

Package ‘Relatedness’

November 17, 2017

Type Package

Title Maximum Likelihood Estimation of Relatedness using EM Algorithm

Version 2.0

Date 2017-11-17

Author Fabien Laporte, Tristan Mary-Huard

Maintainer Fabien Laporte <fabien.laporte@inra.fr>

Description Inference of relatedness coefficients from a bi-allelic genotype matrix using a Maximum Likelihood estimation, Laporte, F., Charcosset, A. and Mary-Huard, T. (2017) <doi:10.1111/biom.12634>.

License AGPL-3

NeedsCompilation yes

Repository CRAN

Date/Publication 2017-11-17 10:51:45 UTC

R topics documented:

Relatedness-package	2
BoucleEMacc	2
Cross	3
Delta	3
Frequencies	4
Genotype	4
RelCoef	5
RelCoefLine	7
RelComb	8

Index	10
--------------	-----------

Relatedness-package *Maximum Likelihood Estimation of Relatedness using EM Algorithm*

Description

Inference of relatedness coefficients from a bi-allelic genotype matrix using a Maximum Likelihood estimation, Laporte, F., Charcosset, A. and Mary-Huard, T. (2017) <doi:10.1111/biom.12634>.

Details

This package infers the relatedness distribution coefficients for all couple of individuals in a set from their genotype, provided in a bi-allelic genotype matrix. The main function is 'RelCoef' which infers those coefficients. The arguments of this function are a genotype matrix for individuals and an frequency matrix that displays the allelic frequency at each marker in each population. Alternatively, a parental genotype matrix and a crossing matrix can be used. Additional information about structure membership can also be provided via a ParentPop vector (for more details see the help of 'RelCoef'). The main matrix is written with C language, make sure you can use this code.

Author(s)

Fabien Laporte, Tristan Mary-Huard

Maintainer: Fabien Laporte <fabien.laporte@inra.fr>

Examples

```
require('Relatedness')
data(Genotype)
data(Frequencies)
data(Cross)
RelCoef(IndividualGenom=matrix(0,ncol=0,nrow=0),ParentalLineGenom=Genotype,
Freq=Frequencies,Crossing=Cross,ParentPop=rep(1,20),Phased=TRUE,NbCores=2)
```

BoucleEMacc

C code for the EM

Description

A C code used in the function 'RelCoeff'.

Cross	<i>Crossing matrix</i>
-------	------------------------

Description

The crossing matrix for the example.

Usage

```
data("Cross")
```

Format

The format is: int [1:5, 1:2] 1 2 3 4 5 6 7 8 9 10

Examples

```
data(Cross)
head(Cross)
```

Delta	<i>List of Relatedness Coefficients</i>
-------	---

Description

A list of relatedness coefficients obtained with the example of *RelCoef*.

Usage

```
data("Delta")
```

Format

The format is: List of 15 \$ Delta1 : num [1:4, 1:4] 0.1872 0 0 0 0.0199 ... \$ Delta2 : num [1:4, 1:4] 0.00 0.00 0.00 0.00 3.55e-05 ... \$ Delta3 : num [1:4, 1:4] 0 0 0 0 0.0472 ... \$ Delta4 : num [1:4, 1:4] 0 0 0 0 0.0322 ... \$ Delta5 : num [1:4, 1:4] 0 0 0 0 0.0871 ... \$ Delta6 : num [1:4, 1:4] 0 0 0 0 0.0395 ... \$ Delta7 : num [1:4, 1:4] 0 0 0 0 0.0429 ... \$ Delta8 : num [1:4, 1:4] 0 0 0 0 0.0386 ... \$ Delta9 : num [1:4, 1:4] 0.8128 0 0 0 0.0202 ... \$ Delta10: num [1:4, 1:4] 0 0 0 0 0.000731 ... \$ Delta11: num [1:4, 1:4] 0 0 0 0 0.028 ... \$ Delta12: num [1:4, 1:4] 0 0 0 0 0.0849 ... \$ Delta13: num [1:4, 1:4] 0 0 0 0 0.0174 ... \$ Delta14: num [1:4, 1:4] 0 0 0 0 0.0437 ... \$ Delta15: num [1:4, 1:4] 0 0 0 0 0.498 ...

Examples

```
data(Delta)
print(Delta$Delta7)
```

 Frequencies

Allele Frequencies

Description

The allele frequencies matrix for the example with 5000 markers and one population.

Usage

```
data("Frequencies")
```

Format

The format is: num [1:5000, 1:2] 0.268 0.786 0.804 0.238 0.235 ...

Examples

```
data(Frequencies)
head(Frequencies)
```

Genotype

Genotype Matrix

Description

The Parental Line Genom matrix for the example with 10 parental lines genotyped with 5000 markers.

