Package 'GWASpoly'

August 12, 2016

Title Genome-wide Association Studies for Autopolyploids

Version 1.3

Author Umesh R. Rosyara, Jeffrey B. Endelman

Maintainer Jeffrey Endelman <endelman@wisc.edu>

Description Extends the GWAS functionality of the rrBLUP package to handle autopoly-ploids. Marker dosage and polyploid genetic models are supported.

Imports rrBLUP, methods

Suggests parallel

License GPL-3

URL http://potatobreeding.cals.wisc.edu/software

NeedsCompilation no

R topics documented:

---

get.QTL   *Extracts significant QTL*

Description

Extracts significant QTL

Usage

```
get.QTL(data, traits = NULL, models = NULL)
```

Arguments

data	Object of class GWASpoly.thresh.
traits	Vector of trait names.
models	Vector of model names.

Details

The function returns which markers exceed the significance threshold.

Value

Data frame with the following columns: Trait, Model, Threshold, Marker, Chrom, Position, Ref, Alt, Score, Effect.

References

Rosyara et al. (in press) Software for genome-wide association studies in autopolyploids and its application to potato. Plant Genome

Examples

```
##see Vignette at http://potatobreeding.cals.wisc.edu/software
```

GWASpoly   *GWAS*

Description

Genome-wide association studies

Usage

```
GWASpoly(data, models, traits = NULL, params = NULL, n.core = 1, quiet = F)
```

Arguments

`data`	Object of class `GWASpoly.K`.
`models`	Vector of model names for the marker effect. See Details for options.
`traits`	Vector of trait names. If NULL, all traits are analyzed.
`params`	List of model parameters, created using `set.params`. If NULL, default values from `set.params` are used.
`n.core`	Number of cores to use for multicore processing (only Linux and MacOS).
`quiet`	TRUE/FALSE whether to suppress output charting the progress of the analysis

Details

The following marker-effect models are available:

- "additive" = Indicates the marker effect is proportional to the dosage of the alternate allele.

- "X-dom" = where X can be any integer between 1 and ploidy/2 and refers to the allele dosage needed for complete dominance (e.g., "1-dom" = simplex dominance, "2-dom" = duplex dominance). The software tries both dominance patterns for a given dosage model, e.g., whether the reference or alternate allele is dominant.

- "diplo-general" = All heterozygotes have the same effect.

- "diplo-additive" = All heterozygotes have the same effect, constrained to be halfway between the homozygous effects.

- "general" = There are no constraints on the effects of the different dosage levels.

To specify additional model parameters, such as the inclusion of fixed effects (Q matrix) and the minimum minor allele frequency, use `set.params`.

Value

Object of class `GWASpoly.fitted`.

References

Rosyara et al. (in press) Software for genome-wide association studies in autopolyploids and its application to potato. Plant Genome

Examples

```
##see Vignette at http://potatobreeding.cals.wisc.edu/software
```

GWASpoly-class   *Class* GWASpoly

Description

Base class for the GWASpoly package.

Slots

ploidy: Ploidy level (e.g., 2 for diploid, 4 for tetraploid)

map: Data frame with five columns: Marker, Chrom, Position, Ref, Alt

pheno: Data frame of phenotypes, with genotype identifiers as rownames.

geno: Numeric matrix (lines x markers) with marker data coded as the dosage of the alternate allele.

fixed: Data frame of factors and covariates to model as fixed effects.

Examples

```
showClass("GWASpoly")
```

GWASpoly.fitted-class   *Class* GWASpoly.fitted

Description

Extends the class GWASpoly.K by the addition of GWAS results, using the function GWASpoly.

Slots

ploidy: Ploidy level (e.g., 2 for diploid, 4 for tetraploid)

map: Data frame with five columns: Marker, Chrom, Position, Ref, Alt

pheno: Data frame of phenotypes, with genotype identifiers as rownames.

geno: Numeric matrix (lines x markers) with marker data coded as the dosage of the alternate allele.

fixed: Data frame of factors and covariates to model as fixed effects.

K: Positive semidefinite covariance (kinship) matrix.

scores: List of data.frames, one for each trait. The data frames contain the -log10(p) results, organized as markers x models.

effects: List of data.frames, one for each trait. The data frames contain the marker effect estimates, organized as markers x models. Effect estimates are not provided for the diplo-general or general models. Effect estimates for the additive model are based on marker coding 0,1,2,..ploidy. For the diplo-additive model, the coding is 0,1,2. For dominance models, the coding is 0,1.

params: List with the following model parameters:

- fixed = Names of fixed effects
- fixed.type = Vector of fixed effect types: either "numeric" (covariate) or "factor"
- n.PC = Number of principal components included as covariates
- min.MAF = Minimum minor allele frequency
- max.geno.freq = Maximum genotype frequency (after applying dominance relations)
- P3D = TRUE/FALSE whether the P3D approximation was used
- models = Names of marker-effect models

Examples

```
showClass("GWASpoly.fitted")
```

GWASpoly.K-class   *Class* GWASpoly.K

Description

Extends the class GWASpoly by the addition of a kinship matrix, using the function set.K.

