

Package ‘LDcorSV’

January 20, 2025

Type Package

Title Linkage Disequilibrium Corrected by the Structure and the Relatedness

Version 1.3.3

Imports parallel

Author David Desrousseaux, Florian Sandron, Aurélie Siberchicot, Christine Cierco-Ayrolles and Brigitte Mangin

Maintainer Aurélie Siberchicot <aurelie.siberchicot@univ-lyon1.fr>

Description Four measures of linkage disequilibrium are provided: the usual r^2 measure, the r^2_S measure (r^2 corrected by the structure sample), the r^2_V (r^2 corrected by the relatedness of genotyped individuals), the r^2_{VS} measure (r^2 corrected by both the relatedness of genotyped individuals and the structure of the sample).

License GPL (≥ 2)

LazyLoad yes

Encoding UTF-8

Suggests testthat

NeedsCompilation no

Repository CRAN

Date/Publication 2020-08-26 09:00:07 UTC

Contents

LDcorSV-package	2
data.test	2
Info.Locus	3
Inv.proj.matrix.sdp	4
LD.Measures	5
Measure.R2	6
Measure.R2S	7
Measure.R2V	9
Measure.R2VS	10

Index**12**

LDcorSV-package	<i>LDcorSV</i>
-----------------	----------------

Description

The package provides a set of functions which aim is to propose four measures of linkage disequilibrium: - Measure.R2: the usual r^2 measure.

- Measure.R2S: r^2 corrected by the structure of the sample (r_S^2).

- Measure.R2V: r^2 corrected by the relatedness of genotyped individuals (r_V^2).

- Measure.R2VS: r^2 corrected by both the relatedness of genotyped individuals and the structure of the sample (r_{VS}^2).

- LD.Measures: this function computes the four measures of linkage disequilibrium (r^2 , r_V^2 , r_S^2 and r_{VS}^2) for a set of loci and gives extra information about them.

Author(s)

David Desrousseaux, Florian Sandron, Aurélie Siberchicot, Christine Cierco-Ayrolles and Brigitte Mangin

Maintainer: aurelie.siberchicot@univ-lyon1.fr

References

Mangin, B., Siberchicot, A., Nicolas, S., Doligez, A., This, P., Cierco-Ayrolles, C. (2012). Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. *Heredity*, 108 (3), 285-291. DOI: 10.1038/hdy.2011.73

data.test	<i>data.test</i>
-----------	------------------

Description

data.test is a list of 3 elements:

- Geno: allelic doses of 20 markers on a chromosome of 91 *Vitis vinifera* plants.

- V.WAIS: kinship matrix of 91 plants of *Vitis vinifera*.

- S.2POP: structure population matrix of 91 plants of *Vitis vinifera* in two sub-populations.

Usage

data(data.test)

Format

A list containing the following components:

- Geno: matrix (91 x 20) of numerical values
- V.WAIS: matrix (91 x 91) of numerical values
- S.2POP: matrix (91 x 1) of numerical values

Examples

```
data(data.test)

# Allelic doses of 20 markers on a chromosome of 91 Vitis vinifera plants
Geno <- data.test[[1]]
Geno

# Kinship matrix of 91 plants of Vitis vinifera
V.WAIS <- data.test[[2]]
V.WAIS

# Structure population matrix of 91 plants of Vitis vinifera
# in two sub-populations
S.2POP <- data.test[[3]]
S.2POP
```

Info.Locus

Information on loci

Description

For a locus, this function computes the minor allelic frequency, the frequency of heterozygous genotypes and the missing value frequency.

Usage

```
Info.Locus(locus, data="G")
```

Arguments

locus	Numeric vector of allelic doses.
data	Value equal to "G" or "H" depending on the type of data (Genotype or Haplotype). Default value is "G".

Value

The returned value is a numeric vector of three values which are respectively the minor allelic frequency, the frequency of heterozygous genotypes (NA if haplotype data) and the missing value frequency.