Usage

```
data("Genotype")
```

Format

The format is: num [1:5000, 1:10] 0 1 0 0 0 1 1 0 1 1 ...

Examples

```
data(Genotype)
head(Genotype)
```

RelCoef *Relatedness Coefficients Estimation for individuals*

Description

This function performs Maximum Likelihood estimation for the relatedness coefficients between individuals based on a bi-allelic genotype matrix. Alternatively, a parental genotype matrix and a crossing matrix can be used. In that case information about structure can also be taken into account via a ParentPop vector.

Usage

```
RelCoef(IndividualGenom = matrix(0, nrow=0, ncol=0),
        ParentalLineGenom = matrix(0, nrow=0, ncol=0),
        Freq = matrix(0, nrow=0, ncol=0),
        Crossing = matrix(0, nrow=0, ncol=0), ParentPop = rep(0,0),
        Combination = list(), Phased = FALSE, Details = FALSE,
        NbInit = 5, Prec = 10^(-4), NbCores = NULL)
```

Arguments

IndividualGenom	Genotype matrix of individuals. Each individual is described by 2 columns. Each row corresponds to a marker. Entries of matrix IndividualGenom should be either 0 or 1. Either IndividualGenom or ParentalLineGenom has to be provided.
ParentalLineGenom	Genotype matrix of parental lines. Each parental line is described by one column with rows corresponding to markers. Entries of ParentalLineGenome should be either 0 or 1.
Freq	Allelic frequencies for allele 1 at each markers and for all populations (one column per population, one line per marker).
Crossing	Required when argument ParentalLineGenom is provided. A 2-column matrix where each row corresponds to a crossing between 2 parents. Parents should be numbered according to their order of appearance in the ParentalLineGenom matrix.
ParentPop	Only available if ParentalLineGenom is displayed. A vector of numbers corresponding to population membership for the parental lines.
Combination	If provided, a list of vector with two components. The jth vector is composed with the number of the first hybrid and the number of the second hybrid of the jth couple to study.
Phased	A Boolean with value TRUE if observations are phased.
Details	A Boolean variable. If TRUE, the relatedness mode graph is displayed.
NbInit	Number of initial values for the EM algorithm.
Prec	Convergence precision parameter for the EM algorithm.
NbCores	Number of cores used by the algorithm (Default is the number of cores available minus one). Only available for linux and Mac.

Details

Argument `IndividualGenom` should be used if the available data consist in genotypic information only. By default the data are assumed to be unphased and the function returns 9 relatedness coefficients. If data are phased, use argument `Phased = TRUE` to obtain the 15 relatedness coefficients. Note that in that case the ordering of the 2 columns per individual in `IndividualGenome` does matter. Alternatively, if the genotyped individuals are hybrids resulting from the crossing of parental lines (or combinations of parental gametes), it is possible to provide a `ParentalLineGenom` and a `Crossing` matrix directly. Additionally, the population membership of the parents can be provided via argument `ParentPop`. Whatever the arguments used to enter the genotypic data, the allelic frequencies of the markers have to be provided using argument `Freq`. Arguments `NbInit` and `Prec` are tuning parameters for the EM algorithm used for likelihood maximization.

Value

By default, relatedness coefficients are displayed for all couple of genotyped individuals (or hybrids). In that case the function returns a list of matrices, each corresponding to a specific relatedness coefficients (details about relatedness coefficients can be obtained by displaying the relatedness mode graph with argument `Details`). Element (i,j) of matrix k corresponds to the k th estimated relatedness coefficient for the couple of individuals i and j . Alternatively, if a list of couples is specified with argument `Combination`, the function returns a list of vectors (each vector corresponding to an relatedness coefficient). In that case element i of vector k corresponds to the k th relatedness coefficient of the i th couple specified in `Combination`.

Warning

In absence of population structure, some relatedness coefficients are not identifiable. Since an EM algorithm is run for each couple of individuals, the procedure can be time consuming for large panels.

Author(s)

Fabien Laporte, 'UMR Genetique Quantitative et Evolution' INRA France.

Examples

```
require('Relatedness')
data(Genotype)
data(Frequencies)
data(Cross)
RelatednessCoefficient <- RelCoef(IndividualGenom=matrix(0,ncol=0,nrow=0),
  ParentalLineGenom=Genotype,
  Freq=Frequencies,Crossing=Cross,
  ParentPop=rep(1,8),Phased=TRUE,NbCores=2)
print(RelatednessCoefficient$Delta3)
```

Description

This function performs Maximum Likelihood estimation for the relatedness coefficients between lines based on a bi-allelic genotype matrix.

