Manual for iGS

iGS: an Integrated Graphical User Interface Software for Genomic Selection

Version 0.1

yao.zhou@genetics.ac.cn

Contents

Data format

To simplify the usage of this software, we only support one data format, other common formats (HapMap, vcf and plink ped format) could be easily transformed by using other tools (like blink, R).

Phenotype data

Header must be provided in phenotype data. And the first column will be taken as the individual name. Multiple phenotypes are supported and users could select the certain phenotype for analysis. Missing phenotype was indicated by NA, and will be predicted.

Taxa	phe1	phe2	phe3
Ind1	223.75855	88.291641	47.405383
Ind2	240.02525	NA	49.017183
Ind3	NA	97.474841	48.928283
Ind4	228.80225	89.714241	48.544283
Ind5	250.11015	114.00184	NA
Ind6	239.54475	114.08254	NA
Ind7	245.35535	102.22144	48.693983
Ind8	231.31385	107.11114	48.928283

Genotype data

Only numeric genotype data (GD) are supported by our software currently. And the two homozygous genotypes are coded as 0 and 2, and the heterozygous are coded as 1. No missing data is allowed for genotype data. Each column represents a SNP, and each row represents an individual. The first column contains the individual names. If there are 2 individuals with 6 SNPs could be represented as:

```
Ind1  0  0  0  1  1  1
Ind2  1  1  2  0  1  2
```

Header represents the genotype names is optional.

Covariate data

Covariate data will be treated as fixed effects in the model. No missing data is allowed. The first row is the name of covariates, and there is **no taxa column**.

Important note: The taxa in phenotype, genotype and covariate data should be the same order.

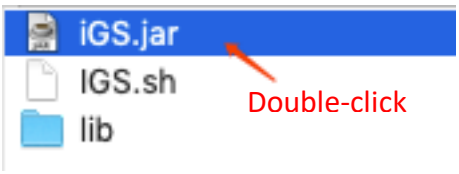
Genetic relationship matrix

For polygenic model, the design matrix should be the genetic relationship matrix estimated from other software, like PEPIS[1]. If only additive kinship was provided, the results will be the same as gBLUP model.

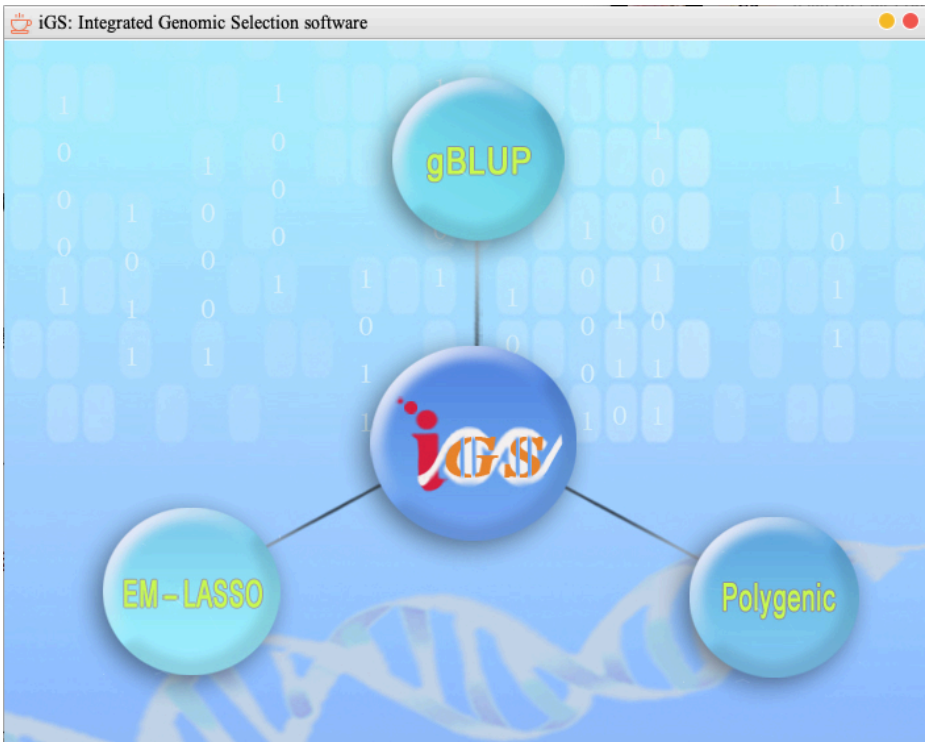
1. Zhang W.C et al., (2016) PEPIS: A Pipeline for Estimating Epistatic Effects in Quantitative Trait Locus Mapping and Genome-Wide Association Studies, *PLoS Comput Biol,* 25;12(5):e1004925

