Package 'rNeighborGWAS'

April 15, 2020

Title Testing Neighbor Effects in Marker-Based Regressions

Version 1.0.0

Description
To incorporate neighbor genotypic identity into genome-wide association studies, the package provides a set of functions for variation partitioning and association mapping. The theoretical background of the method is described in Sato et al. (2019) <doi:10.1101/845735>.

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 7.1.0

Maintainer Yasuhiro Sato <sato.yasuhiro.36c@kyoto-u.jp>

Depends R (>= 2.10)

Imports gaston, Matrix, stats, graphics, RcppParallel

Suggests knitr, rmarkdown

VignetteBuilder knitr

NeedsCompilation no

Author Yasuhiro Sato [aut, cre] (<https://orcid.org/0000-0002-6466-723X>),
Eiji Yamamoto [aut],
Kentaro K. Shimizu [aut],
Atsushi J. Nagano [aut]

Repository CRAN

Date/Publication 2020-04-15 13:20:02 UTC

R topics documented:

`calc_PVEnei` *Calculating proportion of phenotypic variation explained (PVE) by neighbor effects*

Description

A function to calculate PVE by neighbor effects for a series of neighbor distance using a mixed model.

Usage

```
calc_PVEnei(
  pheno,
  geno,
  smap,
  scale_seq,
  addcovar = NULL,
  grouping = rep(1, nrow(smap)),
  response = "quantitative",
  n_core = 1L
)
```

Arguments

`pheno`	A numeric vector including phenotypes for individuals
`geno`	An individual x marker matrix. Bialleles (i.e., A or a) must be converted into -1 or 1 digit.
`smap`	A matrix showing a spatial map for individuals. The first and second column include spatial points along a x-axis and y-axis, respectively.
`scale_seq`	A numeric vector including a set of the maximum spatial distance between a focal individual and neighbors to define neighbor effects. A scalar is also allowed.
`addcovar`	An optional matrix including additional non-genetic covariates. It contains no. of individuals x no. of covariates.
`grouping`	A integer vector assigning each individual to a group. This argument can be useful when a "smap" contains different experimental replicates. Default setting means that all individuals are belong to a single group.
`response`	An option to select if the phenotype is a `"quantitative"` trait subject to linear models, or a `"binary"` trait subject to logistic models.
`n_core`	No. of cores for a multi-core computation. This does not work for Windows OS. Default is a single-core computation.

Details

This function uses mixed models via the gaston package (Perdry & Dandine-Roulland 2020). If "binary" is selected, logistic.mm.aireml() is called via the gaston package. In such a case, PVEnei below is given by the variance component parameter σ (i.e., not a proportional value) and p-values are not provided.

Value

A numeric matrix including a given spatial scale, PVE by neighbor effects, and p-values.

- scale Maximum neighbor distance given as an argument
- PVEnei Proportion of phenotypic variation explained (PVE) by neighbor effects
- p-value p-value by a likelihood ratio test between models with or without neighbor effects

Author(s)

Yasuhiro Sato (<sato.yasuhiro.36c@kyoto-u.jp>)

References

Perdry H, Dandine-Roulland C. (2020) gaston: Genetic Data Handling (QC, GRM, LD, PCA) & Linear Mixed Models. https://CRAN.R-project.org/package=gaston

Examples

```
set.seed(1)
g <- matrix(sample(c(-1,1),100*1000,replace = TRUE),100,1000)
gmap <- cbind(c(rep(1,nrow(g)/2),rep(2,nrow(g)/2)),c(1:ncol(g)))
x <- runif(nrow(g),1,100)
y <- runif(nrow(g),1,100)
smap <- cbind(x,y)
grouping <- c(rep(1,nrow(g)/2), rep(2,nrow(g)/2))
pheno <- nei_simu(geno=g,smap=smap,scale=44,grouping=grouping,n_causal=50,pveB=0.4,pve=0.8)

fake_nei <- list()
fake_nei[[1]] <- g
fake_nei[[2]] <- gmap
fake_nei[[3]] <- smap
fake_nei[[4]] <- data.frame(pheno,grouping)
names(fake_nei) <- c("geno","gmap","smap","pheno")

min_s <- min_dist(fake_nei$smap, fake_nei$pheno$grouping)
scale_seq <- c(min_s, quantile(dist(fake_nei$smap),c(0.2*rep(1:5))))

pve_out <- calc_PVEnei(geno=fake_nei$geno, pheno=fake_nei$pheno[,1],
                       smap=fake_nei$smap, scale_seq=scale_seq,
                       addcovar=as.matrix(fake_nei$pheno$grouping),
                       grouping=fake_nei$pheno$grouping)
delta_PVE(pve_out)
```

delta_PVE *Estimating the effective scale of neighbor effects*

Description

A function to calculate ΔPVE that estimates the effective scale of neighbor effects.

