

Final Project
(Option A: GS on real traits Only)
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: You can use your own data or a public dataset with both genotype and phenotype.

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.

The efforts failed in finding a universal genomic prediction model that is the best across different populations and traits of interest. It was found that the best model varies at least across genetic architecture of traits. For example, gBLUP¹⁻³ is in favor of polygenetic traits while SUPER BLUP in Mendelian traits. Compressed BLUP has advantage over other methods for traits with low heritability⁴. It is valuable to investigate the best genomic prediction model for specific populations and traits so that the optimum decision can be made to maximize breeding progress^{5,6}.

Please pay attention to the systematic bias in assessment of prediction accuracy⁷. Your investigation should be based on the comparisons on least one of the following factors.

1. GS models

The best model depends on specific population and traits. The corresponding investigation is important to select the most appropriate model for genomic prediction at the population for the specific traits.

2. Marker density

The advantage of non-gBLUP methods over gBLUP is reduced with low marker density.

3. Training sample size

As sample size changes, the ability of identifying specific impactful genomic regions is changed accordingly. This change could affect the advantages of genomic prediction models.

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. Prediction of Maize Single-Cross Performance Using RFLPs and Information from Related Hybrids. *Crop Sci* **34**, 20–25 (1994).
2. Zhang, Z., Todhunter, R. J., Buckler, E. S. & Van Vleck, L. D. Technical note: Use of marker-based relationships with multiple-trait derivative-free restricted maximal likelihood. *J Anim Sci* **85**, 881–885 (2007).
3. VanRaden, P. M. Efficient methods to compute genomic predictions. *J Dairy Sci* **91**, 4414–4423 (2008).
4. Wang, J. *et al.* Expanding the BLUP alphabet for genomic prediction adaptable to the genetic architectures of complex traits. *Heredity (Edinb)* **121**, 648–662 (2018).
5. Dong, H. *et al.* Evaluation of the Potential for Genomic Selection to Improve Spring Wheat Resistance to Fusarium Head Blight in the Pacific Northwest. *Front Plant Sci* **9**, 911 (2018).
6. Cui, Z. *et al.* Assessment of the potential for genomic selection to improve husk traits in maize. *G3: Genes, Genomes, Genetics* **10**, (2020).
7. Zhou, Y., Isabel Vales, M., Wang, A. & Zhang, Z. Systematic bias of correlation coefficient may explain negative accuracy of genomic prediction. *Briefings in Bioinformatics* (2016).