Package: mlmm.gwas (via r-universe)

August 23, 2024

Type Package

Title Pipeline for GWAS Using MLMM

Version 1.0.6

Author Fanny Bonnafous [aut], Alexandra Duhnen [aut], Louise Gody [aut], Olivier Guillaume[aut, cre], Brigitte Mangin [aut], Prune Pegot-Espagnet [aut], Vincent Segura [aut], Bjarni J. Vilhjalmsson [aut], Clement Mabire [aut], Timothee Flutre [aut]

Maintainer Clement Mabire <clement.mabire@inra.fr>

Description Pipeline for Genome-Wide Association Study using Multi-Locus Mixed Model from Segura V, Vilhjálmsson BJ et al. (2012) <doi:10.1038/ng.2314>. The pipeline include detection of associated SNPs with MLMM, model selection by lowest eBIC and p-value threshold, estimation of the effects of the SNPs in the selected model and graphical functions.

License GPL-3

Encoding UTF-8

LazyData false

Depends R (>= 3.3.0)

Imports multcompView(>= 0.1-7), multcomp(>= 1.4-8), coxme(>= 2.2-5), sommer(>= 3.2), Matrix(>= 1.2-10)

RoxygenNote 6.0.1

Suggests knitr, rmarkdown

VignetteBuilder knitr

NeedsCompilation no

Repository CRAN

Date/Publication 2019-08-05 15:20:05 UTC

Contents

`eBIC_allmodels` *Compute eBIC and BIC criteria*

Description

Compute log likelihood, BIC and eBIC.

The model with the smallest eBIC should be selected.

Usage

```
eBIC_allmodels(Y, selec_XX, KK, nb.tests, cofs = NULL, female = NULL,
  male = NULL, lambda=NULL)
```

Arguments

`Y`	A numeric named vector where the names are individuals names and the values their phenotype.
`selec_XX`	A list of length one, two or three matrices depending on the models. Use helper function `frommlmm_toebic` to get this argument.
`KK`	a list of one, two or three matrices depending on the models - additive: a n by n matrix, where n=number of individuals, with rownames()=colnames()=individual names - additive+dominance: two n by n matrices, where n=number of individuals, with rownames()=colnames()=individual names - female+male: a n.female by n.female matrix, with rownames()=colnames()=female names and a n.male by n.male matrix, with rownames()=colnames()=male names - female+male+interaction: the same two matrices as the model female+male and a n by n matrix, where n=number of individuals, with rownames()=colnames()=individual names
`nb.tests`	number of computed tests (total number of SNPs)
`cofs`	A n by q matrix, where n=number of individuals, q=number of fixed effect,
`female`	A factor of levels female names and length n, only for the last two models

male	A factor of levels male names and length n, only for the last two models
lambda	penalty used in the computation of the eBIC; if NULL, the default will be 1 - 1/(2k) with L=n^k where L=total number of SNPs (see function "lambda.calc")

Value

A matrix with a line for each mlmm step and 4 columns : BIC, ajout, eBIC_0.5 and LogL.

Examples

```
### Additive model ###
## Not run:
data("mlmm.gwas.AD")

XX = list(Xa)
KK = list(K.add)

# GWAS
res_mlmm <- mlmm_allmodels(floweringDateAD, XX, KK)
manhattan.plot(res_mlmm)

# Model selection
sel_XX <- frommlmm_toebic(XX, res_mlmm)
res.eBIC <- eBIC_allmodels(floweringDateAD, sel_XX, KK, ncol(Xa))

# Effects estimations with the selected model
sel_XXclass <- fromeBICtoEstimation(sel_XX, res.eBIC)
eff.estimations <- Estimation_allmodels(floweringDateAD, sel_XXclass, KK)
genotypes.boxplot(Xa, floweringDateAD, effects = eff.estimations)

## End(Not run)

### Additive + dominance model
## Not run:
data("mlmm.gwas.AD")

XX = list(Xa, Xd)
KK = list(K.add, K.dom)

# GWAS
res_mlmm <- mlmm_allmodels(floweringDateAD, XX, KK)
manhattan.plot(res_mlmm)

# Model selection
sel_XX <- frommlmm_toebic(XX, res_mlmm)
res.eBIC <- eBIC_allmodels(floweringDateAD, sel_XX, KK, ncol(Xa))
#the selected model is the null model

## End(Not run)


### Female+Male model
```

```
## Not run:
data("mlmm.gwas.FMI")

XX = list(Xf, Xm)
KK = list(K.female, K.male)

# GWAS
res_mlmm <- mlmm_allmodels(floweringDateFMI, XX, KK, female = female, male = male)
manhattan.plot(res_mlmm)

# Model selection
sel_XX <- frommlmm_toebic(XX, res_mlmm)
res.eBIC <- eBIC_allmodels(floweringDateFMI, sel_XX, KK, ncol(Xf), female = female, male = male)
#the selected model is the null model

## End(Not run)



### Female+Male+Interaction model
## Not run:
data("mlmm.gwas.FMI")

XX = list(Xf, Xm, Xfm)
KK = list(K.female, K.male, K.hybrid)

# GWAS
res_mlmm <- mlmm_allmodels(floweringDateFMI, XX, KK, female = female, male = male)
manhattan.plot(res_mlmm)

# Model selection
sel_XX <- frommlmm_toebic(XX, res_mlmm)
res.eBIC <- eBIC_allmodels(floweringDateFMI, sel_XX, KK, ncol(Xf), female = female, male = male)
#the selected model is the null model

## End(Not run)
```

