Package ‘mas’

November 9, 2022

Type Package

Title Membership Association Studies

Version 0.1

Date 2022-11-04

Author Alencar Xavier [aut, cre],
Shizhong Xu [aut, ctb]

Maintainer Alencar Xavier `<alenxav@gmail.com>`

Description Genome-wide association analysis with membership information, variance adjustment, and correlated responses.

License GPL-3

Imports Rcpp, truncdist

LinkingTo Rcpp, RcppEigen

Depends R (>= 3.2.0), methods (>= 3.2.0)

NeedsCompilation yes

Repository CRAN

Date/Publication 2022-11-09 10:20:02 UTC

R topics documented:

`mas-package` *Membership Association Studies*

Description

Genome-wide association analysis with membership information, variance adjustment, and correlated responses.

Details

The DESCRIPTION file:

Package:	mas
Type:	Package
Title:	Membership Association Studies
Version:	0.1
Date:	2022-11-04
Authors@R:	c(person("Alencar", "Xavier", role=c("aut", "cre"), email = "alenxav@gmail.com"),person("Shizhong", "Xu",
Author:	Alencar Xavier [aut, cre], Shizhong Xu [aut, ctb]
Maintainer:	Alencar Xavier <alenxav@gmail.com>
Description:	Genome-wide association analysis with membership information, variance adjustment, and correlated response
License:	GPL-3
Imports:	Rcpp, truncdist
LinkingTo:	Rcpp, RcppEigen
Depends:	R (>= 3.2.0), methods (>= 3.2.0)

Index of help topics:

`gwas`	`Genome-Wide Association Studies`
`mas-package`	`Membership Association Studies`
`y`	`Soybean dataset`

Author(s)

Alencar Xavier [aut, cre], Shizhong Xu [aut, ctb] Maintainer: Alencar Xavier <alenxav@gmail.com>

References

Wei, J. and Xu, S., 2016. A random-model approach to QTL mapping in multiparent advanced generation intercross (MAGIC) populations. Genetics, 202(2), pp.471-486.

Xavier, A. and Habier, D., 2022. A new approach fits multivariate genomic prediction models efficiently. Genetics Selection Evolution, 54(1), pp.1-15.

Examples

```
# load the toy dataset
data( soy )

# run gwas
fit1 = gwas(y, Z, pop)
# adjust variances
fit2 = CorrectBeavis( fit1 )

# Compare before and after correction
plot( fit1, h2=TRUE, col=8, pch=20) # display QTL h2
plot( fit2, h2=TRUE, add=TRUE, pch=20, type='o') # adjusted QTL h2
legend('topleft',pch=16,col=c(8,1),c('Before correction','After Beavis correction'))
```

`gwas` *Genome-Wide Association Studies*

Description

Genome-wide association analysis that allows for integrating population membership into the analysis and information from correlated traits.

Usage

```
GWAS(Y,GEN,M,maxit=500,logtol=-8,cores=1,verb=TRUE)
CorrectBeavis(fit)
gwas(y,GEN,m,...)
```

Arguments

`Y`	Numeric matrix with phenotypic information. Columns are traits, rows are individuals. NA are allowed.
`GEN`	Numeric matrix of genotypic information. Columns are markers, rows are individuals. NA are not allowed.
`M`	Numeric matrix with membership information. Columns are population classes, rows are individuals. NA are not allowed.
`maxit`	Integer. Maximum number of iterations.
`logtol`	Scalar. Convergerce parameter of null-model in log scale.
`cores`	Integer. Number of CPUs to use OpenMP.
`verb`	Logical. Verbose: Print process status?
`fit`	Object of class GenModel, output of the GWAS function.
`y`	Numeric vector or matrix with phenotypic information.
`m`	Factor, vector, or numeric matrix with membership information.
`...`	Arguments to be passed to GWAS.

Details

The function `gwas` is a wrapper for `GWAS` that accepts vectors as inputs, suitable for single-trait analysis and discrete membership, such as family, ethnicity or population.

The linear model for the genome-wide association is set as follows:

$$y = Xb + Zu + g + e$$

where y is the response variable, Xb corresponds to the fixed effect term set as the membership matrix, Zu corresponds to the marker-membership interaction term, g is the polygenic term defines by $g = Ma$, where M is the genotypic matrix and g are marker effects, and e is the residual term. The null-hypothesis term cosists of a similar model, but without Zu. The significance threshold is based on the Bonferroni correction.