Author(s)

David Desrousseaux, Florian Sandron, Aurélie Siberchicot, Christine Cierco-Ayrolles and Brigitte Mangin

Examples

```
data(data.test)
Geno <- data.test[[1]]
info <- apply(Geno, 2, Info.Locus)
info
```

Inv.proj.matrix.sdp *Inv.proj.matrix.sdp*

Description

This function computes the Moore-Penrose pseudo-inverse of a symmetric matrix. A single value decomposition is performed, the non positive eigen values are set to zero, then the pseudo-inverse is computed.

Usage

```
Inv.proj.matrix.sdp(matrix)
```

Arguments

matrix symmetric matrix

Value

The returned value is the pseudo-inverse matrix.

Author(s)

David Desrousseaux, Florian Sandron, Aurélie Siberchicot, Christine Cierco-Ayrolles and Brigitte Mangin

Examples

```
data(data.test)
V.WAIS <- data.test[[2]]
Inv.V.WAIS <- Inv.proj.matrix.sdp(V.WAIS)
Inv.V.WAIS
```

Description

This function estimates for a set of loci:

- the usual measure of linkage disequilibrium (r^2)
- the measure of linkage disequilibrium corrected by the structure of the sample (r_S^2).
- the measure of linkage disequilibrium corrected by the relatedness of genotyped individuals (r_V^2).
- the measure of linkage disequilibrium corrected by both, the relatedness of genotyped individuals and the structure of the sample (r_{VS}^2).

This function gives extra informations on the studied loci.

Usage

```
LD.Measures(donnees, V = NA, S = NA, data = "G", supinfo = FALSE, na.presence=TRUE)
```

Arguments

donnees	Numeric matrix (N x M), where N is the number of genotypes (or haplotypes) and M is the number of markers. Matrix values are the allelic doses: - (0,1,2) for genotypes. - (0,1) for haplotypes. Row names correspond to the ID of individuals. Column names correspond to the ID of markers. Missing values are allowed.
V	Numeric matrix (N x N), where N is the number of genotypes (or haplotypes). Matrix values are coefficients of genetic variance-covariance between every pair of individuals. Row and column names must correspond to the ID of individuals. No missing value.
S	Numeric matrix (N x (1-P)), where N is the number of genotypes (or haplotypes) and P the number of sub-populations. Matrix values are the probabilities (between 0 and 1) for each genotype (or haplotype) to belong to each sub-populations. Row names must correspond to the ID of individuals. Column names correspond to the ID of sub-populations. The matrix must be inversible, if the structure is with P sub-populations, only P-1 columns are expected. No missing value.
data	Value equal to "G" or "H" depending on the type of data (Genotype or Haplotype). Default value is "G".
supinfo	Boolean indicating whether you wish to get information about the loci. If supinfo=TRUE, for each locus, the Minor Allelic Frequency (MAF), the frequency of heterozygous genotypes (only if the data are genotypes) and the missing value frequency are computed. By default, supinfo=FALSE.
na.presence	Boolean indicating the presence of missing values in data. If na.presence=FALSE (no missing data), computation of r_V^2 and r_{VS}^2 is largely optimized. By default, na.presence=TRUE.

Value

The returned value is a dataframe of size $(M(M-1))/2$ rows and C columns, where M is the number of markers and C is a number between 3 and 12 depending on options chosen by user. The first three columns contain respectively the name of the first marker, the name of the second marker and the estimated value of the usual measure of linkage disequilibrium (r^2) between these two markers.

If only V is different from NA, the fourth column contains the estimated value of the measure of linkage disequilibrium corrected by the relatedness of genotyped individuals ($r_{V'}^2$).

If only S is different from NA, the fourth column contains the estimated value of the measure of linkage disequilibrium corrected by relatedness corrected by the structure of the sample (r_S^2).

If V and S are simultaneously different from NA, the fourth, fifth and sixth columns respectively contain the estimated values of $r_{V'}^2$, r_S^2 and $r_{V'S}^2$ (r^2 corrected by both the relatedness of genotyped individuals and the structure of the sample).

If `Supinfo=TRUE`, then the last six columns contain information for both loci : the MAF, the frequency of heterozygous genotype (NA if haplotype data) and the missing value frequency.