`Estimation_allmodels` *Compute estimated effects*

Description

Estimate the effect of selected SNPs.

Usage

```
Estimation_allmodels(Y, selec_XXclass, KK, cofs = NULL, female = NULL,
  male = NULL)
```

Arguments

Y	A numeric named vector where the names are individuals names and the values their phenotype. The names of Y will be matched to the row names of X.
selec_XXclass	A n by mk data.frame of factors with rownames()=individual names, and colnames()=mk selected SNP names additive+dominance: three levels factor female+male+interaction: four levels factor Use function fromeBICtoEstimation to get this armument.
KK	a list of one, two or three matrices depending on the models - additive: a n by n matrix, where n=number of individuals, with rownames()=colnames()=individual names - additive+dominance: two n by n matrices, where n=number of individuals, with rownames()=colnames()=individual names - female+male: a n.female by n.female matrix, with rownames()=colnames()=female names and a n.male by n.male matrix, with rownames()=colnames()=male names - female+male+interaction: the same two matrices as the model female+male and a n by n matrix, where n=number of individuals, with rownames()=colnames()=individual names
cofs	A n by q matrix, where n=number of individuals, q=number of fixed effect, with rownames()=individual names and with column names, forbidden head of column names for this matrix "eff1_" and usage of special characters as "*","/","&"
female	A factor of levels female names and length n, only for the last two models
male	A factor of levels male names and length n, only for the last two models

Value

A dataframe with 3 colum: BLUE, Tukey.Class and Frequency. The firt line name is "mu", the names of the other lines are in the form markername_allele.

Examples

```
### Additive model ###
## Not run:
data("mlmm.gwas.AD")

XX = list(Xa)
KK = list(K.add)

# GWAS
res_mlmm <- mlmm_allmodels(floweringDateAD, XX, KK)
manhattan.plot(res_mlmm)

# Model selection
sel_XX <- frommlmm_toebic(XX, res_mlmm)
res.eBIC <- eBIC_allmodels(floweringDateAD, sel_XX, KK, ncol(Xa))

# Effects estimations with the selected model
sel_XXclass <- fromeBICtoEstimation(sel_XX, res.eBIC)
eff.estimations <- Estimation_allmodels(floweringDateAD, sel_XXclass, KK)
```

```
genotypes.boxplot(Xa, floweringDateAD, effects = eff.estimations)

## End(Not run)

### Additive + dominance model
## Not run:
data("mlmm.gwas.AD")

XX = list(Xa, Xd)
KK = list(K.add, K.dom)

# GWAS
res_mlmm <- mlmm_allmodels(floweringDateAD, XX, KK)
manhattan.plot(res_mlmm)

# Model selection
sel_XX <- frommlmm_toebic(XX, res_mlmm)
res.eBIC <- eBIC_allmodels(floweringDateAD, sel_XX, KK, ncol(Xa))
#the selected model is the null model

## End(Not run)


### Female+Male model
## Not run:
data("mlmm.gwas.FMI")

XX = list(Xf, Xm)
KK = list(K.female, K.male)

# GWAS
res_mlmm <- mlmm_allmodels(floweringDateFMI, XX, KK, female = female, male = male)
manhattan.plot(res_mlmm)

# Model selection
sel_XX <- frommlmm_toebic(XX, res_mlmm)
res.eBIC <- eBIC_allmodels(floweringDateFMI, sel_XX, KK, ncol(Xf), female = female, male = male)
#the selected model is the null model

## End(Not run)



### Female+Male+Interaction model
## Not run:
data("mlmm.gwas.FMI")

XX = list(Xf, Xm, Xfm)
KK = list(K.female, K.male, K.hybrid)

# GWAS
res_mlmm <- mlmm_allmodels(floweringDateFMI, XX, KK, female = female, male = male)
manhattan.plot(res_mlmm)
```

```
# Model selection
sel_XX <- frommlmm_toebic(XX, res_mlmm)
res.eBIC <- eBIC_allmodels(floweringDateFMI, sel_XX, KK, ncol(Xf), female = female, male = male)
#the selected model is the null model

## End(Not run)
```

`fromeBICtoEstimation` *Helper function that create the selec_XXclass argument of the Estimation_allmodels function*

Description

Function that create the selec_XXclass argument of the `Estimation_allmodels` function from the output of the `eBIC_allmodels` function or from the output of the `threshold_allmodels` function.

