Manual for Genomic Selection using PCR and PLSR

This is a user manual on genomic selection using principal component regression (PCR) and partial least squares regression (PLSR) to reproduce the results of our analyses.

(1) Download all files and folders from github to a local working folder. The simulated data and the rice data are stored in a subfolder named 'DataIn'.

(2)Change the directory to the working folder using **setwd( )**.

(3)Use **source("sim.hat.R")** or **source("sim.cv.R")** in the R Console to implement the HAT method or the cross-validation based method to analyze the simulated data. The value of the 'model' variable in the R source files should be changed corresponding to regression methods, for example, "pcr" or "blup" can be used in **sim.hat.R**, and "pcr" or "plsr" can be used in **sim.cv.R**.

(4) Use **source("rice.hat.R")** or **source("rice.cv.R")** in the R Console to implement the HAT method or the cross-validation based method to analyze the rice data. Similarly, the value of the 'model' variable in the R source files should be changed according to the regression methods. The rice dataset includes phenotype data of four agronomic traits, 1000 metabolite traits, 24,973 expression traits, and genotype data of 1619 markers for an RIL population with 210 lines. Change the filenames in **read.csv( )** in the R source files to analyze different types of phenotype data.