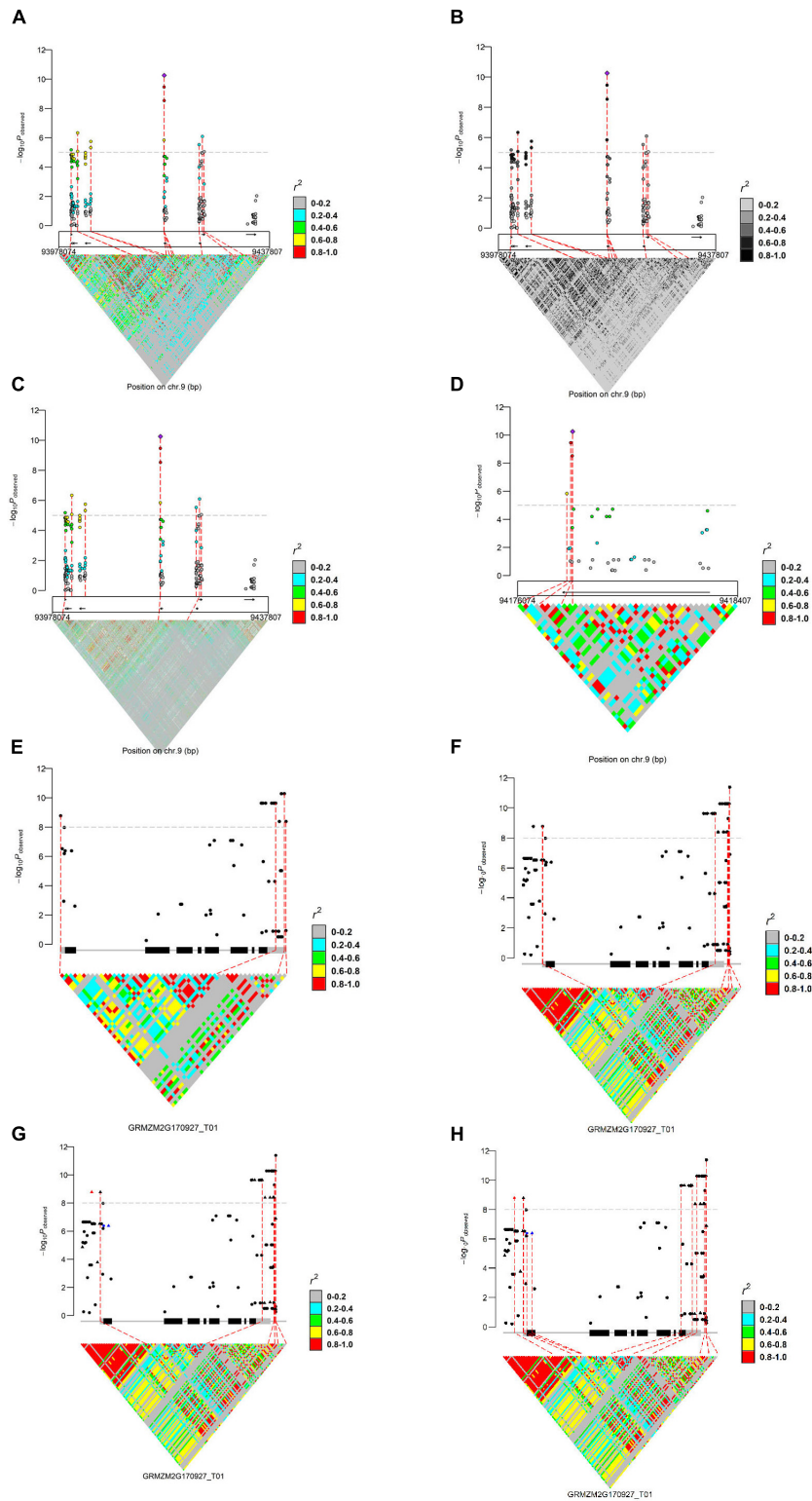


FIGURE 1 | Regional marker-trait associations plot (**A**: regional plot; **B**: LD color scale; **C**: LD matrix from another genotypes; and **D**: zoom-out) and a single gene-based marker-trait associations plot (**E**: single genic plot; **F**: adjusting flanking sequence; **G**: highlighting markers; and **H**: linking markers), using previously published data (Wang et al., 2016). For (**A–D**), the upper layer represents the marker-trait associations and LD of the most significant loci with the others; the middle layer shows the filtered gene models indicated by arrows (MaizeGDB release 5b.60); the bottom layer shows LD matrix. In (**E–H**), the main transcript structure of *ZmVPP1* and the LD matrix are shown. Genotypes used to generate these plots were derived from Chia et al. (2012), Li et al. (2013), and Wang et al. (2016).

