Package 'Demerelate'

March 3, 2017

Version 0.9-3

Date 2017-03-02

Title Functions to Calculate Relatedness on Diploid Genetic Data

Maintainer Philipp Kraemer <Philipp.Kraemer@uni-oldenburg.de>

Imports grDevices, graphics, stats, utils, methods, fts, sfsmisc, vegan, mlogit

Depends R (>= 2.15.0)

Description Functions to calculate pairwise relatedness on diploid genetic datasets. Different estimators for relatedness can be combined with information on geographical distances. Information on heterozygosity, allele- and genotype diversity as well as genetic F-statistics are provided for each population.

License GPL (>= 2)

URL https://www.r-project.org

NeedsCompilation no

Author Philipp Kraemer [aut, cre], Gabriele Gerlach [aut]

Repository CRAN

Date/Publication 2017-03-03 18:01:40

R topics documented:

`Demerelate-package` *Demerelate — Algorithms to estimate pairwise relatedness within populations based on allele sharing*

Description

The package Demerelate provides several functions to calculate relatedness of individuals based on diploid genetic markers. Following indices can be calculated:

1. B_{xy} (number of alleles shared) as described in Li and Horvitz 1953 (Bxy).
2. S_{xy} (number of alleles shared) as described in Lynch 1988 (Sxy).
3. M_{xy} (genotype sharing) as described in Blouin et al. 1996 (Mxy).
4. Li is based on the equations from Li et al. 1993 (Li).
5. The estimator r_{xy} based on Queller and Goodnight 1989 adapted to pairwise comparisons as described in Oliehoek et al. 2006 (rxy).
6. l_{xy} is calculated based on Lynch and Ritland 1999 (lxy, lxy.w).
7. The estimator $loiselle$ is based on Loiselle et al. 1995 (loiselle).
8. The estimator $wang.fin$ is based on Wang 2002 for a finite sample (wang.fin, wang.fin.w).
9. The estimator $wang$ is based on Wang 2002 including bias correction for sample size (wang, wang.w).
10. The estimator $ritland$ is based on Ritland 1996 (ritland).
11. The estimator $morans.fin$ is based on Hardy and Vekemans 1999 ommiting correction for sample size (morans.fin).
12. The estimator $morans$ is based on Hardy and Vekemans 1999 with correction for sample size bias (morans.fin, morans.w).

Detailed information can be found in the references or the descriptions of either subfunction.

For each estimator populations of randomized offsprings and randomized non related individuals are created from a reference population. By default, all empirical data will be used as a reference; however, functions can be forced to use specific data as reference. Be careful when using several populations or specific references. All populations need to be free of any pronounced population structure. References must contain every allele of the tested populations! Any violation of this may lead to strange results or interruptions during calculations. Based on these reference populations all

allele sharing indices and relatedness estimators can be analysed using a logistic regression model. Thresholds for individuals being full-siblings, half-siblings or non-related for each population are calculated. χ^2 statistics are used to calculate whether populations contain more siblings than expected in randomly drawn populations of the same size as the empirical population. Additionally, for different sets of references (full siblings, half siblings and non-related) in comparison to all empirical populations pairwise T-tests are used to calculate whether populations are significantly different in mean relatedness. These results may be biased if the statistical conditions for performing a T-test are not given in the data. You should take care of that before drawing conclusions from these results. Therefore pairwise relatedness for empirical populations as well as randomized populations are passed to the local directory and can be used for downstream analyses. Several plots from each analysis can be exported as .pdf files for visualization of relatedness within populations. For every population geographical locations can be combined with genetic data to find out whether geographic distance has an effect on relatedness. A mantel correlation (Oksanen et al. 2013) plotted via scatterplot including relatedness thresholds from logistic regression are plotted for visualization.

Additionally, basic information on datasets can be calculated, for example allele and genotype frequencies, F_{is} values within populations or analysis on relative informativity of loci for relatedness analysis. For a quick overview, while omitting statistics, all allele sharing indices and relatedness estimators are calculated for every pairwise combination in one analysis.

Details

Package: Demerelate
Version: 0.9-3
Date: 2017-03-02
Depends: R (>= 2.15.0), Formula
Imports: fts, mlogit, sfsmisc, vegan
Suggests: MASS
License: GPL (>= 2)
URL: http://www.r-project.org

Demerelate	– Head function to calculate pairwise relatedness
Loci.test	– Function to analyse datasets of diploid genetic information
Emp.calc	– A quick overview of relatedness by omitting statistics
F.stat	– Function to calculate F_{is} values for single populations
demerelpop	– Example data set from randomized diploid markers for three populations
demerelref	– Example data set from randomized diploid markers as reference
demereldist	– Example data set from randomized relative geographic information

Author(s)