Usage

```
fromeBICtoEstimation(XX, res.eBIC, res.threshold)
```

Arguments

`XX`	A list of length one, two or three matrices depending on the model. Matrices are n by m matrix, where n=number of individuals, m=number of SNPs, with rownames(X)=individual names, and colnames(X)=SNP names. - additive: a single matrix - additive+dominance: two matrices - female+male: two matrices with the female one first - female+male+interaction: three matrices with the female first, the male then the interaction
`res.eBIC`	output of the `eBIC_allmodels` function
`res.threshold`	output of the `threshold_allmodels` function

See Also

`eBIC_allmodels` `Estimation_allmodels`

Examples

```
### Additive model ###
## Not run:
data("mlmm.gwas.AD")

XX = list(Xa)
KK = list(K.add)
```

```
# GWAS
res_mlmm <- mlmm_allmodels(floweringDateAD, XX, KK)
manhattan.plot(res_mlmm)

# Model selection
sel_XX <- frommlmm_toebic(XX, res_mlmm)
res.eBIC <- eBIC_allmodels(floweringDateAD, sel_XX, KK, ncol(Xa))

# Effects estimations with the selected model
sel_XXclass <- fromeBICtoEstimation(sel_XX, res.eBIC)
eff.estimations <- Estimation_allmodels(floweringDateAD, sel_XXclass, KK)
genotypes.boxplot(Xa, floweringDateAD, effects = eff.estimations)

## End(Not run)

### Additive + dominance model
## Not run:
data("mlmm.gwas.AD")

XX = list(Xa, Xd)
KK = list(K.add, K.dom)

# GWAS
res_mlmm <- mlmm_allmodels(floweringDateAD, XX, KK)
manhattan.plot(res_mlmm)

# Model selection
sel_XX <- frommlmm_toebic(XX, res_mlmm)
res.eBIC <- eBIC_allmodels(floweringDateAD, sel_XX, KK, ncol(Xa))
#the selected model is the null model

## End(Not run)


### Female+Male model
## Not run:
data("mlmm.gwas.FMI")

XX = list(Xf, Xm)
KK = list(K.female, K.male)

# GWAS
res_mlmm <- mlmm_allmodels(floweringDateFMI, XX, KK, female = female, male = male)
manhattan.plot(res_mlmm)

# Model selection
sel_XX <- frommlmm_toebic(XX, res_mlmm)
res.eBIC <- eBIC_allmodels(floweringDateFMI, sel_XX, KK, ncol(Xf), female = female, male = male)
#the selected model is the null model

## End(Not run)
```

```
### Female+Male+Interaction model
## Not run:
data("mlmm.gwas.FMI")

XX = list(Xf, Xm, Xfm)
KK = list(K.female, K.male, K.hybrid)

# GWAS
res_mlmm <- mlmm_allmodels(floweringDateFMI, XX, KK, female = female, male = male)
manhattan.plot(res_mlmm)

# Model selection
sel_XX <- frommlmm_toebic(XX, res_mlmm)
res.eBIC <- eBIC_allmodels(floweringDateFMI, sel_XX, KK, ncol(Xf), female = female, male = male)
#the selected model is the null model

## End(Not run)
```

`frommlmm_toebic` *Helper that create the selec_XX argument of eBIC_allmodels()*

Description

Helper function that create the selec_XX argument of `eBIC_allmodels` from the output of `mlmm_allmodels`.

Usage

```
frommlmm_toebic(XX, res.mlmm)
```

Arguments

`XX`	A list of length one, two or three matrices depending on the model. Matrices are n by m matrix, where n=number of individuals, m=number of SNPs, with rownames(X)=individual names, and colnames(X)=SNP names. Use the same XX you used with the `mlmm_allmodels` function
`res.mlmm`	Output from the `mlmm_allmodels` function