Author(s)

David Desrousseaux, Florian Sandron, Aurélie Siberchicot, Christine Cierco-Ayrolles and Brigitte Mangin

References

Mangin, B., Siberchicot, A., Nicolas, S., Doligez, A., This, P., Cierco-Ayrolles, C. (2012). Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. *Heredity*, 108 (3), 285-291. DOI: 10.1038/hdy.2011.73

Examples

```
data(data.test)
Geno <- data.test[[1]]
V.WAIS <- data.test[[2]]
S.2POP <- data.test[[3]]

LD <- LD.Measures(Geno, V = V.WAIS, S = S.2POP, data = "G", supinfo = TRUE, na.presence = TRUE)
head(LD)
```

Measure.R2

r² measure

Description

This function estimates the usual measure of linkage disequilibrium (r^2) between two loci.

Usage

```
Measure.R2(biloci, na.presence = TRUE)
```

Arguments

- biloci** Numeric matrix (N x 2), where N is the number of genotypes (or haplotypes)
Matrix values are the allelic doses:
- (0,1,2) for genotypes.
- (0,1) for haplotypes.
Rows names correspond to the ID of individuals.
Columns names correspond to the ID of markers.
- na.presence** Boolean indicating the presence of missing values in data.
If na.presence=FALSE (no missing data), computation of r_V^2 and r_{VS}^2 is largely optimized.
By default, na.presence=TRUE.

Value

The returned value is the estimated value of the usual measure of linkage disequilibrium (r^2) or NA if less than 5 individuals have non-missing data at both loci

Author(s)

David Desrousseaux, Florian Sandron, Aurélie Siberchicot, Christine Cierco-Ayrolles and Brigitte Mangin

References

Hill, W.G, Robertson, A. (1968). Linkage disequilibrium in finite populations. Theoretical and Applied Genetics, 38, 226-231. DOI: 10.1007/BF01245622

Examples

```
data(data.test)
Geno <- data.test[[1]]
Measure.R2(Geno)
```

Measure.R2S

r^2_S measure

Description

This function estimates the novel measure of linkage disequilibrium which is corrected by the structure of the sample.

Usage

```
Measure.R2S(biloci, struc, na.presence=TRUE)
```

Arguments

- biloci** Numeric matrix (N x 2), where N is the number of genotypes (or haplotypes)
Matrix values are the allelic doses:
- (0,1,2) for genotypes.
- (0,1) for haplotypes.
Row names correspond to the ID of individuals.
Column names correspond to the ID of markers.
- struc** Numeric matrix (N x (P-1)), where N is the number of genotypes (or haplotypes)
and P the number of sub-populations.
Matrix values are the probabilities for each genotypes (or haplotypes) to belong
to each sub-populations.
Row names must correspond to the ID of individuals and must be ranged as in
the biloci matrix.
Column names correspond to the ID of sub-populations.
The matrix must be inversible, if the structure is with P sub-populations, only
P-1 columns are expected.
No missing value.
- na.presence** Boolean indicating the presence of missing values in data.
If na.presence=FALSE (no missing data), computation of r_V^2 and r_{VS}^2 is largely
optimized.
By default, na.presence=TRUE.

Value

The returned value is the estimated value of the measure of linkage disequilibrium corrected by the structure of the sample or NA if less than 5 individuals have non-missing data at both loci.

Author(s)

David Desrousseaux, Florian Sandron, Aurélie Siberchicot, Christine Cierco-Ayrolles and Brigitte Mangin

References

Mangin, B., Siberchicot, A., Nicolas, S., Doligez, A., This, P., Cierco-Ayrolles, C. (2012). Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. *Heredity*, 108 (3), 285-291. DOI: 10.1038/hdy.2011.73

Examples

```
data(data.test)
Geno <- data.test[[1]]
S.2POP <- data.test[[3]]
Measure.R2S(Geno, S.2POP)
```

Measure.R2V	r^2_V measure
-------------	-----------------

Description

This function estimates the novel measure of linkage disequilibrium which is corrected by the relatedness of genotyped individuals.