Usage

1. Double click the iGS.jar under windows and Mac operation system:

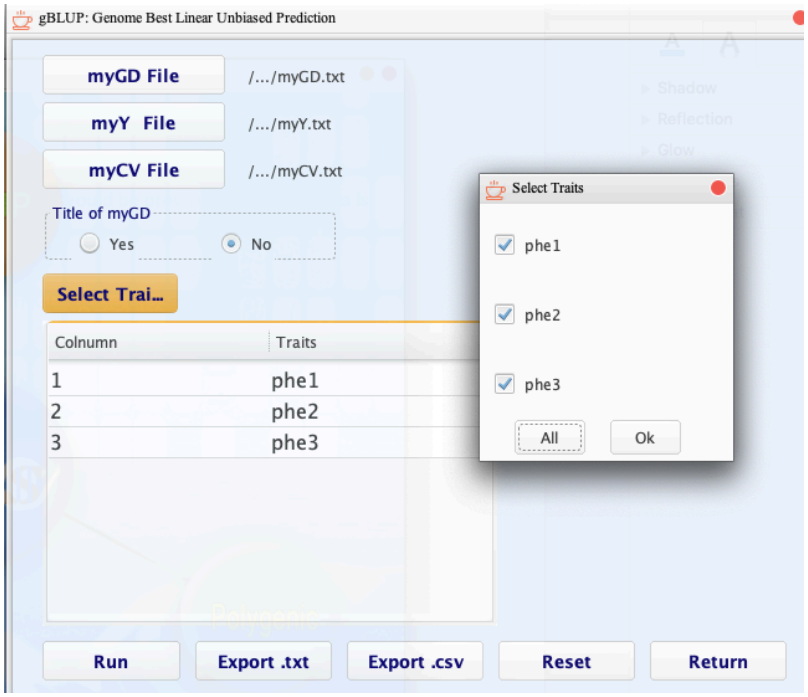


2. then choose the model from the main interface.



gBLUP

1. Load data: Genotype and phenotype data should be provided, covariate data is optional.
2. If the Genotype with header, **title of myGD** should choose Yes.
3. All phenotypes will be chosen as default, **select traits** button can be used for choosing a certain phenotype.
4. Click run and waiting for analysis;
5. When analysis finished, results could be save to txt or csv format.



EM-LASSO

1. Load data: Genotype and phenotype data should be provided, covariate data is optional. Genotype information (GM) data is never used in the EM-LASSO model for prediction.
2. If the Genotype with header, **title of myGD** should choose Yes.
3. All phenotypes will be chosen as default, **select traits** button can be used for choosing a certain phenotype.
4. Click run and waiting for analysis;
5. When analysis finished, results could be save to txt or csv format.

Polygenic model

1. Load data: design matrix and phenotype data should be provided, covariate data is optional.
2. All phenotypes will be chosen as default, **select traits** button can be used for choosing a certain phenotype.
3. REML algorithm should be selected (fs: fisher scoring; nr: new-raphson; ai: ai-reml)
4. Click run and waiting for analysis;
5. When analysis finished, results could be save to txt or csv format.