See Also

`mlmm_allmodels` `eBIC_allmodels`

Examples

```
### Additive model ###
## Not run:
data("mlmm.gwas.AD")

XX = list(Xa)
```

```
KK = list(K.add)

# GWAS
res_mlmm <- mlmm_allmodels(floweringDateAD, XX, KK)
manhattan.plot(res_mlmm)

# Model selection
sel_XX <- frommlmm_toebic(XX, res_mlmm)
res.eBIC <- eBIC_allmodels(floweringDateAD, sel_XX, KK, ncol(Xa))

# Effects estimations with the selected model
sel_XXclass <- fromeBICtoEstimation(sel_XX, res.eBIC)
eff.estimations <- Estimation_allmodels(floweringDateAD, sel_XXclass, KK)
genotypes.boxplot(Xa, floweringDateAD, effects = eff.estimations)

## End(Not run)

### Additive + dominance model
## Not run:
data("mlmm.gwas.AD")

XX = list(Xa, Xd)
KK = list(K.add, K.dom)

# GWAS
res_mlmm <- mlmm_allmodels(floweringDateAD, XX, KK)
manhattan.plot(res_mlmm)

# Model selection
sel_XX <- frommlmm_toebic(XX, res_mlmm)
res.eBIC <- eBIC_allmodels(floweringDateAD, sel_XX, KK, ncol(Xa))
#the selected model is the null model

## End(Not run)


### Female+Male model
## Not run:
data("mlmm.gwas.FMI")

XX = list(Xf, Xm)
KK = list(K.female, K.male)

# GWAS
res_mlmm <- mlmm_allmodels(floweringDateFMI, XX, KK, female = female, male = male)
manhattan.plot(res_mlmm)

# Model selection
sel_XX <- frommlmm_toebic(XX, res_mlmm)
res.eBIC <- eBIC_allmodels(floweringDateFMI, sel_XX, KK, ncol(Xf), female = female, male = male)
#the selected model is the null model

## End(Not run)
```

```
### Female+Male+Interaction model
## Not run:
data("mlmm.gwas.FMI")

XX = list(Xf, Xm, Xfm)
KK = list(K.female, K.male, K.hybrid)

# GWAS
res_mlmm <- mlmm_allmodels(floweringDateFMI, XX, KK, female = female, male = male)
manhattan.plot(res_mlmm)

# Model selection
sel_XX <- frommlmm_toebic(XX, res_mlmm)
res.eBIC <- eBIC_allmodels(floweringDateFMI, sel_XX, KK, ncol(Xf), female = female, male = male)
#the selected model is the null model

## End(Not run)
```

`genotypes.boxplot` *Boxplots representation of the distributions of phenotypes according to allelic classes*

Description

For each allele class of a given loci, display as boxplot the distributions of the phenotypes of individuals with this allele class.

For instance, it can be used as a simple representation of the effects of one QTL.

Usage

```
genotypes.boxplot(X, Y, markers = "all", effects = NULL,
  genotypes = c("00", "01|10", "11"), tukeyTextCol = NA, tukeyTextCex = 1,
  tukeyCol = c("#2ecc71", "#3498db", "#9b59b6", "#6c7a89", "#f2ca27",
  "#e67e22", "#e74c3c", "#c08d57"), tukeyPch = c(1, 3, 2, 4:8),
  tukeyCex = 1, ...)
```

Arguments

`X`	A matrix where rownames are individuals names, colnames are markers names, and values are genotypes. Genotypes are encoded as allelic dosage (0, 1, 2) or as any numeric values as long as the smallest and highest values correspond to homozygous and the mean of these smallest and highest values to heterozygous. Other values (imputated genotypes) will be rounded to the nearest.
`Y`	A numeric named vector where the names are individuals names and the values their phenotype. The names of Y will be matched to the row names of X.

markers	A vector of names of markers, a plot will be drawn for each of them. "all" is a special value meaning a plot will be drawn for all markers in the estimations object, or in the matrix X if the estimations object is not provided.
effects	A GWAS.EFFECTS oject, created with `Estimation_allmodels` function.
genotypes	A length 3 string vector, used as labels for the genotypes.
tukeyTextCol	Colors of the letters of the Tukey classes.
tukeyTextCex	Size of the letters of the Tukey classes.
tukeyCol	Color of the symbols of the Tukey classes.
tukeyPch	Symbols of the Tukey classes.
tukeyCex	Size of the symbols of the Tukey classes.
...	Additional arguments are passed to the `boxplot` function.

Details

A plot is drawn for each of marker of the markers vector.

In each of thes plots, a boxplot is drawn for each allelic classes. Theses boxplots represent the distribution of the phenotypes of individuals with these allelic classes.

If the effects parameter is not NULL, the Tukey classes of the effects of markers will be represented as a symbol and/or a letter in the boxplot. The ordinates of these symbols is the average of the phenotype of individuals with the allele.