Slots

ploidy: Ploidy level (e.g., 2 for diploid, 4 for tetraploid)

map: Data frame with five columns: Marker, Chrom, Position, Ref, Alt

pheno: Data frame of phenotypes, with genotype identifiers as rownames.

geno: Numeric matrix (lines x markers) with marker data coded as the dosage of the alternate allele.

fixed: Data frame of factors and covariates to model as fixed effects.

K: Positive semidefinite kinship matrix.

Examples

```
showClass("GWASpoly.K")
```

GWASpoly.thresh-class   *Class* GWASpoly.thresh

Description

Extends the class GWASpoly.fitted by the addition of significance thresholds, using the function set.threshold.

Slots

ploidy: Ploidy level (e.g., 2 for diploid, 4 for tetraploid)

map: Data frame with five columns: Marker, Chrom, Position, Ref, Alt

pheno: Data frame of phenotypes, with genotype identifiers as rownames.

geno: Numeric matrix (lines x markers) with marker data coded as the dosage of the alternate allele.

fixed: Data frame of factors and covariates to model as fixed effects.

K: Positive semidefinite covariance (kinship) matrix.

scores: List of data.frames, one for each trait. The data frames contain the -log10(p) results, organized as markers x models.

effects: List of data.frames, one for each trait. The data frames contain the marker effect estimates, organized as markers x models. Effect estimates are not provided for the diplo-general or general models. Effect estimates for the additive model are based on marker coding 0,1,2,..ploidy. For the diplo-additive model, the coding is 0,1,2. For dominance models, the coding is 0,1.

params: List with the following model parameters:

- fixed = Names of fixed effects
- fixed.type = Vector of fixed effect types: either "numeric" (covariate) or "factor"
- n.PC = Number of principal components included as covariates
- min.MAF = Minimum minor allele frequency
- max.geno.freq = Maximum genotype frequency (after applying dominance relations)
- P3D = TRUE/FALSE whether the P3D approximation was used
- models = Names of marker-effect models

threshold: Matrix (traits x models) of significance thresholds.

Examples

```
showClass("GWASpoly.thresh")
```

---

manhattan.plot   *Manhattan plot*

Description

Manhattan plot

Usage

```
manhattan.plot(data, trait, model, cex = 1, y.max = NULL, filename = NULL)
```

Arguments

data	Object inheriting from class GWASpoly.fitted.
trait	Name of trait to plot
model	Name of model to plot
cex	Controls size of points
y.max	Controls the upper limit for y = -log10(p). If NULL, y.max is determined from the data.
filename	If NULL, plot appears on screen. Providing a file name will generate a postscript file.

Details

If a threshold has been set with set.threshold, it is displayed as well.

Value

NULL

References

Rosyara et al. (in press) Software for genome-wide association studies in autopolyploids and its application to potato. Plant Genome

Examples

```
##see Vignette at http://potatobreeding.cals.wisc.edu/software
```

---

qq.plot *QQ plot*

Description

Generates a quantile-quantile plot of the observed vs. expected -log10(p) values.

Usage

```
qq.plot(data, trait, model, cex = 1, filename = NULL)
```

Arguments

data	Object inheriting from class GWASpoly.fitted.
trait	Name of trait to plot
model	Name of model to plot
cex	Controls size of points
filename	If NULL, plot appears on screen. Providing a file name will generate a postscript file.

Details

QQ plots are used to check the inflation of p-values in GWAS.

Value

NULL

References

Rosyara et al. (in press) Software for genome-wide association studies in autopolyploids and its application to potato. Plant Genome

Examples

```
##see Vignette at http://potatobreeding.cals.wisc.edu/software
```

---

read.GWASpoly *Read in data for GWASpoly*

Description

Reads marker and phenotype data

Usage

```
read.GWASpoly(ploidy, pheno.file, geno.file, format, n.traits, delim = ",")
```

Arguments

`ploidy`	Integer specifying the ploidy level (e.g., 2 for diploid, 4 for tetraploid)
`pheno.file`	Name of the phenotype file.
`geno.file`	Name of the genotype file.
`format`	Format for the marker data, either "numeric", "AB", or "ACGT" (see details).
`n.traits`	Integer specifying the number of traits in the phenotype file.
`delim`	Character to indicate the delimiter in the data files (e.g., "," for csv, "\t" for tab-delimited).