TABLE 3 | R commands used to generate **Figure 1**.

Scale	Operation	Command	Output
Regional plot	Lead SNP LD + triangle LD + directioned gene length + linking line	<code>IntRegionalPlot(chr = 9,left = 94178074-200000,right = 94178074 + 200000,gtf = gtf,association = association,hapmap = hapmap_am368,hapmap_ld = hapmap_am368,threshold = 5, leadsnp_size = 2)</code>	Figure 1A
	LD color scale	<code>IntRegionalPlot(chr = 9,left = 94178074-200000,right = 94178074 + 200000,gtf = gtf,association = association,hapmap = hapmap_am368,hapmap_ld = hapmap_am368,threshold = 5,leadsnp_size = 2,color02 = "gray81",color04 = "gray61",color06 = "gray41",color08 = "gray11",color10 = "gray1")</code>	Figure 1B
	Triangle LD from another set of genotype data	<code>IntRegionalPlot(chr = 9,left = 94178074-200000,right = 94178074 + 200000,gtf = gtf,association = association,hapmap = hapmap_am368,hapmap_ld = hapmap2,threshold = 5, leadsnp_size = 2)</code>	Figure 1C
	Zoom-in or -out	<code>IntRegionalPlot(chr = 9,left = 94178074-2000,right = 94178074 + 6000,gtf = gtf,association = association,hapmap = hapmap_am368,hapmap_ld = hapmap_am368,threshold = 5,leadsnp_size = 2)</code>	Figure 1D
Single genic plot	Triangle LD + gene structure + linking line	<code>IntGenicPlot("GRMZM2G170927_T01",gtf,association = zmvpp1_association,hapmap = zmvpp1_hapmap,leadsnp = FALSE,triangleLD = TRUE,threshold = 8)</code>	Figure 1E
	Flanking sequence	<code>IntGenicPlot("GRMZM2G170927_T01",gtf,association = zmvpp1_association,hapmap = zmvpp1_hapmap,leadsnp = FALSE,triangleLD = TRUE, threshold = 8,up = 500,down = 600)</code>	Figure 1F
	Highlight marker by shape and/or color	<code>IntGenicPlot("GRMZM2G170927_T01",gtf,association = zmvpp1_association,hapmap = zmvpp1_hapmap,leadsnp = FALSE,triangleLD = TRUE,threshold = 8,up = 500,down = 600,marker2highlight = marker2highlight)</code>	Figure 1G
	Selected markers linking from association to LD	<code>IntGenicPlot("GRMZM2G170927_T01",gtf,association = zmvpp1_association,hapmap = zmvpp1_hapmap,leadsnp = FALSE,triangleLD = TRUE,threshold = 8,up = 500,down = 600,marker2highlight = marker2highlight,link2gene = marker2link,link2LD = marker2link)</code>	Figure 1H

relationships (Behrouzi et al., 2017), which will be extended to IntAssoPlot in future development. On the other hand, by uploading IntAssoPlot at GitHub, we can make IntAssoPlot publicly accessible and invite the scientific community to contribute and enhance the capabilities of the package.

CONCLUSION

At present, it is difficult to summarize GWAS results, genomic context, and linkage disequilibrium (LD) in a single view. The currently available tools cannot integrate, analyze, and display LD using genotypic markers different from those in association analysis. Therefore, we were prompted to present IntAssoPlot, an R package, aimed to solve these challenges. The output image of an IntAssoPlot plot, which is in a publication-ready format, draws a line connecting the P -values on a $-\log_{10}$ scale,

the genome annotation, and the LD matrix, calculated from genotypes same or different from those in GWAS analysis. Inspection of the output image could reveal: (1) the extent of LD within a region with the most significant associated variants, providing important biological evidence to infer the candidate gene, (2) the most significantly associated genetic feature located within the candidate gene, and (3) the LD between the loci, demonstrating the mechanism of functional variation for further experimental validation.

Functions in IntAssoPlot are designed to be user-friendly, with attention to the quality of visualization, owing to the characteristics of R programming language. Documentation, outlining the detailed functionality, and parameter usage of each function, is available for each of the functions in IntAssoPlot. Detailed examples of the usage of IntAssoPlot have been created based on the analysis of data from a previously reported study on the genetic basis of maize drought tolerance

(Wang et al., 2016). The IntAssoPlot package is available at <https://github.com/whweve/IntAssoPlot>.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the <https://github.com/whweve/IntAssoPlot>.

AUTHOR CONTRIBUTIONS

FH, SD, and HW contributed to implementation, documentation, and manuscript writing. FQ contributed to conception, project design, oversight, and manuscript review. All authors read and approved the final manuscript.