Examples

```
### Additive model ###
## Not run:
data("mlmm.gwas.AD")

XX = list(Xa)
KK = list(K.add)

# GWAS
res_mlmm <- mlmm_allmodels(floweringDateAD, XX, KK)
manhattan.plot(res_mlmm)

# Model selection
sel_XX <- frommlmm_toebic(XX, res_mlmm)
res.eBIC <- eBIC_allmodels(floweringDateAD, sel_XX, KK, ncol(Xa))

# Effects estimations with the selected model
sel_XXclass <- fromeBICtoEstimation(XX, res.eBIC)
eff.estimations <- Estimation_allmodels(floweringDateAD, sel_XXclass, KK)
genotypes.boxplot(Xa, floweringDateAD, effects = eff.estimations)

## End(Not run)
```

`manhattan.plot` *Manhattan plot*

Description

Draw a Manhattan plot of the association p-values of the markers.

Usage

```
manhattan.plot(res.mlmm, map = NULL, steps = 1, hideCofactors = FALSE,
  chrToPlot = "all", unit = "cM", ...)
```

Arguments

`res.mlmm`	Output object from `mlmm_allmodels`.
`map`	Dataframe with 3 columns : markers names, chromosome or scaffold names and position (any unit is allowed: cM, Mpb etc.).
`steps`	An integer. The iteration number of the forward approach. If a vector of length >= 2 is passed, several plots will be drawn. By default, only step 1 is drawn.
`hideCofactors`	If TRUE, the coffactors (fixed effects) won't be drawn
`chrToPlot`	Names of the chromosomes or scaffolds to plot. Use this if you want to zoom on a particular chromosome.
`unit`	Unit of the positions in the map.
`...`	additional arguments can be passed to the plot function.

Details

Draws a manhattan plot ie. plot -log(p-value) vs marker position

If a map is passed, markers position will be used as x axis. If not, the indices of markers inside the res.mlmm object will be used instead

If there are cofactors (as in all but the first step of the forward approch), the cofactors markers will be plotted too (symbol: star).

If a map is passed, markers not in the map or in the map but not assigned to a chromosome will be assigned to a virtual chromosome 0.

Markers in the map, assigned to a chromosome, but with missing position, will be ploted at the end of the chromosome.

See Also

`mlmm_allmodels`

Examples

```
### Additive model ###
## Not run:
data("mlmm.gwas.AD")

XX = list(Xa)
KK = list(K.add)

# GWAS
res_mlmm <- mlmm_allmodels(floweringDateAD, XX, KK)
manhattan.plot(res_mlmm)

# Model selection
sel_XX <- frommlmm_toebic(XX, res_mlmm)
res.eBIC <- eBIC_allmodels(floweringDateAD, sel_XX, KK, ncol(Xa))

# Effects estimations with the selected model
sel_XXclass <- fromeBICtoEstimation(sel_XX, res.eBIC)
eff.estimations <- Estimation_allmodels(floweringDateAD, sel_XXclass, KK)
genotypes.boxplot(Xa, floweringDateAD, effects = eff.estimations)

## End(Not run)

### Additive + dominance model
## Not run:
data("mlmm.gwas.AD")

XX = list(Xa, Xd)
KK = list(K.add, K.dom)

# GWAS
res_mlmm <- mlmm_allmodels(floweringDateAD, XX, KK)
manhattan.plot(res_mlmm)

# Model selection
sel_XX <- frommlmm_toebic(XX, res_mlmm)
res.eBIC <- eBIC_allmodels(floweringDateAD, sel_XX, KK, ncol(Xa))
#the selected model is the null model

## End(Not run)


### Female+Male model
## Not run:
data("mlmm.gwas.FMI")

XX = list(Xf, Xm)
KK = list(K.female, K.male)

# GWAS
res_mlmm <- mlmm_allmodels(floweringDateFMI, XX, KK, female = female, male = male)
manhattan.plot(res_mlmm)
```

```
# Model selection
sel_XX <- frommlmm_toebic(XX, res_mlmm)
res.eBIC <- eBIC_allmodels(floweringDateFMI, sel_XX, KK, ncol(Xf), female = female, male = male)
#the selected model is the null model

## End(Not run)



### Female+Male+Interaction model
## Not run:
data("mlmm.gwas.FMI")

XX = list(Xf, Xm, Xfm)
KK = list(K.female, K.male, K.hybrid)

# GWAS
res_mlmm <- mlmm_allmodels(floweringDateFMI, XX, KK, female = female, male = male)
manhattan.plot(res_mlmm)

# Model selection
sel_XX <- frommlmm_toebic(XX, res_mlmm)
res.eBIC <- eBIC_allmodels(floweringDateFMI, sel_XX, KK, ncol(Xf), female = female, male = male)
#the selected model is the null model

## End(Not run)
```

---

mlmm.gwas *mlmm.gwas*

Description

Pipeline for GWAS using Multi Locus Mixed Model (MLMM).

