

Package ‘polyBreedR’

July 19, 2020

Title Using genome-wide markers for polyploid breeding

Version 0.1

Author Jeffrey B. Endelman

Maintainer Jeffrey Endelman <endelman@wisc.edu>

Description Using genome-wide markers for polyploid breeding

Depends R (>= 3.5.0)

License GPL-3

LazyData true

RoxygenNote 7.1.0

Encoding UTF-8

Imports AGHmatrix, ggplot2, ggrepel, pedigree, grDevices, utils, tidyr

Suggests knitr, rmarkdown

VignetteBuilder knitr

R topics documented:

A_mat	2
check_ploidy	2
check_trio	3
geno_call	4
get_pedigree	5
GvsA	5
G_coeff	6
readXY	7
update_alias	7
Index	8

A_mat	<i>Additive relationship matrix from pedigree</i>
-------	---

Description

Additive relationship matrix from pedigree

Usage

```
A_mat(ped, ploidy, order.ped = TRUE)
```

Arguments

ped	Pedigree in three column format: id, mother, father
ploidy	2 or 4
order.ped	TRUE/FALSE does the pedigree need to be ordered so that progeny follow parents

Details

This is a wrapper that prepares the pedigree in the format required for R package AGHmatrix by Amadeu et al. (2016) (cite them if you use this function). A random bivalents model for tetraploid meiosis is assumed.

Value

Additive relationship matrix (dim: indiv x indiv)

References

Amadeu et al. (2016) Plant Genome 9, doi:10.3835/plantgenome2016.01.0009

check_ploidy	<i>Check ploidy</i>
--------------	---------------------

Description

Fraction of simplex or triplex markers

Usage

```
check_ploidy(geno, map)
```

Arguments

geno	Genotype matrix (markers x indiv)
map	Data frame with marker map (Marker, Chrom, Position)

Details

For every indiv in the genotype matrix, the fraction of markers per chromosome called as simplex or triplex is calculated, which should be low for diploids. A small amount of missing genotype data can be tolerated.

Value

- List containing
- mat** Matrix (indiv x chrom) of results
- plot** ggplot2 barplot

check_trio	<i>Check markers for parent-offspring trio</i>
------------	--

Description

Check markers for parent-offspring trio

Usage

check_trio(parentage, geno, ploidy)

Arguments

- parentage Data frame with three columns: id, mother, father
- geno Matrix of allele dosages: markers x indiv
- ploidy 2 or 4

Details

Computes the percentage of markers at which the two parents and offspring have incompatible allele dosages (for tetraploids, the random bivalents model is used). For dihaploid offspring of a single tetraploid parent, use ploidy = 4 and "haploid" for the father in parentage, as well as a diploid (0,1,2) genotype for the offspring. A small amount of missing genotype data can be tolerated.

Value

Data frame with the percentage of incompatible markers for each trio

geno_call	<i>Genotype calls</i>
-----------	-----------------------

Description

Genotype calls based on a normal mixture model

Usage

```
geno_call(  
  data,  
  filename,  
  model.ploidy = 4,  
  sample.ploidy = 4,  
  min.posterior = 0,  
  transform = TRUE  
)
```

Arguments

data	matrix (markers x id) of input values for the normal mixture model
filename	CSV filename with the model parameters
model.ploidy	2 or 4 (default)
sample.ploidy	2 or 4 (default)
min.posterior	minimum posterior probability (default 0) for genotype call
transform	TRUE (default) or FALSE whether to apply arcsin square root transformation

Details

The first column of the CSV input file should be the SNP ID, followed by columns for the normal distribution means, standard deviations, and mixture probabilities. Genotype calls are based on the maximum a posteriori (MAP) method. If the posterior probability of the MAP genotype is less than `min.posterior`, then NA is returned for that sample. By default, an arcsin square root transformation is applied to the input values to match the approach used by R package `fitPoly`. To use a tetraploid mixture model for diploid samples, set `sample.ploidy = 2` and `model.ploidy = 4`.

