Overview

About

GCTB is a software tool that comprises Bayesian mixed linear models for complex trait analyses using genome-wide SNPs. It was developed to simultaneously estimate the joint effects of all SNPs and the genetic architecture parameters for a complex trait, including SNP-based heritability, polygenicity and the joint distribution of effect sizes and minor allele frequencies.

Credits

Jian Zeng developed the software tool with supports from Jian Yang and Futao Zhang.

Questions and Help Requests

If you have any bug reports or questions, please send an email to Jian Zeng (j.zeng@uq.edu.au) or Jian Yang (jian.yang@uq.edu.au).

Citations

Zeng, J. et al. Widespread signatures of negative selection in the genetic architecture of human complex traits *bioRxiv* (2017)

Download

Executable Files

gctb_1.0_Linux.zip

gctb_1.0_Mac.zip

Source code

Standard version

MPI version

The MPI version implements a distributed computing strategy that allow the analysis to be scalable to very large sample sizes. A significant improvement in computing time is expected for a sample size > 10,000. The MPI version needs to be compiled on user's machine. See HowToCompile.txt in the tarball for instructions of compilation.

Update log

1. 1 Dec, 2017: first release.

Basic options

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.

Data management

--keep test.indi.list

Specify a list of individuals to be included in the analysis.

--chr 1

Include SNPs on a specific chromosome in the analysis, e.g. chromosome 1.

--extract test.snplist

Specify a list of SNPs to be included in the analysis.

--exclude test.snplist

Specify a list of SNPs to be excluded from the analysis.

--mpheno 2

If the phenotype file contains more than one trait, by default, GCTB takes the first trait for analysis (the third column of the file) unless this option is specified. For example, **--mpheno** 2 tells GCTB to take the second trait for analysis (the fourth column of the file).

--covar test.qcovar

Input quantitative covariates from a plain text file, e.g. test.qcovar. Each quantitative covariate is recognized as a continuous variable.

MCMC settings

--seed 123

Specify the seed for random number generation, e.g. 123. Note that giving the same seed value would result in exactly the same results between two runs.

--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.

--no-mcmc-bin

Suppress the output of MCMC samples of SNP effects.

Bayes alphabet

--bayes S

Specify the Bayes alphabet for the analysis, e.g. S. Different alphabet launch different models, which principally differ in the prior specification for the SNP effects. The available alphabet include C: Each SNP effect is assumed to have an i.i.d mixture prior of a normal distribution N(0, σ2) with a probability π and a point mass at zero with a probability 1-π. S: Similar to C but the variance of SNP effects is related to minor allele frequency (p) through a parameter S, i.e. σj2 = [2p(1-p)]Sσ2. N: nested BayesC. SNPs within a 0.2 Mb non-overlapping genomic region are collectively considered as a window (specify the distance by **--wind** 0.2). This nested approach speeds up the analysis by fast "jumping" over windows with zero effect. NS: nested BayesS.

--fix-pi

An option to fix the π to a constant (the value is specified by the option --pi below). The default setting is to treat π as random and estimated from the data.

--pi 0.05

A starting value for the sampling of π when it is estimated from the data, or a given value for π when it is fixed. The default value is 0.05.

--hsq 0.5

A starting value for the sampling of SNP-based heritability, which may improve the mixing of MCMC algorithm if it starts with a good estimate. The default value is 0.5.

--S 0

A starting value for the sampling of the parameter S (relationship between MAF and variance of SNP effects) in BayesS, which may improve the mixing of MCMC algorithm if it starts with a good estimate. The default value is 0.

Examples

Standard version of gctb:

```
gctb --bfile test --phen test.phen --bayes S --pi 0.1 --hsq 0.5 --chain-length 25000 --bu
rn-in 5000 --out test > test.log 2>&1
```

MPI version of gctb (when using intelMPI libraries and two nodes):

```
mpirun -f $PBS_NODEFILE -np 2 gctb_mpi --bfile test --phen test.phen --bayes S --pi 0.1 -
-hsq 0.5 --chain-length 25000 --burn-in 5000 --out test > test.log 2>&1
```

The output files include:

test.log: a text file of running status, intermediate results, posterior mean and standard deviation of key parameters

test.snpRes: a text file of posterior means of SNP effects

test.mcmcsamples.hsq: a text file of MCMC samples for SNP-based heritability

test.mcmcsamples.pi: a text file of MCMC samples for π

test.mcmcsamples.S: a text file of MCMC samples for S

test.mcmcsamples.NNZsnp: a text file of MCMC samples for the number of non-zero SNP effects

test.mcmcsamples.NNZwind: a text file of MCMC samples for the number of non-zero window effects (if BayesN or BayesNS is used)

test.mcmcsamples.FixedEffects: a text file of MCMC samples for the covariates fitted in the model

test.mcmcsamples.SnpEffects: a binary file of MCMC samples for the SNP effects