Package ‘NAM’
March 11, 2015

Type  Package
Title  Nested Association Mapping Analysis
Version  1.1
Date  2014-03-03
Author  Alencar Xavier, William Muir, Katy Rainey, Tiago Pimenta, Qishan Wang, Shizhong Xu.
Maintainer  Alencar Xavier <xaviera@purdue.edu>
Description  Designed for association studies in nested association mapping (NAM) panels, but also handling biparental and random panels. It includes functions for genome-wide associations mapping of multiple populations, marker quality control, solving mixed models and finding variance components through REML and Gibbs sampling.
License  GPL-3
Imports  Rcpp, Matrix, missForest
LinkingTo  Rcpp
Depends  R (>= 3.1.0)
NeedsCompilation  yes
Repository  CRAN
Date/Publication  2015-03-11 07:29:10

R topics documented:

NAM-package ......................................................... 2
Bayesian Gibbs Sampler ............................................. 2
Fst ................................................................. 4
gwas ............................................................. 5
Internals ........................................................... 6
manhattan ......................................................... 7
reference ........................................................ 8
REML .............................................................. 9
snpH2 ........................................................... 10
snpQC ............................................................. 11
tpod ............................................................. 12

Index  13
**Bayesian Gibbs Sampler**

---

**Description**

Designed for association studies in nested association mapping (NAM) panels, but also handling biparental and random panels. It includes functions for genome-wide associations mapping of multiple populations, marker quality control, solving mixed models and finding variance components through REML and Gibbs sampling.

**Details**

Package: NAM  
Type: Package  
Version: 1.1  
Date: 2015-03-10  
License: GPL-3

**Author(s)**

Alencar Xavier, William Muir, Katy Rainey, Tiago Pimenta, Shizhong Xu  
Maintainer: Alencar Xavier <xaviera@purdue.edu>

**See Also**

Functions: gibbs, reml, Fst, gwas/gwas2, manhattan, reference, snpQC, snpH2

---

**Bayesian Gibbs Sampler**

---

**Description**

Univariate mixed model solver through Gibbs Sampling.

**Usage**

```r
gibbs(y,X=NULL,Z=NULL,iK=NULL,Iter=1500,Burn=500,Thin=4,DF=5)
```
Bayesian Gibbs Sampler

Arguments

\( y \)  
Numeric vector of observations \((n)\) describing the trait to be analyzed. NA is allowed.

\( x \)  
Formula or incidence matrix \((n \times p)\) for fixed effect. NA is not allowed.

\( z \)  
Formula or list of numeric matrix \((n \times p)\) with incidence matrices for random effect. NA is not allowed.

\( iK \)  
Numeric matrix or list of numeric matrices \((p \times p)\) corresponding to the the inverse kinship matrix of each random effect with \(p\) parameters.

\( \text{Iter} \)  
Integer. Number of iterations or samples to be generated.

\( \text{Burn} \)  
Integer. Number of iterations or samples to be discarded.

\( \text{Thin} \)  
Integer. Thinning parameter, used to save memory by storing only one every 'Thin' samples.

\( \text{DF} \)  
Integer. Hyper-parameter regarding degrees of freedom.

Details

Solve Gaussian mixed models in the Bayesian framework as described by García-Cortés and Sorensen (1996) and Sorensen and Gianola, D. (2002) with conjugated priors. The hyper-prior solution for the scale parameter of inverse Chi squared is calculated as proposed by de los Campos et al. (2013).

Value

The function gibbs returns a list with variance components distribution a posteriori (Posterior.VC) and mode estimated (VC.mode), a list with the posterior distribution of regression coefficients (Posterior.Coeff) and the posterior mode (Coeff.mode) and mean (Coeff.mean), and the fitted values using the mean (Fit.mean) and mode (Fit.mode) of posterior coefficients.

Author(s)

Alencar Xavier

References


Examples

```r
# Fitting Mixed Model
data(tpod)
S = seq(1, 350, 50)
```
Fst

Description

Genetic variation associated with markers distributed among subpopulations. The function generates a plot for structure diagnosis.

Usage

Fst(gen,fam)

Arguments

gen Numeric matrix containing the genotypic data. A matrix with \( n \) rows of observations and \( m \) columns of molecular markers. SNPs must be coded as \( 0, 1, 2 \), for founder homozygous, heterozygous and reference homozygous. NA is allowed.

fam Numeric vector of length \( n \) indicating which subpopulations (i.e. family) each observation comes from. NA is not allowed.

Details

Fst (Wright 1943) represents the differentiation among populations for the given locus. It is also a measure of genetic divergence between admixed populations. Neutral markers have an expected Fst 0.05, but may change according to the population. Outlier Fst supports signatures of selection. Permutation threshold is commonly used to verify statistical significance.

Value

Wright’s FST.

Author(s)

Alencar Xavier and William Muir
gwas

References

Examples

data(tpod)
Fst(gen=gen,fam=fam)

gwas

Empirical Bayes Genome Wide Association Mapping

Description
The gwas function calculates the likelihood ratio for each marker under the empirical Bayesian framework. The method also works with multiple populations.