Details

This is a fork of the MLMM / MultLocMixMod package by Vincent Segura and Bjarni J. Vilhjalmsson.

The main differencies from the original package are:

- abandon of the multi-Bonferroni model selection
- abandon of the backward model search
- eBIC modified to be adapted to the rate between number of individuals and number of markers.
- function added to select significant SNP according to a p-value threshold at each mlmm step
- new models supported: additive+dominance, male+female and male+female+interaction. These models are described in Bonnafous et al. (2017).
- graphical functions: a new Manhattan plot and a boxplot representation of markers effects.

A vignette presents the usage of this package with the additive model.

References

Bonnafous, F., Fievet, G., Blanchet, N. et al. Theor Appl Genet (2018) 131: 319.

Segura V, Vilhjálmsson BJ, Platt A, Korte A, Seren Uuml, Long Q, et al. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nat Genet. 2012;44:825–830.

`mlmm.gwas.AD` *Dataset for examples with the mlmm.gwas package, additive and additive+dominance models*

Description

- Species: Helianthus annuus
- Individuals: 125
- Markers: 500

Usage

```
data(mlmm.gwas.AD)
```

Format

- floweringDateAD: a named numeric of length 444.
- Xa: a 110x500 numeric matrix
- Xd: a 110x500 numeric matrix
- K.add: a 110x110 numeric matrix
- K.dom: a 110x110 numeric matrix

Details

Variables:

- floweringDateAD: flowering dates in °C.day, time since sowing.
- Xa: genotype matrix (additive)
- Xd: genotype matrix (dominance)
- K.add: "kinship" matrix (additive)
- K.dom: "kinship" matrix (dominance)

Source

Bonnafous & al. (2017)

`mlmm.gwas.FMI` *Dataset for examples with the mlmm.gwas package, male+female and male+female+interaction models*

Description

- Species: Helianthus annuus
- Individuals: 303
- Markers: 500

Usage

```
data(mlmm.gwas.FMI)
```

Format

- floweringDateFMI: a named numeric vector of length 303.
- female: a factor of length 303
- male: a factor of length 303
- hybrid: a factor of length 303
- Xf: a 303x500 numeric matrix
- Xm: a 303x500 numeric matrix
- Xfm: a 303x500 numeric matrix
- K.female: 36x36 numeric matrix
- K.male: 36x36 numeric matrix
- K.hybrid: 303x303 numeric matrix

Details

Variables:

- floweringDateFMI: flowering dates in °C.day, time since sowing.
- female: names of the female parent of the individuals
- male: names of the male parent of the individuals
- hybrid: names of the hybrids (name of female and male)
- Xf: female genotype matrix (additive)
- Xm: male genotype matrix (dominance)
- Xfm: female-male interaction genotype matrix (dominance)
- K.female: female "kinship" matrix (additive)
- K.male: male "kinship" matrix (dominance)
- K.hybrid: hybrid "kinship" matrix (dominance)

Source

Bonnafous & al. (2017)

`mlmm_allmodels` *Multi-Locus Mixed-Model*

Description

Carry GWAS correcting for population structure while including cofactors through a forward regression approach.

possible models: additive, additive+dominance, female+male, female+male+interaction

For additive model, look at the example below or at this vignette. For other models, read Bonnafous et al. (2017).

Usage

```
mlmm_allmodels(Y, XX, KK, nbchunks = 2, maxsteps = 20, cofs = NULL,
  female = NULL, male = NULL, threshold=NULL)
```

Arguments

`Y`	A numeric named vector where the names are individuals' names and the values their phenotype. The names of Y will be matched to the row names of X.
`XX`	A list of length one, two or three matrices depending on the model. Matrices are n by m matrix, where n=number of individuals, m=number of SNPs, with rownames(X)=individual names, and colnames(X)=SNP names. - additive: a single matrix - additive+dominance: two matrices - female+male: two matrices with the female one first - female+male+interaction: three matrices with the female first, the male then the interaction
`KK`	a list of one, two or three matrices depending on the models - additive: a n by n matrix, where n=number of individuals, with rownames()=colnames()=individual names - additive+dominance: two n by n matrices, where n=number of individuals, with rownames()=colnames()=individual names - female+male: a n.female by n.female matrix, with rownames()=colnames()=female names and a n.male by n.male matrix, with rownames()=colnames()=male names - female+male+interaction: the same two matrices as the model female+male and a n by n matrix, where n=number of individuals, with rownames()=colnames()=individual names
`nbchunks`	An integer defining the number of chunks of matrices to run the analysis, allows to decrease the memory usage. minimum=2, increase it if you do not have enough memory
`maxsteps`	An integer >= 3. Maximum number of steps desired in the forward approach. The forward approach breaks automatically once the pseudo-heritability is close to 0, however to avoid doing too many steps in case the pseudo-heritability does not reach a value close to 0, this parameter is also used.

cofs	A n by q matrix, where n=number of individuals, q=number of fixed effect, with rownames()=individual names and with column names, forbidden head of column names for this matrix "eff1_" and usage of special characters as "*","/","&"
female	A factor of levels female names and length n, only for the last two models
male	A factor of levels male names and length n, only for the last two models
threshold	a value to declare the significant p value. The default value is Bonferroni 0.05

Details

Each of the data arguments must be sorted in the same way, according to the individual name.

