Final Project
(Option B: GWAS+GS on real or simulated traits)
Genomic Prediction Workshop
September 6-10, 2023
Beijing, China
Website: https://zzlab.net/Beijing2023GS
Instructors: Dr. Zhiwu Zhang and Dr. Jiabo Wang

Due on November 10, 2023, 5PM (Beijing time)

Data files: Choose a dataset from the recommended list (https://zzlab.net/Beijing2023GS/Doc/PublicData.pdf), or a dataset outside the list (please specify source, 5 extra points), or your own data that are released to the public (please specify source, 10 extra points).

Hand in: Email wangjiaboyifeng@163.com with subject of "Beijing2023GS Project". The email should contain two links, one to GitHub hosting your R code and data source. The other links to preprint.org hosting your manuscript. The email should identify your real name and indicate if extra credit is applicable.

Marker Assisted Selection (MAS) is the earliest format of molecular breeding based on linkage analysis or Genome-Wide Association Study (GWAS). Genomic Prediction, or Genomic Selection (GS) in animal or plant breeding, uses all available markers regardless of if they are statistically significant or not to predict genetic merit of individuals, especially for complex traits. Literature claims that incorporating GWAS results in genomic selection (GS) would improve accuracy than conducting GS and MAS along (e.g. two fold accuracy increase reported by Ravelombola et al., 2020). However, many of the reports have an invalidate procedure for such claim. As reviewed by McGowan et al. (2021) on Plant Breeding Reviewer (can be accessed by the links in Reference), invalidate procedure can artificially create fake accuracy improvement for such incorporation. There is critical needs to conduct validate procedure if there are conditions that the incorporation improves genomic prediction accuracy and provide guidance to apply the incorporation.

Your investigation should be based on the comparisons of the incorporation to both GS and MAS to address at least one of the following topics.

1. Marker density
When marker density is too low, there are barely chances to identified associated markers in GWAS. The value of incorporation is dramatically reduced. Appropriate marker density is essential for the incorporation.

2. Multiple real traits
Traits are different in regard of genetic architecture, including number of genes and heritability. The knowledge of incorporation is beneficial to breeding on the specific traits.

3. Simulated traits

Simulation has advantages of knowing the genetic architecture underlying a trait, including the true breeding values of individuals. Number of genes and heritability can be varied to find appropriate combinations to incorporate GWAS into GS.

4. GWAS models

There are substantial variations of statistical power among different GWAS models. A model with high statistical power provides valuable prior knowledge and less noise for the incorporation in GS.

5. GS models

Some of GS models also have additional function for GWAS. Therefore, the benefit of incorporating GWAS varies among GS models.

6. Incorporation procedures

There are multiple procedures to incorporate GWAS results in GS, including fitting the associated markers as covariates in BLUP and Bayesian models (Spindel, 2016) or building kinship using associated markers in gBLUP (Zhang et al. 2014).

Your R code should be bug free (10 points) and contain comments to illustrate the purpose (10 points). Your manuscript should contain the following sections:
- Title (5 points)
- Abstract (10 points)
- Introduction (10 points)
- Method (15 points)
- Results (15 points)
- Discussion (10 points)
- Conclusion (10 points)
- References (5 points)

Reference
1. Bernardo, R. 1994. **Prediction of maize single-cross performance using RFLPs and information from related hybrids**. Crop Sci. 34: 20–25. **https://doi.org/10.2135/cropsci1994.0011183X003400010003x**
2. Brian Rice and Alexander E. Lipka, **Evaluation of RR-BLUP Genomic Selection Models that Incorporate Peak Genome-Wide Association Study Signals in Maize and Sorghum**, The Plant Genome, 2019, doi: 10.3835/plantgenome2018.07.0052.
3. Bertrand C.Y Collard and David J Mackill, **Marker-assisted selection: an approach for precision plant breeding in the twenty-first century**. Philosophical Transactions of the Rolyal Society B, 2007. https://doi.org/10.1098/rstb.2007.2170
4. Matthew McGowan, Jiabo Wang, Haixiao Dong, Xiaolei Liu, Yi Jia, Xianfeng Wang, Hiroyoshi Iwata, Yutao Li, Alexander E Lipka, and **Zhiwu Zhang***. **Ideas in Genomic Selection with the Potential to Transform Plant Molecular Breeding: A Review**. *Plant Breeding Review Vol 45* (eds. Irwin Goldman), John Wiley & Sons, Inc. 2021. pp. 273-320, **https://doi.org/10.1002/9781119828235.ch7**, also available as Preprint, Publication, and Google Book.

5. Spindel, JE, H Begum, D Akdemir, B Collard, E Redoña, J-L Jannink and S McCouch. **Genome-wide prediction models that incorporate de novo GWAS are a powerful new tool for tropical rice improvement**. Heredity,2016. https://www.nature.com/articles/hdy2015113.

6. VanRaden, P. M. 2008. **Efficient methods to compute genomic predictions**. J. Dairy Sci. 91: 4414–4423. DOI: 10.3168/jds.2007-0980

7. Waltram Second Ravelombola, Jun Qin, Ainong ShiID, Liana Nice, Yong Bao, Aaron Lorenz, James H. Orf, Nevin D. Young, Senyu Chen, **Genome-wide association study and genomic selection for tolerance of soybean biomass to soybean cyst nematode infestation**. PLOS One, 2020, https://doi.org/10.1371/journal.pone.0235089.

8. Yao Zhou, MI Vales, Aoxue Wang, Zhiwu Zhang. **Systematic bias of correlation coefficient may explain negative accuracy of genomic prediction**. *Briefings in Bioinformatics* 2016, https://doi.org/10.1093/bib/bbx133.

9. Zhe Zhang, Ulrike Ober, Malena Erbe, Hao Zhang, Ning Gao, Jinlong He, Jiaqi Li. **Improving the Accuracy of Whole Genome Prediction for Complex Traits Using the Results of Genome Wide Association Studies.** PLOS One. https://doi.org/10.1371/journal.pone.0093017.

10. Zhang Z, Todhunter RJ, Buckler ES, Van Vleck LD. **Technical note: Use of marker-based relationships with multiple-trait derivative-free restricted maximal likelihood**. *J Anim Sci* 2007, **85**:881–885, https://doi.org/10.2527/jas.2006-656.