Usage

gwas(y,gen,fam=NULL,chr=NULL,window=NULL)

Arguments

\textbf{y} \hspace{1cm} \text{Numeric vector of observations (}n\text{) describing the trait to be analyzed. NA is allowed.}

\textbf{gen} \hspace{1cm} \text{Numeric matrix containing the genotypic data. A matrix with }n\text{ rows of observations and (}m\text{) columns of molecular markers. SNPs must be coded as 0, 1, 2, for founder homozygous, heterozygous and reference homozigous. NA is allowed.}

\textbf{fam} \hspace{1cm} \text{Numeric vector of length (}n\text{) indicating which subpopulations (i.e. family) each observation comes from. Default assumes that all observations are from the same populations.}

\textbf{chr} \hspace{1cm} \text{Numeric vector indicating the number of markers in each chromosome. The sum of \textit{chr} must be equal to the number of columns in \textit{gen}. Default assumes that all markers are from the same chromosome.}

\textbf{window} \hspace{1cm} \text{Numeric. If specified, genetic distance between markers is used for moving window strategy (Wang 2015). Window must be specified in Morgans (e.g. 0.05 would represent 5cM). Genetic distance is calculated assuming that individuals are RILs.}

Details
Special incidence matrix is recreated to optimize the information provided by the subpopulations. Each locus is recoded as a vector with length \textit{f} equal to number of subpopulations, or NAM families.

For example, a locus heterozygous from an individual from subpopulation 2 is coded as [ 1, 0, 1, ... , \textit{f} ] , a locus homozigous for the reference allele from any subpopulation is coded as [ 2, 0, 0, ... , \textit{f} ]
[ ] and a locus homozigous for the founder allele from an individual from subpopulation 1 is coded as [ 0, 2, 0, ... , f ].

The base model for genome scanning includes the fixed effect ($Xb$), the marker ($Zu$), the polygene ($g$) and the residuals ($e$).

If the window term is specified, the model for genome scanning includes three extra terms, the left side genome ($Zu[ k - 1 ]$), the right side genome ($Zu[ k + 1 ]$) and window polygene ($-g[ k ]$).

The polygenic term is calculated only once (Zhang et al 2010) using eigendecomposition (Zhou ans Stephens 2012). Efficient inversion of capacitance matrix is obtained through the Woodbury matrix identities.

In order to analyze large dataset, one can avoid memory issues by using the function gwas2 that has the same arguments, except that the argument 'window' is not implemented.

Value

The function nam returns a list containing the method deployed (Method), predicted parameters and statistical test (PolyTest), genetic map for NAM panels (MAP) and the marker names (SNPs).

Author(s)

Alencar Xavier, Tiago Pimenta, Qishan Wang and Shizhong Xu

References


Examples

data(tpod)
test=gwas(y=y,gen=gen[,1:240],fam=fam,chr=chr[1:12])
manhattan(test,type="h",lwd=3)

Description

Functions written in C++ to speed up gwas and gibbs.

Author(s)

Alencar Xavier and Tiago Pimenta
**Description**

Generates a Manhattan plot for a `gwas` object.

**Usage**

```r
manhattan(gwas, colA = 2, colB = 4, alpha=NULL, GenDist = FALSE, OtherDist = NULL, ...)```

**Arguments**

- `gwas`: An output of the function `gwas`.
- `colA`: Color of even chromosomes.
- `colB`: Color of odd chromosomes.
- `alpha`: If specified, it provides the negative log p-value based on a chi squared distribution with a significance threshold for the given alpha.
- `GenDist`: Logical. If true, it displays the Manhattan plot with the estimated genetic map for NAM populations.
- `OtherDist`: Numeric vector of length \( m \), equal to the number of columns in `gen`. If provided, it changes the X coordinates used for the Manhattan plot.
- `...`: Other arguments for plotting function.

**Author(s)**

Alencar Xavier

**Examples**

```r
data(tpod)
test=gwas(y=y,gen=gen,fam=fam,chr=chr)
manhattan(gwas=test, colA=3, colB=1, type="h", lwd=2)
manhattan(gwas=test, GenDist=TRUE, pch=20)
```
Changing the Reference Genotype

Description

Function changes the reference genotype. For NAM populations, this function must be used when genotypes are coded according to the reference genome instead of the standard parent.

Usage

```r
reference(gen, ref = NULL)
```

Arguments

- **gen**: Numeric matrix containing the genotypic data. A matrix with `n` rows of observations and `(m)` columns of molecular markers. SNPs must be coded as 0, 1, 2.
- **ref**: Numeric vector of length `n` with elements coded as 0, 1, 2, it represents the genotypic information of a new reference genotype. Default assumes that the more frequent allele represents the reference genome.

Details

If genotypes are coded based on the reference genome, NAM analysis are optimized by using the standard parent as reference to allele coding.

Value

Returns a recoded `gen` matrix

Author(s)

Alencar Xavier

Examples