Value

a list with one element per step of the forward approach. Each element of this list is a named vector of p-values, the names are the names of the markers, with "selec_" as prefix for the markers used as fixed effects.

See Also

manhattan.plot

Examples

```
### Data for additive and additive+dominance models

data("mlmm.gwas.AD")

### Additive model ###

XX = list(Xa)
KK = list(K.add)

# GWAS
res_mlmm <- mlmm_allmodels(floweringDateAD, XX, KK)
manhattan.plot(res_mlmm)

# Model selection
sel_XX <- frommlmm_toebic(XX, res_mlmm)
res.eBIC <- eBIC_allmodels(floweringDateAD, sel_XX, KK, ncol(Xa))

# Marker selection
res.threshold <- threshold_allmodels(threshold=NULL, res_mlmm)

# Effects estimations with the selected model
sel_XXclass <- fromeBICtoEstimation(sel_XX, res.eBIC, res.threshold)
eff.estimations <- Estimation_allmodels(floweringDateAD, sel_XXclass, KK)
genotypes.boxplot(Xa, floweringDateAD, effects = eff.estimations)

## Not run:
### Additive + dominance model
```

```
XX = list(Xa, Xd)
KK = list(K.add, K.dom)

# GWAS
res_mlmm <- mlmm_allmodels(floweringDateAD, XX, KK)
manhattan.plot(res_mlmm)

# Model selection
sel_XX <- frommlmm_toebic(XX, res_mlmm)
res.eBIC <- eBIC_allmodels(floweringDateAD, sel_XX, KK, ncol(Xa))
#the selected model is the null model

# Marker selection
res.threshold <- threshold_allmodels(threshold=NULL, res_mlmm)

### Data for female+male and female+male+interaction

data("mlmm.gwas.FMI")

### Female+Male model

XX = list(Xf, Xm)
KK = list(K.female, K.male)

# GWAS
res_mlmm <- mlmm_allmodels(floweringDateFMI, XX, KK, female = female, male = male)
manhattan.plot(res_mlmm)

# Model selection
sel_XX <- frommlmm_toebic(XX, res_mlmm)
res.eBIC <- eBIC_allmodels(floweringDateFMI, sel_XX, KK, ncol(Xf), female = female, male = male)
#the selected model is the null model

# Marker selection
res.threshold <- threshold_allmodels(threshold=NULL, res_mlmm)


### Female+Male+Interaction model

XX = list(Xf, Xm, Xfm)
KK = list(K.female, K.male, K.hybrid)

# GWAS
res_mlmm <- mlmm_allmodels(floweringDateFMI, XX, KK, female = female, male = male)
manhattan.plot(res_mlmm)

# Model selection
sel_XX <- frommlmm_toebic(XX, res_mlmm)
res.eBIC <- eBIC_allmodels(floweringDateFMI, sel_XX, KK, ncol(Xf), female = female, male = male)
#the selected model is the null model

# Marker selection
res.threshold <- threshold_allmodels(threshold=NULL, res_mlmm)
```

```
## End(Not run)
```

`run_entire_gwas_pipeline` *MLMM, model selection and effects estimation*

Description

Internaly run functions of the mlmm.gwas package:

- `mlmm_allmodels` (GWAS)
- `frommlmm_toebic`
- `eBIC_allmodels` (model selection)
- `threshold_allmodels`
- `fromeBICtoEstimation`
- `Estimation_allmodels` (effects estimation)

Usage

```
run_entire_gwas_pipeline(Y, XX, KK, nbchunks = 2, maxsteps = 20,
  cofs = NULL, female = NULL, male = NULL,threshold=NULL,lambda=NULL)
```

