Package 'rNeighborGWAS'

April 7, 2021

Title Testing Neighbor Effects in Marker-Based Regressions

Version 1.2.3

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. (2021) <doi:10.1038/s41437-020-00401-w>.

License GPL-3

Encoding UTF-8

RoxygenNote 7.1.1

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

Suggests knitr, rmarkdown, testthat

VignetteBuilder knitr

Imports gaston, Matrix, RcppParallel, stats, graphics

NeedsCompilation no

Author Yasuhiro Sato [aut, cre] (<https://orcid.org/0000-0002-6466-723X>),
Eiji Yamamoto [aut],
Kentaro K. Shimizu [aut] (<https://orcid.org/0000-0002-6483-1781>),
Atsushi J. Nagano [aut] (<https://orcid.org/0000-0001-7891-5049>)

Repository CRAN

Date/Publication 2021-04-07 20:00:02 UTC

R topics documented:

`calc_PVEnei` *Calculating phenotypic variation explained 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 = c("quantitative", "binary"),
  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 an 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 positive 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 replaced by the ratio of phenotypic variation explained (RVE) by neighbor effects as RVE_nei = σ_2^2/σ_1^2 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
- PVEself Proportion of phenotypic variation explained (PVE) by self effects. RVE is returned when response = "binary"
- PVEnei Proportion of phenotypic variation explained (PVE) by neighbor effects. RVE is returned when response = "binary"
- p-value p-value by a likelihood ratio test between models with or without neighbor effects (when s is not zero); or between a null model and model with self effects alone (when s = 0). NA when response = "binary"

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`

gaston2neiGWAS *Convert gaston's bed.matrix data to rNeighborGWAS genotype data.*

Description

A function convert a bed.matrix dataset to rNeighborGWAS genotype data.

Usage

```
gaston2neiGWAS(x)
```

Arguments

x	A bed.matrix created using the gaston package (Perdry & Dandine-Roulland 2020).

Details

This function converts genotype data into -1, 0, or 1 digit as the rNeighborGWAS format. Zero indicates heterozygotes.

Value

A list including an individual x marker matrix, a data.frame including chromosome numbers in the first column, and SNP positions in the second column, and a numeric vector including phenotypes for individuals.

- geno Individual x marker matrix
- gmap Data.frame including chromosome numbers in the first column, and SNP positions in the second column
- pheno Numeric vector including phenotypes for individuals

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

```
data("TTN", package="gaston")
x <- gaston::as.bed.matrix(TTN.gen, TTN.fam, TTN.bim)
g <- gaston2neiGWAS(x)
```

`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 an x-axis and y-axis, respectively.
`grouping`	A positive 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 = rep(1, nrow(smap)),
  response = c("quantitative", "binary"),
  model = c("lmm", "lm"),
  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 an 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 positive 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 = c("exp", "gaussian"),
  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 an 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`	An option to set a distance decay coefficient α in a dispersal kernel. Default is set at Inf, meaning no distance decay.
`kernel`	An option to select either `"exp"` or `"gaussian"` for a negative exponential kernel or Gaussian kernel, respectively.
`grouping`	A positive 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.

Details

Default setting is recommended for `alpha` and `kernel` arguments unless spatial distance decay of neighbor effects needs to be modeled. If `alpha` is not `Inf`, output variables are weighted by a distance decay from a focal individual to `scale`. For the type of dispersal kernel in the distance decay, we can choose a negative exponential or Gaussian kernel as a fat-tailed or thin-tailed distribution, respectively. See Nathan et al. (2012) for detailed characteristics of the two dispersal kernels.

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_lm *Standard linear models for testing self and neighbor effects*

Description

A function to provide coefficients and p-values of self and neighbor effects for each marker.

Usage

```
nei_lm(
  geno,
  g_nei,
  pheno,
  addcovar = NULL,
  response = c("quantitative", "binary"),
  n_core = 1L,
  asym = FALSE
)
```

Arguments

geno	An individual x marker matrix. Bialleles (i.e., A or a) must be converted into -1 or 1 digit.
g_nei	An output of nei_coval() object, namely an individual x marker matrix including neighbor genotypic identity.
pheno	A numeric vector including phenotypes for individuals
addcovar	An optional matrix including additional non-genetic covariates. It contains no. of individuals x no. of covariates.
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.
asym	If TRUE, asymmetric neighbor effects are also tested and returned as "beta_sxn" and "p_sxn".

Details

This function is a subset of neiGWAS(). nei_lm() gives detailed results when the option model="lm" is selected in neiGWAS().

Value

A data.frame including coefficients and p-values of self and neighbor effects, without the chromosome numbers and marker position.

- beta_self coefficient for self effects
- beta_self coefficient for neighbor effects
- p_self p-value for self effects by a likelihood ratio test between a null and standard linear model
- p_nei p-value for neighbor effects by a likelihood ratio test between models with or without neighbor effects

Author(s)

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

See Also

neiGWAS

nei_lmm *Mixed models for testing self and neighbor effects*

Description

A function to provide coefficients and p-values of self and neighbor effects for each marker.

Usage

```
nei_lmm(
  geno,
  g_nei,
  pheno,
  addcovar = NULL,
  response = c("quantitative", "binary"),
  n_core = 1L,
  asym = FALSE
)
```

Arguments

geno	An individual x marker matrix. Bialleles (i.e., A or a) must be converted into -1 or 1 digit.
g_nei	An output of `nei_coval()` object, namely an individual x marker matrix including neighbor genotypic identity.
pheno	A numeric vector including phenotypes for individuals
addcovar	An optional matrix including additional non-genetic covariates. It contains no. of individuals x no. of covariates.
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.
asym	If TRUE, asymmetric neighbor effects are also tested and returned as "beta_sxn" and "p_sxn".

Details

This function is a subset of `neiGWAS()`. `nei_lmm()` gives detailed results but requires more computational time.

Value

A data.frame including coefficients and p-values of self and neighbor effects, without the chromosome numbers and marker position.

- `beta_self` coefficient for self effects

- `beta_self` coefficient for neighbor effects
- `p_self` p-value for self effects by a likelihood ratio test between a null and standard GWAS model
- `p_nei` p-value for neighbor effects by a likelihood ratio test between models with or without neighbor effects

Author(s)

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

See Also

neiGWAS

`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 = rep(1, nrow(smap)),
  kernel = c("exp", "gaussian"),
  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 an 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 positive 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 "exp" or Gaussian "gaussian" for a dispersal kernel of neighbor effects.
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 an explanatory variable for neighbor effects
eigenK_self	Products of eigen() with self covariance matrices that are used as explanatory variables for the phenotype.
eigenK_nei	Products of eigen() with neighbor covariance matrices that are used as explanatory variables for the phenotype.
b_ratio	Ratio for contributions of eigenK_self and eigenK_nei to the phenotype.
pveB	Proportion of variance explained by genetic effects attributable to the fixed effects (i.e., b_.. vector).
pve	Proportion of variance explained by all genetic effects (i.e., b_.. and eigenK_..)

Value

A list of simulated phenotypes

- y Simulated phenotype values
- beta_self major self-genetic effects
- beta_nei major neighbor effects
- sigma_self self polygenic effects
- sigma_nei neighbor polygenic effects
- epsilon residuals
- res_pveB realized proportion of variation explained by major-effect genes
- res_pve realized 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 = c("exp", "gaussian"))
```

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 "exp" or Gaussian "gaussian" for a dispersal kernel of neighbor effects.

Value

A numeric scalar for a distance decay weight.

Author(s)

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

Index