Details

The first column of the phenotype file contains the genotype identifier, columns 2 through (n.traits + 1) contain trait values, and subsequent columns contain the levels (for factors) or numeric values (for covariates) of any fixed effects.

The first three columns of the genotype file are (1) marker name, (2) chromosome, and (3) position. Subsequent columns contain the marker data for each individual in the population.

Marker data can be coded in one of three formats:

- "numeric" = markers are coded based on the dosage of the alternate allele, taking on integer values between 0 and ploidy (fractional values not allowed)
- "AB" format (e.g., AAAB, ABBB for tetraploids)
- "ACGT" format (e.g., AAAT, GGCC for tetraploids)

Only bi-allelic markers are allowed. Missing marker data will be imputed with the population mode (most frequent value) for each marker.

Value

An object of class GWASpoly.

References

Rosyara et al. (in press) Software for genome-wide association studies in autopolyploids and its application to potato. Plant Genome

Examples

```
##see Vignette at http://potatobreeding.cals.wisc.edu/software
```

---

set.K   *Creates kinship (relationship) matrix*

Description

Creates the kinship matrix for GWAS

Usage

```
set.K(data, K = NULL)
```

Arguments

`data`	Object of class `GWASpoly`.
`K`	Optional: User-supplied kinship matrix

Details

By default, K is computed as MM^T, where M is the centered genotype matrix (lines x markers). For GWAS, the overall scaling of K is irrelevant. At present, K is scaled such that the mean of its diagonal elements is 1. Alternatively, the user can supply any positive semidefinite K (with row.names that match the genotype identifiers in data@geno).

Value

Object of class `GWASpoly.K`.

References

Rosyara et al. (in press) Software for genome-wide association studies in autopolyploids and its application to potato. Plant Genome

Examples

```
##see Vignette at http://potatobreeding.cals.wisc.edu/software
```

---

set.params	*Set parameters for GWAS*

Description

Set parameters for GWAS

Usage

```
set.params(fixed=NULL,fixed.type=NULL,n.PC=0,MAF=0.05,geno.freq=0.95,P3D=T)
```

Arguments

`fixed`	Vector of names of fixed effects
`fixed.type`	Vector of effect types ("numeric" or "factor"), corresponding to the effects listed in "fixed".
`n.PC`	Number of principal components to include as covariates.
`MAF`	Minimum minor allele frequency
`geno.freq`	Maximum genotype frequency (after applying dominance relations)
`P3D`	TRUE/FALSE whether to use the P3D approximation (variance components not re-estimated for every marker).

Details

The list returned by the function should be passed to `GWASpoly`.

Value

A list with the following components

fixed	Names of fixed effects
fixed.type	Types of fixed effects
n.PC	Number of principal components to include as covariates
min.MAF	Minimum minor allele frequency
max.geno.freq	Maximum genotype frequency (after applying dominance relations)
P3D	TRUE/FALSE whether to use the P3D approximation

References

Rosyara et al. (in press) Software for genome-wide association studies in autopolyploids and its application to potato. Plant Genome

Examples

```
##see Vignette at http://potatobreeding.cals.wisc.edu/software
```

---

set.threshold   *Sets the significance threshold.*

Description

Sets the significance threshold.

Usage

```
set.threshold(data, method, level = 0.05, n.permute = 1000, n.core = 1)
```

Arguments

data	Object of class GWASpoly.fitted.
method	One of the following: "Bonferroni","FDR","permute"
level	Genome-wide false positive rate for the Bonferroni or permutation methods; false discovery rate for method FDR.
n.permute	Number of permutations for method "permute"
n.core	Number of cores to use for multicore processing (only Linux and MacOS).

Details

The FDR method is based on version 1.30.0 of the qvalue package.

Value

Object of class GWASpoly.thresh.

References

Rosyara et al. (in press) Software for genome-wide association studies in autopolyploids and its application to potato. Plant Genome

Examples

```
##see Vignette at http://potatobreeding.cals.wisc.edu/software
```

---

write.GWASpoly   *Write GWASpoly results to file*

Description

Write GWASpoly results to file

Usage

```
write.GWASpoly(data, trait, filename, what = "scores", delim = ",")
```

Arguments

data	Object inheriting from class GWASpoly.fitted.
trait	Name of trait to output
filename	Name of file
what	Two valid options: "scores" or "effects"
delim	Delimiter to use in the output file (default is comma).

Details

By default the "scores", which are the -log10(p) results, are written to a comma-separated values file. If what = "effects", the estimated effects for the markers are written to file instead of the scores. (Effects are not returned for the general and diplo-general models. See GWASpoly.fitted.)

Value

NULL

References

Rosyara et al. (in press) Software for genome-wide association studies in autopolyploids and its application to potato. Plant Genome

Examples

```
##see Vignette at http://potatobreeding.cals.wisc.edu/software
```

Index