Usage

```
delta_PVE(res, fig = TRUE, ...)
```

Arguments

res	Output results of calc_PVEnei().
fig	TRUE/FALSE to plot the results (or not). Default is TRUE.
...	Arguments to be passed to plot().

Value

Estimated effective scale and proportion of phenotypic variation explained by neighbor effects at that scale.

Author(s)

Yasuhiro Sato (<sato.yasuhiro.36c@kyoto-u.jp>)

See Also

calc_PVEnei

min_dist *Calculating the minimum distance*

Description

A function to calculate a Euclidian distance including at least one neighbor for all individuals.

Usage

```
min_dist(smap, grouping = rep(1, nrow(smap)))
```

Arguments

smap	A matrix showing a spatial map for individuals. The first and second column include spatial points along a x-axis and y-axis, respectively.
grouping	A integer vector assigning each individual to a group. This argument can be useful when a "smap" contains different experimental replicates. Default setting means that all individuals are belong to a single group.

Value

Return a scalar of the minimum Euclidian distance that allows all individuals to have at least one neighbor.

Author(s)

Yasuhiro Sato (<sato.yasuhiro.36c@kyoto-u.jp>)

Examples

```
set.seed(1)
g <- matrix(sample(c(-1,1),100*1000,replace = TRUE),100,1000)
gmap = cbind(c(rep(1,nrow(g)/2),rep(2,nrow(g)/2)),c(1:ncol(g)))
x <- runif(nrow(g),1,100)
y <- runif(nrow(g),1,100)
smap <- cbind(x,y)
grouping <- c(rep(1,nrow(g)/2), rep(2,nrow(g)/2))
pheno <- nei_simu(geno=g,smap=smap,scale=44,grouping=grouping,n_causal=50,pveB=0.4,pve=0.8)

fake_nei <- list()
fake_nei[[1]] <- g
fake_nei[[2]] <- gmap
fake_nei[[3]] <- smap
fake_nei[[4]] <- data.frame(pheno,grouping)
names(fake_nei) <- c("geno","gmap","smap","pheno")

min_s <- min_dist(fake_nei$smap, fake_nei$pheno$grouping)
```

---

neiGWAS *Genome-wide association mapping of neighbor effects*

Description

A function to test neighbor effects for each marker and to calculate p-values at a given reference scale.

Usage

```
neiGWAS(
  geno,
  pheno,
  gmap,
  smap,
  scale,
  addcovar = NULL,
  grouping,
  response = "quantitative",
  model = "lmm",
  n_core = 1L
)
```

Arguments

geno	An individual x marker matrix. Bialleles (i.e., A or a) must be converted into -1 or 1 digit.
pheno	A numeric vector including phenotypes for individuals
gmap	A matrix or data.frame including chromosome numbers in the first column, and SNP positions in the second column.
smap	A matrix showing a spatial map for individuals. The first and second column include spatial points along a x-axis and y-axis, respectively.
scale	A numeric scalar indicating the maximum spatial distance between a focal individual and neighbors to define neighbor effects.
addcovar	An optional matrix including additional non-genetic covariates. It contains no. of individuals x no. of covariates.
grouping	A integer vector assigning each individual to a group. This argument can be useful when a "smap" contains different experimental replicates. Default setting means that all individuals are belong to a single group.
response	An option to select if the phenotype is a `"quantitative"` trait subject to linear models, or a `"binary"` trait subject to logistic models.
model	An option to select linear mixed model `"lmm"` or linear model `"lm"`. Default setting is to use a mixed model.
n_core	No. of cores for a multi-core computation. This does not work for Windows OS. Default is a single-core computation.

Details

This function calls a mixed model via the `gaston` package. If `"lmm"` with `"binary"` is selected, p-values are based on Wald tests. This is because the logistic mixed model is based on a pseudo-likelihood and thus likelihood ratio tests are not applicable. See Chen et al. (2016) for the theory.