```r
data(tpod)
gen = reference(gen = gen, ref = NULL)
```
Restricted Maximum Likelihood

Description

Univariate REML estimators and variance components for a single random variable fitted by an EMMA-like algorithm.

Usage

`reml(y, X=NULL, Z=NULL, K=NULL)`

Arguments

- `y` Numeric vector of observations (n) describing the trait to be analyzed. NA is allowed.
- `X` Formula or incidence matrix (n by p) for fixed effect. NA is not allowed.
- `Z` Formula or numeric matrix (n by p) that corresponds to the incidence matrix of random effect. NA is not allowed.
- `K` Numeric matrix (p by p). Kinship matrix for random effect with p parameters. NA is not allowed.

Details

Solve mixed models with a single random effects minimizing the log restricted maximum likelihood (REML) using the EMMA algorithm (Kang et al 2008). Prediction of random coefficients are performed according to VanRaden (2008). A sample strategy is also available as MCreml.

Value

The function `reml` returns a list with variance components and heritability (VC), fixed effect coefficients and standard variations (Fixed) and estimated breeding values (EBV).

Author(s)

Alencar Xavier, Tiago Pimenta and Shizhong Xu

References


Examples

# Fitting a random model
data(tpod)
FIT = reml(y=y, Z=~as.factor(fam))

# Fitting GBLUP
G = tcrossprod(gen)
G = G/mean(diag(G))
GBLUP = reml(y=y, K=G)

### snpH2

#### SNP heritability

Description

Calculates the ability of markers to carry a gene.

Usage

snpH2(gen)

Arguments

gen Numeric matrix containing the genotypic data. A matrix with n rows of observations and (m) columns of molecular markers. Missing values not allowed.

Value

Numeric vector containing the heritability of each markers.

Author(s)

Alencar Xavier

References


Examples

data(tpod)
Heritability = snpH2(gen)
plot(Heritability)
Description

A function for quality control. It may be used to count/remove neighbor repeated SNPs and markers with MAF lower than a given threshold. This function is also used for imputations.

Usage

snpQC(gen, psy=1, MAF=0.05, remove=TRUE, impute=FALSE)

Arguments

- **gen**: Numeric matrix containing the genotypic data. A matrix with \( n \) rows of observations and \( m \) columns of molecular markers. SNPs must be coded as 0, 1, 2, for founder homozygous, heterozygous and reference homozygous. NA is allowed.
- **psy**: Tolerance parameter for repeated markers. Default is 1, which removes only SNPs 100% equal to its following neighbor.
- **MAF**: Minor Allele Frequency. Default is 0.05. Useful to inform or remove markers below the MAF threshold.
- **remove**: Remove SNPs that are redundant or pursue low MAF: TRUE/FALSE.
- **impute**: If TRUE, impute missing values using Random Forest implemented in the package missForest. Methods are described in Rutkoski et al (2013).

Value

Returns the genomic matrix without missing, redundant or low MAF markers.

Author(s)

Alencar Xavier, Katy Rainey and William Muir

References


Examples

data(tpod)
gen=snpQC(gen=gen, psy=1, MAF=0.05, remove=TRUE, impute=FALSE)
Description
Two biparental crosses phenotyped for the percentage of pods containing four seeds

Usage
```r
data(tpod)
```

Details
Soybean nested association panel with 2 families (`fam`) containing 196 individuals. Genotypic matrix (`gen`) have 376 SNP across 20 chromosome (`chr`). Phenotypic information (`y`) regards the proportion of tetra-seed pods. Data provided by Rainey Lab for Soybean Breeding and Genetics, Purdue University.

Author(s)
Alencar Xavier and Katy Rainey
Index

> Topic BLUP
  Bayesian Gibbs Sampler, 2
  gwas, 5
  NAM-package, 2
  REML, 9
> Topic Fst
  Fst, 4
> Topic GIBBS
  Bayesian Gibbs Sampler, 2
  NAM-package, 2
> Topic GWAS
  gwas, 5
  NAM-package, 2
> Topic NAM
  gwas, 5
  NAM-package, 2
> Topic REML
  NAM-package, 2
  REML, 9
.Random.seed (Internals), 6
Bayesian Gibbs Sampler, 2
calcSize (Internals), 6
chr (tpod), 12
fam (tpod), 12
Fst, 4
funI (Internals), 6
funX (Internals), 6
gen (tpod), 12
gibbs (Bayesian Gibbs Sampler), 2
gwas, 5
gwas2 (gwas), 5
inputRow (Internals), 6
Internals, 6
manhattan, 7
MCreml (REML), 9

NAM (NAM-package), 2
NAM-package, 2
phys (tpod), 12
RcppExports (Internals), 6
reference, 8
REML, 9
reml (REML), 9
SAMP (Internals), 6
snph2, 10
snpQC, 11
timesMatrix (Internals), 6
timesVec (Internals), 6
tpod, 12
y (tpod), 12

13