Package: simulMGF (via r-universe)

August 22, 2024

Type Package

Title Simulate SNP Matrix, Phenotype and Genotypic Effects

Version 0.1.1

Author Martin Nahuel Garcia [aut, cre]

Maintainer Martin Nahuel Garcia <`garcia.martin@inta.gob.ar`>

Description Simulate genotypes in SNP (single nucleotide polymorphisms) Matrix as random numbers from an uniform distribution, for diploid organisms (coded by 0, 1, 2), Sikorska et al., (2013) <doi:10.1186/1471-2105-14-166>, or half-sib/full-sib SNP matrix from real or simulated parents SNP data, assuming mendelian segregation. Simulate phenotypic traits for real or simulated SNP data, controlled by a specific number of quantitative trait loci and their effects, sampled from a Normal or an Uniform distributions, assuming a pure additive model. This is useful for testing association and genomic prediction models or for educational purposes.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

URL `https://github.com/mngar/simulMGF`

RoxygenNote 7.2.3

Suggests spelling

Language en-US

Repository https://mngar.r-universe.dev

RemoteUrl https://github.com/mngar/simulmgf

RemoteRef HEAD

RemoteSha 76685a3a972b32bb968b7606658256036af91478

Contents

`simGeno` *Function to simulate SNP matrix*

Description

Simulate SNP matrix coded 0, 1 and 2; with random genotypes.

Usage

```
simGeno(Nind, Nmarkers)
```

Arguments

`Nind`	number of individuals to simulate.
`Nmarkers`	Nmarkers number of SNP markers to generate.

Value

a matrix of dimensions Nind x Nmarkers.

Author(s)

Martin Nahuel Garcia <orcid:0000-0001-5760-986X>

References

Wu, R., Ma, C., & Casella, G. (2007). Statistical genetics of quantitative traits: linkage, maps and QTL. Springer Science & Business Media.

Examples

```
#simulate 100 individuals and 1000 SNPs
set.seed(123)
simGeno(100, 1000)
#[1] "simG was generated"
dim(simG);simG[1:5,1:5]
#[1]  100 1000
#[,1] [,2] [,3] [,4] [,5]
#[1,]    0    1    0    2    2
```

```
#[2,]    2    0    2    0    0
#[3,]    1    1    1    2    2
#[4,]    2    2    1    2    1
#[5,]    2    1    1    1    1
```

`simPheno` *Function to simulate phenotypes*

Description

Simulate a phenotype from a genotype matrix with QTLs with random effects sampled from a Normal distribution.

Usage

```
simPheno(x, Nqtl, Esigma, Pmean, Perror)
```

Arguments

`x`	SNP matrix coded like 0 homozygous; 1 heterozygous; 2 homozygous
`Nqtl`	number of QTLs to simulate
`Esigma`	standard deviation of effects with distribution N~(0,Esigma^2)
`Pmean`	phenotype mean
`Perror`	standard deviation of error (portion of phenotype not explained by genomic information)

Value

An object of class list containing the trait, the markers associated and their effects.

`pheno`	vector with the trait values simulated.
`QTN`	column in the SNP matrix with the SNP associated.
`Meffects`	effects of the associated SNPs.

Author(s)

Martin Nahuel Garcia <orcid:0000-0001-5760-986X>

References

Wu, R., Ma, C., & Casella, G. (2007). Statistical genetics of quantitative traits: linkage, maps and QTL. Springer Science & Business Media.

Examples

```
set.seed(123)
simGeno(100, 1000)
#' #[1] "simG was generated"
simPheno(simG, 50, .8, 12, .5)
#[1] "simP was generated"
str(simP)
#List of 3
#$ pheno   : num [1:100, 1] 24 20.5 15.6 13.6 18.5 ...
#$ QTN     : int [1:50] 568 474 529 349 45 732 416 51 413 514 ...
#$ Meffects: num [1:50] 0.2396 -0.138 0.906 0.0186 1.0687 ...
```

`simulFS` *Function to simulate full sib progeny genotype*

Description

Simulate full sib progeny genotypes from the genotype of the parents (matrixes with the same dimensions). Pair of parents mating will be in the order of the matrixes. We assume that these are diploid organisms.

Usage

```
simulFS(x, y, Nprogeny)
```

Arguments

`x`	genotype matrix of a set of moms
`y`	genotype matrix of a set of dads
`Nprogeny`	Nprogeny number of progeny's genotypes to generate from each pair of parents

Value

a matrix of dimensions (nrow(x)*Nprogeny) x ncol(x)

Author(s)

Martin Nahuel Garcia <orcid:0000-0001-5760-986X>

References

Wu, R., Ma, C., & Casella, G. (2007). Statistical genetics of quantitative traits: linkage, maps and QTL. Springer Science & Business Media.

Examples

```
#simulate 100 individuals and 1000 SNPs
set.seed(123)
simGeno(100, 1000)
#[1] "simG was generated"
#simulate the genotype of 5 FS from 3 pairs of parents
simulFS(simG[1:3,],simG[4:6,],5)
#[1] "simulatedFS was generated"
 dim(simulatedFS)
#[1]   15 1000
# The first 5 individuals are progeny of mom 1 and dad 1, the second 5 individuals
# are progeny of mom 2 and dad 2, and so on.
```

`simulHS` *Function to simulate half sib progeny genotypes*

Description

Simulate half sib progeny from one genotyped parent assuming a random genotype for the other parental. We assume that these are diploid organisms.

