Practical demo

GCTB

A tool for Genome-wide Complex Trait Bayesian analysis

Zeng et al. 2018 *Nature Genetics*

(20 min)

Two traits as example

SLA H^2= 0.074 (0.04 – 0.106)

Height French Atlantic 2018 H^2= 0.273 (0.232 – 0.309)

Quantitative genetic models: Bayesian Mixed Models with *MCMCglmm* R package

$$y_{ijk} = \mu + B_i + \mathbf{P_j} + \mathbf{P(C)_{jk}} + \varepsilon_{ijk}$$

BLUPs

$$H2 = \frac{\sigma^2_{clone}}{\sigma^2_{clone} + \sigma^2_e}$$

Fixed factors: Block

Random factors: Population and clone (nested within population)

Show data file

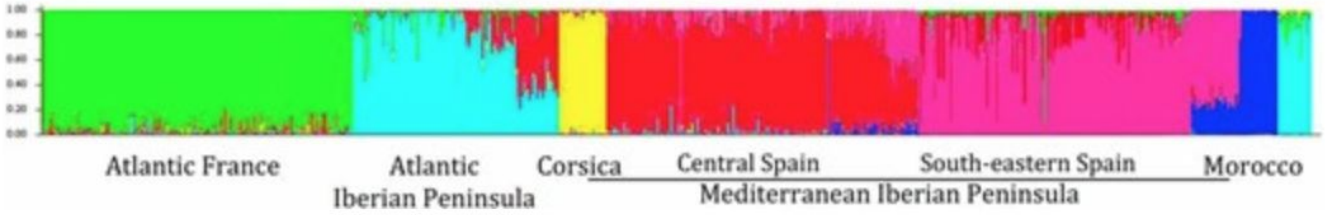
Pipeline for the estimation of polygenicity

Step 0.1 Obtain the genetic values (BLUPs) from mixed models according to your experimental design.

Step 0.2 Correction of your data by population genetic structure

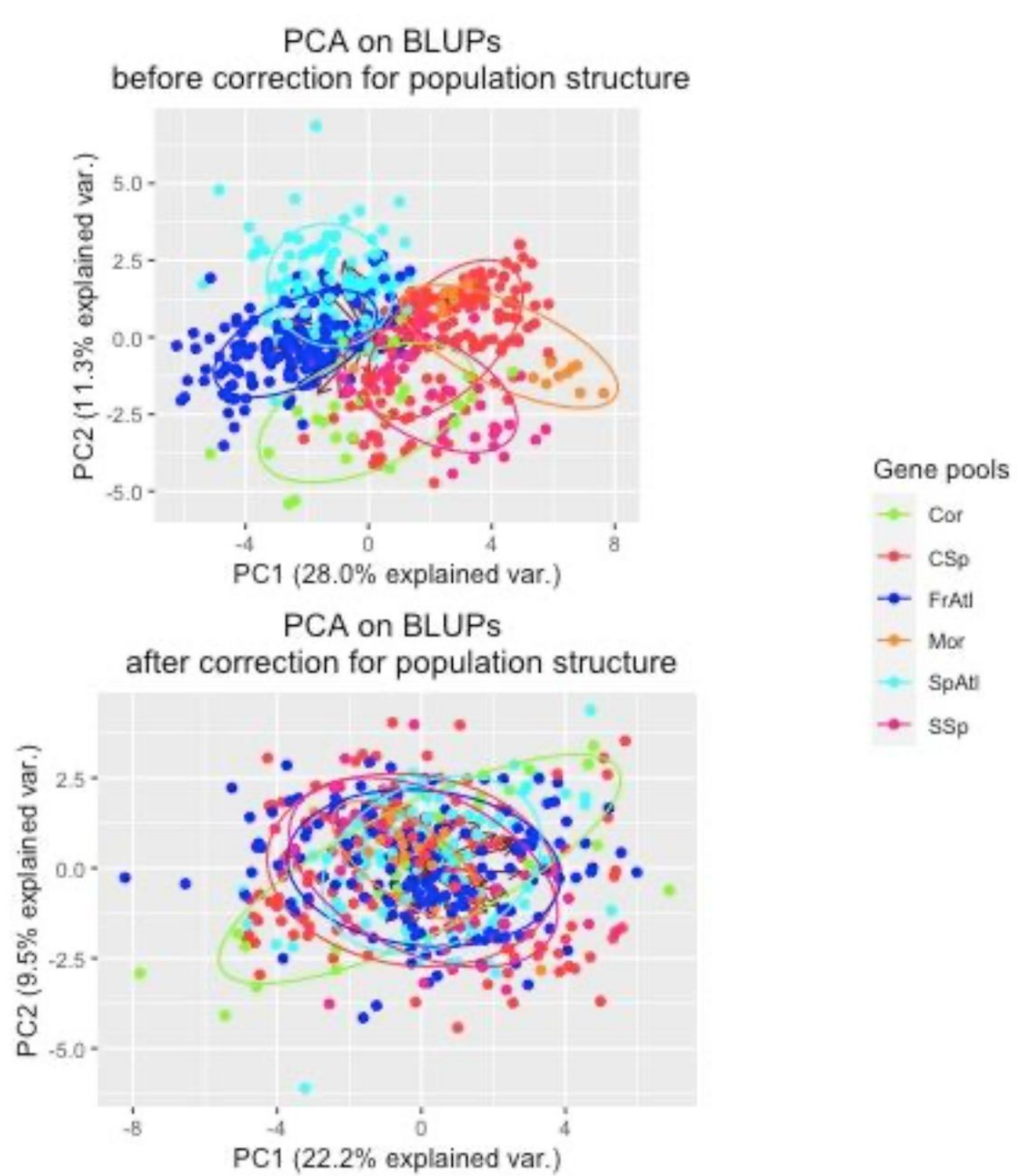
Correction of population genetic structure

Linear models relating the genetic value for each trait to the admixture coefficients for each clone (Q-scores) obtained from STRUCTURE. We extracted the normalized residuals for each trait.