Usage

```
Measure.R2V(biloci, V, na.presence=TRUE, V_inv=NULL)
```

Arguments

biloci	Numeric matrix (N x 2), where N is the number of genotypes (or haplotypes). Matrix values are the allelic doses: - (0,1,2) for genotypes. - (0,1) for haplotypes. Row names correspond to the ID of individuals. Column names correspond to the ID of markers.
V	Numeric matrix (N x N), where N is the number of genotypes (or haplotypes). Matrix values are coefficients of genetic covariance for each pair of individuals. Rows and columns names must correspond to the ID of individuals and must be ranged in the same order as in the biloci matrix. No missing value.
na.presence	Boolean indicating the presence of missing values in data. If na.presence=FALSE (no missing data), computation of r^2_V and r^2_{VS} is largely optimized. By default, na.presence=TRUE.
V_inv	Should stay NULL.

Value

The returned value is the estimated value of the measure of linkage disequilibrium corrected by the relatedness of genotyped individuals or NA if less than 5 individuals have non-missing data at both loci.

Author(s)

David Desrousseaux, Florian Sandron, Aurélie Siberchicot, Christine Cierco-Ayrolles and Brigitte Mangin

References

Mangin, B., Siberchicot, A., Nicolas, S., Doligez, A., This, P., Cierco-Ayrolles, C. (2012). Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. *Heredity*, 108 (3), 285-291. DOI: 10.1038/hdy.2011.73

Examples

```
data(data.test)
Geno <- data.test[[1]]
V.WAIS <- data.test[[2]]
Measure.R2V(Geno, V.WAIS)
```

Measure.R2VS	<i>r²_{VS} measure</i>
--------------	---

Description

This function estimates the novel measure of linkage disequilibrium which is corrected by both the relatedness of genotyped individuals and the structure of the sample.

Usage

```
Measure.R2VS(biloci, V, struc, na.presence = TRUE, V_inv = NULL)
```

Arguments

biloci	Numeric matrix (N x 2), where N is the number of genotypes (or haplotypes) Matrix values are the allelic doses: - (0,1,2) for genotypes. - (0,1) for haplotypes. Row names correspond to the ID of individuals. Column names correspond to the ID of markers.
V	Numeric matrix (N x N), where N is the number of genotypes (or haplotypes). Matrix values are coefficients of genetic variance-covariance for every pair of individuals. Row and column names must correspond to the ID of individuals and must be ranged as in the biloci matrix. No missing value.
struc	Numeric matrix (N x (P-1)), where N is the number of genotypes (or haplotypes) and P the number of sub-populations. Matrix values are the probabilities (between 0 and 1) for each genotypes (or haplotypes) to belong to each sub-populations. Row names must correspond to the ID of individuals and must be ranged as in the biloci matrix. Column names correspond to the ID of sub-populations. The matrix must be inversible, if the structure is with P sub-populations, only P-1 columns are expected. No missing value.
na.presence	Boolean indicating the presence of missing values in data. If na.presence=FALSE (no missing data), computation of r_V^2 and r_{VS}^2 is largely optimized. By default, na.presence=TRUE.
V_inv	Should stay NULL

Value

The returned value is the estimated value of the linkage disequilibrium measure corrected by both the relatedness of genotyped individuals and the structure of the sample or NA if less than 5 individuals have non-missing data at both loci.

Author(s)

David Desrousseaux, Florian Sandron, Aurélie Siberchicot, Christine Cierco-Ayrolles and Brigitte Mangin

References

Mangin, B., Siberchicot, A., Nicolas, S., Doligez, A., This, P., Cierco-Ayrolles, C. (2012). Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. *Heredity*, 108 (3), 285-291. DOI: 10.1038/hdy.2011.73

Examples

```
data(data.test)
Geno <- data.test[[1]]
V.WAIS <- data.test[[2]]
S.2POP <- data.test[[3]]
Measure.R2VS(Geno, V.WAIS, S.2POP)
```

Index

- * **datasets**

- data.test, [2](#)

- * **package**

- LDcorSV-package, [2](#)

data.test, [2](#)

Info.Locus, [3](#)

Inv.proj.matrix.sdp, [4](#)

LD.Measures, [5](#)

LDcorSV (LDcorSV-package), [2](#)

LDcorSV-package, [2](#)

Measure.R2, [6](#)

Measure.R2S, [7](#)

Measure.R2V, [9](#)

Measure.R2VS, [10](#)