Usage

```
RelCoefLine(LineGenom = matrix(0,nrow=0,ncol=0),  
            Freq = matrix(0,nrow=0,ncol=0),  
            LinePop = rep(0,0),  
            Combination = NULL,  
            NbInit = 5, Prec = 10-4, NbCores = NULL)
```

Arguments

LineGenom	Genotype matrix of lines. Each line is described by 1 column. Each row corresponds to a marker. Entries of matrix Genotype should be either 0 or 1.
Freq	Allelic frequencies for allele 1 at each markers and for all populations (one column per population, one line per marker).
LinePop	A vector of numbers corresponding to population membership for the parental lines.
Combination	If provided, a list of vector with two components. The jth vector is composed with the number of the first hybrid and the number of the second hybrid of the jth couple to study.
NbInit	Number of initial values for the EM algorithm.
Prec	Convergence precision parameter for the EM algorithm.
NbCores	Number of cores used by the algorithm (Default is the number of cores available minus one). Only available for linux and Mac.

Value

By default, relatedness coefficients are displayed for all couple of genotyped lines. In that case the function returns a matrix corresponding to the Simple Relatedness Coefficient, i.e. the probability that each couple of lines are related. Element (i,j) of the matrix corresponds to the estimated relatedness coefficient for the couple of lines i and j. Alternatively, if a list of couples is specified with argument Combination, the function returns a list of coefficients (each coefficient corresponding to an relatedness coefficient). In that case element i of the list corresponds to the estimated relatedness coefficient of the ith couple specified in Combination.

Author(s)

Fabien Laporte, 'UMR Genetique Quantitative et Evolution' INRA France.

Examples

```
require('Relatedness')
data(Genotype)
data(Frequencies)
data(Cross)
SimpleRelatedness <- RelCoefLine(LineGenom=Genotype,Freq=Frequencies,
LinePop=rep(1,8),NbCores=2)
print(SimpleRelatedness)
```

RelComb

Computation of Linear Combination of Relatedness Coefficients

Description

Compute any relatedness synthetic criterion based on a linear combination of the relatedness coefficients.

Usage

```
RelComb(Combination, Delta,
Crossing = matrix(0, nrow = 0, ncol = 0),
ParentPop = rep(0, 0),
ShowIdentifiable = TRUE)
```

Arguments

Combination	A vector, with length identical to the length of Delta where the kth element corresponds to the coefficient for the kth relatedness coefficient, or alternatively a character in the following list: 'simple relatedness', 'double relatedness', 'first inbreeding', 'second inbreeding', 'double inbreeding'.
Delta	A list of matrices, each corresponding to a specific relatedness coefficients. Element (i,j) of matrix k corresponds to the kth estimated relatedness coefficient for the couple of individuals i and j.
Crossing	A 2-column matrix where each row corresponds to a crossing between 2 parents. Parents should be numbered.
ParentPop	A vector of numbers corresponding to population membership for the parental lines.
ShowIdentifiable	A boolean describing whether the combination should be displayed only for identifiable cases. Default value is TRUE.

Details

The function can be applied to a list of relatedness coefficients - as produced by the RelCoeff function - to compute any synthetic criterion based on a linear combination of the relatedness coefficients, for all couples. Additional information Crossing and ParentPop are required if they were used in the RelCoeff function to obtain the relatedness coefficients. The function automatically checks the identifiability of the combination to be evaluated. Several classical genetic criteria are implemented by default and can be computed using the Csuros argument. Alternatively, the user can provide a vector of coefficients to be applied through the Combination argument.

Value

If identifiability is satisfied for all pairs of individuals, all criteria are computed and returned in a matrix. If identifiability is not guaranteed for all pairs, the function will return a matrix with NA entries for the potentially non-estimable pairs. This by-default behavior can be bypassed if required by setting ShowIdentifiable to FALSE.

Author(s)

Fabien Laporte, 'UMR Genetique Quantitative et Evolution' INRA France.

References

Csuros M (2014) Non-identifiability of identity coefficients at biallelic loci. *Theoretical Population Biology* 92: 22-24.

Examples

```
require('Relatedness')
data(Delta)
RelatednessComb <- RelComb(Combination='simple relatedness', Delta, ShowIdentifiable = TRUE)
print(RelatednessComb)
```

Index

*Topic **Identifiability**

RelComb, [8](#)

*Topic **Relatedness**

Relatedness-package, [2](#)

RelCoef, [5](#)

RelCoefLine, [7](#)

RelComb, [8](#)

*Topic **datasets**

Cross, [3](#)

Delta, [3](#)

Frequencies, [4](#)

Genotype, [4](#)

BoucleEMacc, [2](#)

Cross, [3](#)

Delta, [3](#)

Frequencies, [4](#)

Genotype, [4](#)

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

Relatedness-package, [2](#)

RelCoef, [5](#)

RelCoefLine, [7](#)

RelComb, [8](#)