STRUCTURE K=6
Jaramillo-Correa et al. 2015 Genetics

$$BLUP = a + \beta_1 Q_1 + \beta_2 Q_2 + \beta_3 Q_3 + \beta_4 Q_4 + \beta_5 Q_5 + \varepsilon$$


PCA on BLUPs
before correction for population structure
PC2 (11.3% explained var.)
5.0
2.5
0.0
-2.5
-5.0
-4
0
4
8
PC1 (28.0% explained var.)
PCA on BLUPs
after correction for population structure
PC2 (9.5% explained var.)
2.5
0.0
-2.5
-5.0
-8
-4
0
4
PC1 (22.2% explained var.)
Gene pools
Cor
CSp
FrAtl
Mor
SpAtl
SSp

Pipeline for the estimation of polygenicity

Step 0.1 Obtain the genetic values (BLUPs) from mixed models according to your experimental design.

Step 0.2 Correction of your BLUPs by population genetic structure

--- need Linux OS (or use a virtual machine if you have another OS).

Step 1 Download CGTB software from Zeng et al. 2018 *Nature Genetics.*

Download plink (software for data formatting).

Step 2 Run plink to obtain binary ped files (input of CGTB).

Step 3 Run CGTB to estimate polygenicity, *GEV* and the coefficient *S*.

--- any OS

Step 4 Formatting results in graphs and tables

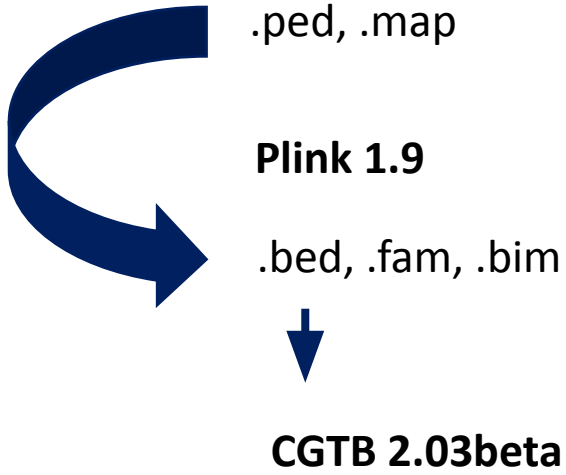
Input data

SNPs: plink pedigree files (https://www.cog-genomics.org/plink/1.9/formats#ped)

-9 missing data

.ped : the first 6 (mandatory) columns

- Family ID
- Individual ID
- Paternal ID
- Maternal ID
- Sex (1=male; 2=female; other=unknown)
- Phenotype



It must be accompanied by a .map file: a text file with no header line, and one line per variant with the following 3-4 fields:

- Chromosome code
- Variant identifier
- Position in morgans or centimorgans (optional)
- Base-pair coordinate

Binary ped files (.bed)

To store the pedigree/phenotype information in separate file (*.fam) and create an extended MAP file (*.bim) (which contains information about the allele names)

- .bed (binary file, genotype information). This is a compressed file, not able to open with a text editor!!!
- .fam (first six columns of .ped)
- .bim (extended MAP file: two extra cols = allele names)

Input data

Phenotypes:

NA missing data No header

.phen : Text file with at least 3 columns (each phenotype in a different column)

- Family ID
- Individual ID
- Phenotype 1
- Phenotype 2
- Phenotype 3…

Run CGTB

MCMC settings
--chain-length 21000 Specify the total number of iterations in MCMC, e.g. 21000 (default).
--burn-in 1000 Specify the number of iterations to be discarded, e.g. 1000 (default).
--out-freq 100 Display the intermediate results for every 100 iterations (default).
--thin 10 Output the sampled values for SNP effects and genetic architecture parameters for every 10 iterations (default). Only non-zero sampled values of SNP effects are written into a binary file.

Bayes alphabeth
--bayes S Specify the Bayesian alphabet for the analysis, Different alphabet launch different models, which differ in the prior specification for the SNP effects.

Input and output
--bfile test Input PLINK binary PED files, e.g. test.fam, test.bim and test.bed (see PLINK user manual for details).
--pheno test.phen Input phenotype data from a plain text file, e.g. test.phen.
--out test Specify output root filename.

Output data

CGTB:

- **.log**: a text file of running status, intermediate output and final results
- **.snpRes**: a text file of posterior statistics of SNP effects
- **.covRes**: a text file of posterior statistics of covariates
- **.parRes**: a text file of posterior statistics of key model parameters
- **.mcmcsamples.SnpEffects**: a binary file of MCMC samples for the SNP effects
- **.mcmcsamples.par**: a text file of MCMC samples for the parameters fitted in the model
- **.mcmcsamples.CovEffects**: a text file of MCMC samples for the covariates fitted in the model