Value

A data.frame including the chromosome number, marker position, and p-values.

- chr Chromosome number
- pos Marker position
- p p-value by a likelihood ratio test between models with or without neighbor effects

Author(s)

Yasuhiro Sato (<sato.yasuhiro.36c@kyoto-u.jp>)

References

Chen H, Wang C, Conomos M. et al. (2016) Control for population structure and relatedness for binary traits in genetic association studies via logistic mixed models. The American Journal of Human Genetics 98: 653-666.

Examples

```
set.seed(1)
g <- matrix(sample(c(-1,1),100*1000,replace = TRUE),100,1000)
gmap <- cbind(c(rep(1,nrow(g)/2),rep(2,nrow(g)/2)),c(1:ncol(g)))
x <- runif(nrow(g),1,100)
y <- runif(nrow(g),1,100)
smap <- cbind(x,y)
grouping <- c(rep(1,nrow(g)/2), rep(2,nrow(g)/2))
pheno <- nei_simu(geno=g,smap=smap,scale=44,grouping=grouping,n_causal=50,pveB=0.4,pve=0.8)

fake_nei <- list()
fake_nei[[1]] <- g
fake_nei[[2]] <- gmap
fake_nei[[3]] <- smap
fake_nei[[4]] <- data.frame(pheno,grouping)
names(fake_nei) <- c("geno","gmap","smap","pheno")

scale <- 43
gwas_out <- neiGWAS(geno=fake_nei$geno, pheno=fake_nei$pheno[,1],
                    gmap=fake_nei$gmap, smap=fake_nei$smap,
                    scale=scale, addcovar=as.matrix(fake_nei$pheno$grouping),
                    grouping=fake_nei$pheno$grouping)

gaston::manhattan(gwas_out)
gaston::qqplot.pvalues(gwas_out$p)
```

nei_coval *Calculating neighbor genotypic identity*

Description

A function to calculate neighbor genotypic identity, with a given reference scale and a degree of distance decay.

Usage

```
nei_coval(
  geno,
  smap,
  scale,
  alpha = Inf,
  kernel = "exp",
  grouping = rep(1, nrow(smap)),
  n_core = 1L
)
```

Arguments

geno	An individual x marker matrix. Bialleles (i.e., A or a) must be converted into -1 or 1 digit.
smap	A matrix showing a spatial map for individuals. The first and second column include spatial points along a x-axis and y-axis, respectively.
scale	A numeric scalar indicating the maximum spatial distance between a focal individual and neighbors to define neighbor effects.
alpha	Distance decay coefficient α in a dispersal kernel. Default is set at Inf, meaning no distance decay.
kernel	Type of dispersal kernel in the distance decay. Select "exp" or "gaussian" for a negative exponential kernel (fat-tailed) or Gaussian kernel (thin-tailed), respectively.
grouping	A integer vector assigning each individual to a group. This argument can be useful when a "smap" contains different experimental replicates. Default setting means that all individuals are belong to a single group.
n_core	No. of cores for a multi-core computation. This does not work for Windows OS. Default is a single-core computation.

Value

A numeric matrix for neighbor covariates, with no. of individuals x markers.

Author(s)

Yasuhiro Sato (<sato.yasuhiro.36c@kyoto-u.jp>)

References

Nathan R, Klein E, Robledo-Arnuncio JJ, Revilla E, (2012) Dispersal kernels: review. In: Clobert J, Baguette M, Benton TG, Bullock JM (Eds.), *Dispersal Ecology and Evolution*. Oxford University Press, pp.186-210.

Examples

```
set.seed(1)
g <- matrix(sample(c(-1,1),100*1000,replace = TRUE),100,1000)
gmap <- cbind(c(rep(1,nrow(g)/2),rep(2,nrow(g)/2)),c(1:ncol(g)))
x <- runif(nrow(g),1,100)
y <- runif(nrow(g),1,100)
smap <- cbind(x,y)
grouping <- c(rep(1,nrow(g)/2), rep(2,nrow(g)/2))

g_nei <- nei_coval(g,smap,44,grouping = grouping)
```

`nei_simu` *Simulating phenotypes with self and neighbor effects*

Description

A function to simulate phenotypes caused by self and neighbor effects, with the proportion of phenotypic variation explained (PVE) by fixed and random effects controlled.