Theorical description of the model is provided by Wei and Xu (2016). Variance components are REML estimates under the univariate case. In the multivariate case, variance and covariance components are obtained through a general-purpose REML approximation (Van Raden and Jung 1987), which yields the exact same solution as REML when all traits are observed in all individuals. The multivariate generalization is described by Xavier and Habier (2022).

The package provides the function `MLM(Y,X,Z)` for users only interested in solving multivariate mixed models, which takes as input the phenotypic matrix (`Y`), design matrices of fixed effects (`X`) and random effect (`Z`), and the same controls (`logtol,cores`) as the function `GWAS`. Unlike the GWAS solved, this function uses an efficient solver without single-value decomposition, as described by Xavier and Habier (2022).

In the `GWAS` function, set `M=matrix(rep(1,nrow(Y)))` when membership is not knwon.

The output of functions `gwas/GWAS` is of class "GenModel", which has two custom functions: `plot` and `print`. The function `plot.GenModel` has a few arguments that allow users to choose the trait (`trait=1`), to overlay plots (`add=FALSE`), and to plot the QTL heritability instead of the association statistic (`h2=FALSE`).

Value

Returns a list of class "GenModel" with two sublists: GBLUP and GWAS. The GBLUP list contains fixed and random effect coefficients, variance components, genetic correlations, heritability. The GWAS list provides the Wald statistics and QTL variances for each marker-trait combination, as well as the regression coefficients for each marker-population combination for the last trait.

References

Van Raden, P.M. and Jung, Y.C., 1988. A general purpose approximation to restricted maximum likelihood: the tilde-hat approach. Journal of Dairy Science, 71(1), pp.187-194.

Wei, J. and Xu, S., 2016. A random-model approach to QTL mapping in multiparent advanced generation intercross (MAGIC) populations. Genetics, 202(2), pp.471-486.

Xavier, A. and Habier, D., 2022. A new approach fits multivariate genomic prediction models efficiently. Genetics Selection Evolution, 54(1), pp.1-15.

Examples

```
# load the toy dataset
data( soy )

# run gwas
fit1 = gwas(y, Z, pop)
# adjust variances
fit2 = CorrectBeavis( fit1 )

# Compare before and after correction
plot( fit1, h2=TRUE, col=8, pch=20) # display QTL h2
plot( fit2, h2=TRUE, add=TRUE, pch=20, type='o') # adjusted QTL h2
legend('topleft',pch=16,col=c(8,1),c('Before correction','After Beavis correction'))
```

`soy` *Soybean dataset*

Description

Toy dataset part of the SoyNAM. Plant height observed in a few soybean families in multiple environements.

Details

Data was obtained using the `BLUP` from the R package SoyNAM. SNPs are coded as 012, where 0 = homozygous as hub parent, 1 = heterozygous, 2 = homozygous as the founder parent. Details of the soybean nested association mappging (SoyNAM) population are provided by Song et al. (2017), Xavier et al. (2018) and Diers et al. (2018).

References

Song, Q., Yan, L., Quigley, C., Jordan, B.D., Fickus, E., Schroeder, S., Song, B.H., Charles An, Y.Q., Hyten, D., Nelson, R. and Rainey, K., 2017. Genetic characterization of the soybean nested association mapping population. The Plant Genome, 10(2), pp.plantgenome2016-10.

Xavier, A., Jarquin, D., Howard, R., Ramasubramanian, V., Specht, J.E., Graef, G.L., Beavis, W.D., Diers, B.W., Song, Q., Cregan, P.B. and Nelson, R., 2018. Genome-wide analysis of grain yield stability and environmental interactions in a multiparental soybean population. G3: Genes, Genomes, Genetics, 8(2), pp.519-529.

Diers, B.W., Specht, J., Rainey, K.M., Cregan, P., Song, Q., Ramasubramanian, V., Graef, G., Nelson, R., Schapaugh, W., Wang, D. and Shannon, G., 2018. Genetic architecture of soybean yield and agronomic traits. G3: Genes, Genomes, Genetics, 8(10), pp.3367-3375.

Examples

```
data(soy)
```

Index