Value

matrix of allele dosages (0,1,2,..ploidy) with dimensions markers x individuals

get_pedigree	<i>Generate pedigree</i>
--------------	--------------------------

Description

Generate pedigree for a set of individuals

Usage

```
get_pedigree(id, pedfile, delim = ",", na.string = "NA", trim = TRUE)
```

Arguments

id	Vector of names of individuals
pedfile	Name of pedigree file
delim	Delimiter for the pedigree file (default is "," for CSV)
na.string	String used for NA in the pedigree file (default is "NA")
trim	TRUE/FALSE whether to trim pedigree (see Details)

Details

Finds ancestors of individuals in a three-column pedigree file (id,mother,father). The id column can be the identifier for an individual or cross. String matches must be exact or based on the naming convention crossID-progenyID. The returned pedigree is ordered using R package pedigree so that offspring follow parents. When trim is TRUE (default), the pedigree is trimmed to remove ancestors with only one offspring (which are not needed to compute the pedigree relationship matrix).

Value

Data frame with columns id, mother, father

GvsA	<i>Plot G vs. A</i>
------	---------------------

Description

Plot marker-based vs. pedigree-based additive relationship coefficients

Usage

```
GvsA(parentage, G, A, filename = NULL, thresh.G = Inf, thresh.A = 0.5)
```

Arguments

parentage	Data frame of individuals to plot, with 3 columns: id,mother,father
G	Genomic relationship matrix
A	Pedigree relationship matrix
filename	Name of PDF file to save the results (optional for one individual)
thresh.G	Threshold above which names are displayed (default Inf)
thresh.A	Threshold above which names are displayed (default 0.5)

Details

Useful for finding and correcting pedigree errors. If the G or A coefficient for an individual exceeds the threshold, its name is displayed in the figure. If parentage contains one individual, by default a ggplot2 variable will be returned, but the result can also be written to file. If multiple individuals are present, a filename is required.

G_coeff	<i>Coefficients to make the additive G matrix</i>
---------	---

Description

Coefficients of the additive effects for bi-allelic markers

Usage

```
G_coeff(geno, ploidy)
```

Arguments

geno	Matrix of allele dosages: markers x indiv
ploidy	2 or 4

Details

Additive effects are based on the traditional orthogonal decomposition of genetic variance in pan-mictic populations (Fisher 1918; Kempthorne 1957; Endelman et al. 2018). The G matrix can be computed from the coefficients and scaling factor according to $G = \text{tcrossprod}(\text{coeff}/\text{scale})$. Missing genotype data is replaced with the population mean.

Value

List containing

coeff Coefficients of the marker effects (dim: indiv x marker)

scale Scaling factor between markers and indiv

References

Fisher (1918) Trans. Roy. Soc. Edin. 52:399-433.

Kempthorne (1957) An Introduction to Genetic Statistics.

Endelman et al. (2018) Genetics 209:77-87.

readXY	<i>Read SNP array intensity data</i>
--------	--------------------------------------

Description

Read SNP array intensity data

Usage

```
readXY(filename, skip, output = "ratio")
```

Arguments

filename	filename
skip	number of lines to skip before the header line with the column names
output	Either "ratio" or "theta"

Details

The first two columns of the tab-delimited input file should be the SNP and Sample ID. Columns labeled "X" and "Y" contain the signal intensities for the two alleles. Use output to specify whether to return the ratio = $Y/(X+Y)$ or theta = $\text{atan}(Y/X)*2/\pi$.

Value

matrix with dimensions markers x individuals

update_alias	<i>Update names based on alias</i>
--------------	------------------------------------

Description

Update names based on data frame with alias and preferred name

Usage

```
update_alias(x, alias, remove.space = TRUE)
```

Arguments

x	Vector of names to update
alias	Data frame with two columns: first is the preferred name and second is the alias
remove.space	TRUE/FALSE

Details

Parameter `remove.space` indicates whether blank spaces should be removed before string matching

Value

Vector with updated names

Index

A_mat, [2](#)

check_ploidy, [2](#)

check_trio, [3](#)

G_coeff, [6](#)

geno_call, [4](#)

get_pedigree, [5](#)

GvSA, [5](#)

readXY, [7](#)

update_alias, [7](#)