Usage

```
nei_simu(
  geno,
  smap,
  scale,
  alpha = Inf,
  grouping = grouping,
  kernel = "exp",
  n_causal,
  pveB,
  pve,
  b_ratio = c(1, 1)
)
```

Arguments

`geno`	An individual x marker matrix. Bialleles (*i.e.*, A or a) must be converted into -1 or 1 digit.
`smap`	A matrix showing a spatial map for individuals. The first and second column include spatial points along a x-axis and y-axis, respectively.

scale	A numeric scalar indicating the maximum spatial distance between a focal individual and neighbors to define neighbor effects.
alpha	Distance decay coefficient α in a dispersal kernel. Default is set at Inf, meaning no distance decay.
grouping	A integer vector assigning each individual to a group. This argument can be useful when a "smap" contains different experimental replicates. Default setting means that all individuals are belong to a single group.
kernel	An option to select a negative exponential kernel "exp" or Gaussian kernel "gaussian".
n_causal	No. of causal markers in a simulated phenotype
pveB	Proportion of phenotypic variation explained by fixed effects.
pve	Proportion of phenotypic variation explained by fixed and random effects.
b_ratio	A vector composed of two numeric scalars that control the ratio of contributions of self or neighbor effects to a phenotype. The first and second element are for self and neighbor effects, respectively.

Value

A vector of simulated phenotype values for all individuals

Author(s)

Yasuhiro Sato (<sato.yasuhiro.36c@kyoto-u.jp>)

Examples

```
set.seed(1)
g <- matrix(sample(c(-1,1),100*1000,replace = TRUE),100,1000)
gmap <- cbind(c(rep(1,nrow(g)/2),rep(2,nrow(g)/2)),c(1:ncol(g)))
x <- runif(nrow(g),1,100)
y <- runif(nrow(g),1,100)
smap <- cbind(x,y)
grouping <- c(rep(1,nrow(g)/2), rep(2,nrow(g)/2))
pheno <- nei_simu(geno=g,smap=smap,scale=44,grouping=grouping,n_causal=50,pveB=0.4,pve=0.8)

fake_nei <- list()
fake_nei[[1]] <- g
fake_nei[[2]] <- gmap
fake_nei[[3]] <- smap
fake_nei[[4]] <- data.frame(pheno,grouping)
names(fake_nei) <- c("geno","gmap","smap","pheno")
```

`qtl_pheno_simu` *Simulating phenotype values with neighbor effects.*

Description

A function to simulate phenotype values with multiple sources of variation controlled

Usage

```
qtl_pheno_simu(
  b_self,
  b_nei,
  eigenK_self,
  eigenK_nei,
  b_ratio = c(1, 1),
  pveB,
  pve
)
```

Arguments

`b_self`	A n x 1 genotype vector to design major additive genetic effect.
`b_nei`	A vector of explanatory variable for beta_2 (neighbor effects)
`eigenK_self`	Products of function 'eigen' with self covariance matrices that were used as explanatory variables for the phenotype.
`eigenK_nei`	Products of function 'eigen' with neighbor covariance matrices that were used as explanatory variables for the phenotype.
`b_ratio`	Ratio for contribution of engenK_self, eigenK_nei and eigenK_sxn to the phenotype.
`pveB`	Proportion of variance explained by genetic effect designed by b_.. vector.
`pve`	Proportion of variance explained by all genetic effects (i.e., b_.. and eigenK_..)

Value

A list of simulated phenotypes

- ySimulated phenotype values
- beta_selfmajor self-genetic effects
- beta_neimajor neighbor effects
- sigma_selfself polygenic effects
- sigma_neineighbor polygenic effects
- epsilonresiduals
- res_pveBrealized proportion of variation explained by major-effect genes
- res_pverealized proportion of variation explained by major-effect genes and polygenic effects

Author(s)

Eiji Yamamoto, and Yasuhiro Sato

`w` *Calculating a distance decay weight*

Description

A function to calculate, with a negative exponential or Gaussian dispersal kernel.

Usage

```
w(s, a, kernel = "exp")
```

Arguments

`s`	A numeric scalar indicating spatial distance at which the distance decay is referred
`a`	A numeric scalar indicating a decay coefficient
`kernel`	An option to select a negative exponential kernel `"exp"` or Gaussian kernel `"gaussian"`.

Value

A numeric scalar for a distance decay weight.

Author(s)

Yasuhiro Sato (<sato.yasuhiro.36c@kyoto-u.jp>)

Index