FUNDING

This work was supported by the Hubei Province Educational Commission (Q20171307), National Natural Science Foundation

REFERENCES

- Barrett, J. C., Fry, B., Maller, J., and Daly, M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21, 263–265. doi: 10.1093/bioinformatics/bth457
- Behrouzi, P., Arends, D., and Wit, E. C. (2017). netgwas: An R package for network-based genome-wide association studies. *arXiv [Preprint]*.
- Bradbury, P. J., Zhang, Z. W., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308
- Chia, J. M., Song, C., Bradbury, P. J., Costich, D., de Leon, N., Doebley, J., et al. (2012). Maize HapMap2 identifies extant variation from a genome in flux. *Nat. Genet.* 44, 803–807. doi: 10.1038/ng.2313
- Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40, D1178–D1186. doi: 10.1093/nar/gkr944
- Hu, B., Jin, J., Guo, A., Zhang, H., Luo, J., and Gao, G. (2015). GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* 31, 1296–1297. doi: 10.1093/bioinformatics/btu817
- Ihaka, R., and Gentleman, R. (1996). R: A language for data analysis and graphics. *J. Comput. Graph. Stat.* 5, 299–314. doi: 10.2307/1390807
- Khrantsova, E. A., and Stranger, B. E. (2017). Assocplots: a Python package for static and interactive visualization of multiple-group GWAS results. *Bioinformatics* 33, 432–434. doi: 10.1093/bioinformatics/btw641
- Kierczak, M., Jablonska, J., Forsberg, S. K., Bianchi, M., Tengvall, K., Pettersson, M., et al. (2015). Cgmisc: enhanced genome-wide association analyses and visualization. *Bioinformatics* 31, 3830–3831. doi: 10.1093/bioinformatics/btv426
- Li, H., Peng, Z., Yang, X., Wang, W., Fu, J., Wang, J., et al. (2013). Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nat. Genet.* 45, 43–50. doi: 10.1038/ng.2484
- Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., et al. (2012). GAPIT: genome association and prediction integrated tool. *Bioinformatics* 28, 2397–2399. doi: 10.1093/bioinformatics/bts444
- Machiela, M. J., and Chanock, S. J. (2015). LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* 31, 3555–3557. doi: 10.1093/bioinformatics/btv402
- Mackay, T. F. C. (2001). Quantitative trait loci in *Drosophila*. *Nat. Rev. Genet.* 2, 11–20. doi: 10.1038/35047544

of China (31701062 and 31701439), Natural Science Foundation of Hubei Province (2019CFB616), and the Opening Fund of Engineering Research Center of Ecology and Agricultural Use of Wetland, Ministry of Education (KF201912). The funders played no role in the study design, data collection, analysis, decision to publish, or in preparation of the manuscript.

ACKNOWLEDGMENTS

The authors thank Shengxue Liu and Xianglan Wang for testing and making suggestions for IntAssoPlot.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00260/full#supplementary-material>

- Pruim, R. J., Welch, R. P., Sanna, S., Teslovich, T. M., Chines, P. S., Gliedt, T. P., et al. (2010). LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 26, 2336–2337. doi: 10.1093/bioinformatics/btq419
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795
- Sanner, M. F. (1999). Python: a programming language for software integration and development. *J. Mol. Graph. Model* 17, 57–61.
- Wang, H., and Qin, F. (2017). Genome-wide association study reveals natural variations contributing to drought resistance in crops. *Front. Plant Sci.* 8:1110. doi: 10.3389/fpls.2017.01110
- Wang, X., Wang, H., Liu, S., Ferjani, A., Li, J., Yan, J., et al. (2016). Genetic variation in ZmVPP1 contributes to drought tolerance in maize seedlings. *Nat. Genet.* 48, 1233–1241. doi: 10.1038/ng.3636
- Wickham, H. (2007). Reshaping data with the reshape package. *J. Stat. Softw.* 21, 1–20. doi: 10.3978/j.issn.2305-5839.2016.01.33
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Berlin: Springer.
- Yu, J., and Buckler, E. S. (2006). Genetic association mapping and genome organization of maize. *Curr. Opin. Biotechnol.* 17, 155–160. doi: 10.1016/j.copbio.2006.02.003
- Zheng, X., Gogarten, S. M., Lawrence, M., Stilp, A., Conomos, M. P., Weir, B. S., et al. (2017). SeqArray—a storage-efficient high-performance data format for WGS variant calls. *Bioinformatics* 33, 2251–2257. doi: 10.1093/bioinformatics/btx145
- Zheng, X., Levine, D. K., Shen, J., Gogarten, S. M., Laurie, C. C., and Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28, 3326–3328. doi: 10.1093/bioinformatics/bts606

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.

Copyright © 2020 He, Ding, Wang and Qin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.