Arguments

`Y`	A numeric named vector where the names are individuals' names and the values their phenotype. The names of Y will be matched to the row names of X.
`XX`	A list of length one, two or three matrices depending on the model. Matrices are n by m matrix, where n=number of individuals, m=number of SNPs, with rownames(X)=individual names, and colnames(X)=SNP names. - additive: a single matrix - additive+dominance: two matrices - female+male: two matrices with the female one first - female+male+interaction: three matrices with the female first, the male then the interaction
`KK`	a list of one, two or three matrices depending on the models - additive: a n by n matrix, where n=number of individuals, with rownames()=colnames()=individual names - additive+dominance: two n by n matrices, where n=number of individuals, with rownames()=colnames()=individual names - female+male: a n.female by n.female matrix, with rownames()=colnames()=female names and a n.male by n.male matrix, with rownames()=colnames()=male names - female+male+interaction: the same two matrices as the model female+male and a n by n matrix, where n=number of individuals, with rownames()=colnames()=individual names

nbchunks	An integer defining the number of chunks of matrices to run the analysis, allows to decrease the memory usage. minimum=2, increase it if you do not have enough memory
maxsteps	An integer >= 3. Maximum number of steps desired in the forward approach. The forward approach breaks automatically once the pseudo-heritability is close to 0, however to avoid doing too many steps in case the pseudo-heritability does not reach a value close to 0, this parameter is also used.
cofs	A n by q matrix, where n=number of individuals, q=number of fixed effect, with rownames()=individual names and with column names, forbidden head of column names for this matrix "eff1_" and usage of special characters as "*","/","&"
female	A factor of levels female names and length n, only for the last two models
male	A factor of levels male names and length n, only for the last two models
threshold	a value to declare the significant p value. The default value is Bonferroni 0.05
lambda	penalty used in the computation of the eBIC; if NULL, the default will be 1 - 1/(2k) with L=n^k where L=total number of SNPs (see function "lambda.calc")

Value

A named list with 2 or 3 elements:

- pval: the return value of `mlmm_allmodels`
- eBic: the return value of `eBIC_allmodels`
- threshold: the return value of `threshold_allmodels`
- effects: the return value of `Estimation_allmodels`, only if there is at least one marker in the model selected by lowest eBIC.

Examples

```
data("mlmm.gwas.AD")
results <- run_entire_gwas_pipeline(floweringDateAD, list(Xa), list(K.add))
```

---

threshold_allmodels *Select significant marker thanks to a p-value threshold*

Description

Select significant marker at each mlmm step according to a threshold.

Usage

```
threshold_allmodels(threshold=NULL, res_mlmm)
```

Arguments

threshold	a value to declare the significant p value. The default value is Bonferroni 0.05
res_mlmm	a list of p-value for each mlmm step. Use helper function `mlmm_allmodels` to get this argument.

Value

A matrix with a line for significant SNP at each mlmm step (according to the defined threshold) and 3 columns : SNP, p-value, step

Examples

```
### Additive model ###
## Not run:
data("mlmm.gwas.AD")

XX = list(Xa)
KK = list(K.add)

# GWAS
res_mlmm <- mlmm_allmodels(floweringDateAD, XX, KK)
manhattan.plot(res_mlmm)

# Model and Marker selection
sel_XX <- frommlmm_toebic(XX, res_mlmm)
res.eBIC <- eBIC_allmodels(floweringDateAD, sel_XX, KK, ncol(Xa))
res.threshold <- threshold_allmodels(threshold, res_mlmm)

# Effects estimations with the selected model
sel_XXclass <- fromeBICtoEstimation(sel_XX, res.eBIC, res.threshold)
eff.estimations <- Estimation_allmodels(floweringDateAD, sel_XXclass, KK)
genotypes.boxplot(Xa, floweringDateAD, effects = eff.estimations)

## End(Not run)

### Additive + dominance model
## Not run:
data("mlmm.gwas.AD")

XX = list(Xa, Xd)
KK = list(K.add, K.dom)

# GWAS
res_mlmm <- mlmm_allmodels(floweringDateAD, XX, KK)
manhattan.plot(res_mlmm)

# Marker selection
res.threshold <- threshold_allmodels(threshold, res_mlmm)

## End(Not run)


### Female+Male model
## Not run:
data("mlmm.gwas.FMI")

XX = list(Xf, Xm)
KK = list(K.female, K.male)
```

```
# GWAS
res_mlmm <- mlmm_allmodels(floweringDateFMI, XX, KK, female = female, male = male)
manhattan.plot(res_mlmm)

# Marker selection
res.threshold <- threshold_allmodels(threshold, res_mlmm)

## End(Not run)



### Female+Male+Interaction model
## Not run:
data("mlmm.gwas.FMI")

XX = list(Xf, Xm, Xfm)
KK = list(K.female, K.male, K.hybrid)

# GWAS
res_mlmm <- mlmm_allmodels(floweringDateFMI, XX, KK, female = female, male = male)
manhattan.plot(res_mlmm)

# Marker selection
res.threshold <- threshold_allmodels(threshold, res_mlmm)

## End(Not run)
```

Index