Usage

```
simulHS(x, Nprogeny)
```

Arguments

`x`	genotype matrix of a set of moms
`Nprogeny`	number of progeny's genotypes to simulate for each mom

Details

The function assume: a diploid organism; mendelian segregation of alleles; and independent segregation.

Value

a matrix of dimensions (nrow(x)*Nprogeny) x ncol(x)

Author(s)

Martin Nahuel Garcia <orcid:0000-0001-5760-986X>

References

Wu, R., Ma, C., & Casella, G. (2007). Statistical genetics of quantitative traits: linkage, maps and QTL. Springer Science & Business Media.

Examples

```
#' #simulate 100 individuals and 1000 SNPs
set.seed(123)
simGeno(100, 1000)
#[1] "simG was generated"
#simulate the genotype of 3 sets 5 HS (one set by mom)
simulHS(simG[1:3,],5)
#[1] "simulatedHS was generated"
dim(simulatedHS)
#[1]   15 1000
```

`simulN` *Function to simulate a SNP matrix, a phenotypic trait and the effects of associated SNPs.*

Description

This function simulate a SNP matrix (coded as 0, 1, 2) and traits with a selected number of QTLs and their effects that will be sampled from a Normal distribution.

Usage

```
simulN(Nind, Nmarkers, Nqtl, Esigma, Pmean, Perror)
```

Arguments

`Nind`	number of individuals to simulate.
`Nmarkers`	number of SNP markers to generate.
`Nqtl`	number of QTLs controlling the trait.
`Esigma`	standard deviation of effects with distribution N~(0,Esigma^2).
`Pmean`	phenotype mean.
`Perror`	standard deviation of error (portion of phenotype not explained by genomic information).

Details

Genotypic data is simulated as the round value sampled from an uniform distribution with interval (-.5,2.5). Phenotypic data are obtained as a linear function defined by:

$$y = Pmean + \sum QTN * Meffects + Perror$$

Value

An object of class list containing the SNP matrix, the trait, the markers associated and their effects.

`geno`	SNP matrix generated.
`pheno`	vector with the trait values simulated.
`QTN`	column in the SNP matrix with the SNP associated.
`Meffects`	effects of the associated SNPs.

Note

The genotype is simulated in the same way of simGeno function. The trait, QTLs and their effects are simulated in the same way of simPheno function.

Author(s)

Martin Nahuel Garcia <orcid:0000-0001-5760-986X>

References

Wu, R., Ma, C., & Casella, G. (2007). Statistical genetics of quantitative traits: linkage, maps and QTL. Springer Science & Business Media.

See Also

simGeno, simPheno, simulU

Examples

```
set.seed(123)
simulN(100, 1000, 50, .9, 12, .5)
#[1] "nsimout was generated"
str(nsimout)
#List of 4
#$ geno    : num [1:100, 1:1000] 0 2 1 2 2 0 1 2 1 1 ...
#$ pheno   : num [1:100, 1] 25.4 21.6 16 13.8 19.4 ...
#$ QTN     : int [1:50] 568 474 529 349 45 732 416 51 413 514 ...
#$ Meffects: num [1:50] 0.2696 -0.1552 1.0192 0.0209 1.2023 ...
```

`simulU` *Function to simulate a random SNP matrix, phenotype and QTLs with their effects*

Description

This function simulate a SNP matrix (coded as 0, 1, 2) and traits with a selected number of QTLs and their effects that will be sampled from a Uniform distribution.

Usage

```
simulU(Nind, Nmarkers, Nqtl, Pmean, Perror)
```

Arguments

`Nind`	number of individuals to simulate.
`Nmarkers`	number of SNP markers to generate.
`Nqtl`	number of QTLs controlling the trait.
`Pmean`	phenotype mean.
`Perror`	standard deviation of error (portion of phenotype not explained by genomic information).

Value

An object of class list containing the SNP matrix, the trait, the markers associated and their effects.

`geno`	SNP matrix generated.
`pheno`	vector with the trait values simulated.
`QTN`	column in the SNP matrix with the SNP associated.
`Meffects`	effects of the associated SNPs.

Author(s)

Martin Nahuel Garcia <orcid:0000-0001-5760-986X>

References

Wu, R., Ma, C., & Casella, G. (2007). Statistical genetics of quantitative traits: linkage, maps and QTL. Springer Science & Business Media.

See Also

simGeno, simulN

Examples

```
set.seed(123)
simulU(100, 1000, 50, 12, .5)
#[1] "usimout was generated"
str(usimout)
#List of 4
#$ geno    : num [1:100, 1:1000] 0 2 1 2 2 0 1 2 1 1 ...
#$ pheno   : num [1:100, 1] 10.3 14.7 11.8 10.2 13.1 ...
#$ QTN     : int [1:50] 568 474 529 349 45 732 416 51 413 514 ...
#$ Meffects: num [1:50] 0.2355 0.0158 -0.1369 -0.1246 0.7426 ...
```

Index