Philipp Kraemer and Gabriele Gerlach

Maintainer: Philipp Kraemer, <philipp.kraemer@uni-oldenburg.de>

References

Armstrong, W. (2012) fts: R interface to tslib (a time series library in c++). by R package version 0.7.7.
Blouin, M., Parsons, M., Lacaille, V. and Lotz, S. (1996) Use of microsatellite loci to classify individuals by relatedness. Molecular Ecology, 5, 393-401.
Croissant, Y. 2011 mlogit: multinomial logit model. R package version 0.2-2.
Hardy, O.J. and Vekemans, X. (1999) Isolation by distance in a contiuous population: reconciliation between spatial autocorrelation analysis and population genetics models. Heredity, 83, 145-154.
Li, C.C., Weeks, D.E. and Chakravarti, A. (1993) Similarity of DNA fingerprints due to chance and relatedness. Human Heredity, 43, 45-52.
Li, C.C. and Horvitz, D.G. (1953) Some methods of estimating the inbreeding coefficient. American Journal of Human Genetics, 5, 107-17.
Loiselle, B.A., Sork, V.L., Nason, J. and Graham, C. (1995) Spatial genetic structure of a tropical understory shrub, Psychotria officinalis (Rubiaceae). American Journal of Botany, 82, 1420-1425.
Lynch, M. (1988) Estimation of relatedness by DNA fingerprinting. Molecular Biology and Evolution, 5(5), 584-599.
Lynch, M. and Ritland, K. (1999) Estimation of pairwise relatedness with molecular markers. Genetics, 152, 1753-1766.
Maechler, M. and many others. (2012) Utilities from Seminar fuer Statistik ETH Zurich. R package version 1.0-20.
Nei, M. (1977) F-statistics and analysis of gene diversity in subdivided populations. Annals of Human Genetics, 41, 225-233.
Nei, M. and Chesser R.K. (1983) Estimation of fixation indices and gene diversities. Annals of Human Genetics, 47, 253-259.
Oliehoek, P. A. et al. (2006) Estimating relatedness between individuals in general populations with a focus on their use in conservation programs. Genetics, 173, 483-496.
Oksanen, J. et al. (2013) vegan: Community Ecology Package. R package version 2.0-8.
Queller, D.C. and Goodnight, K.F. (1989) Estimating relatedness using genetic markers. Evolution, 43, 258-275.
Ritland, K. (1999) Estimators for pairwise relatedness and individual inbreeding coefficients. Genetics Research, 67, 175-185.
Wang, J. (2002) An estimator for pairwise relatedness using molecular markers. Genetics, 160, 1203-1215.
Weir, B.S. and Cockerham, C.C. (1984) Estimating F-Statistics for the analysis of population structure. Evolution, 38, 1358-1370.

`allele.sharing` *Calculates allele sharing rates or similarity estimators for two populations*

Description

Internal function of Demerelate to calculate different allele sharing indices or estimators for any pair of population.

Usage

```
allele.sharing(pop1, pop2, allele.column, data,
               value = NA, ref.pop)
```

Arguments

pop1	Specific dataframe of type inputformat. Population one used for calculations. Inputformat needs to be standard three digits mode for Demerelate.
pop2	Specific dataframe of type inputformat. Population two used for calculations. Inputformat needs to be standard three digits mode for Demerelate.
allele.column	Numeric value - It equals the number of the first column in the dataframe containing allele information. Order of loci in both populations needs to be exactly equal.
data	If set as TRUE allele sharing is only calculated for each pairwise comparison of individuals in the same position in pop1 with this position in pop2 - i.e. the diagonal of a distance matrix. If set as data.frame each pairwise comparison of column 1 with column 2 is calculated - i.e. the lower triangle of a distance matrix.
value	Character defining method to calculate allele sharing or similarity estimates. Can be set as "Bxy", "Sxy", "Mxy", "Li", "rxy", "lxy", "loiselle", "wang.fin", "wang", "ritland", "morans.fin" or "morans".
ref.pop	Reference population used for relatedness calculations.

Details

The function uses each population to build a dataframe with row positions for pairwise comparisons according to their position in pop1 and pop2. Similarity indices are calculated for each pair in one locus defined by the number indicated in allele.column. If allele.column is set as 1 allele information will be taken from the two columns of the first locus.

Value

Function returns the object simil.vector as vector of similarities according to value.

Author(s)

Philipp Kraemer, <philipp.kraemer@uni-oldenburg.de>

See Also

inputformat Demerelate Emp.calc Bxy Sxy Mxy
Li rxy lxy loiselle wang.fin wang ritland
morans.fin

Examples

```
## internal function not intended for direct use
```

`Demerelate` *Demerelate — Algorithms to estimate pairwise relatedness within populations based on allele sharing*

Description

Head function of Demerelate. This function should be called if any estimation of relatedness is intended. Additionally, some F-statistics can be calculated. Default parameters are set for convenient usage. Only an input dataframe containing allelic information is necessary. Geographical distances, reference populations or alterations on statistics can be set by adapting parameters.

Usage

```
Demerelate(inputdata, tab.dist = "NA", ref.pop = "NA",
           object = FALSE, value = "Mxy", Fis = FALSE,
           file.output = FALSE, p.correct = FALSE,
           iteration = 1000, pairs = 1000,
           dis.data = "relative", NA.rm = TRUE,
           genotype.ref = TRUE)
```

Arguments

`inputdata`	R object or external file to be read internally with standard Demerelate inputformat. Dataframe will be split by population information and calculations will run separately. If no reference population information is specified (`ref.pop = "NA"`) all information on loci are used as reference by omitting population information.
`tab.dist`	R object or external file to be read internally with standard Demerelate inputformat. Geographic distances can be defined and will be analysed combined with genetic data. Column three and four of standard inputformat are used for x and y coordinates.
`ref.pop`	R object or external file to be read internally with standard Demerelate inputformat. Custom reference populations will be loaded for the analysis. Population information of reference file will be omitted so that allele frequencies are calculated from the whole dataset. Optionally allele frequencies can be loaded as reference: The object should be then a list of allele frequencies. For each locus a vector with allele frequencies p and allele names as vector names needs to be combined to a list. The last list object is a vector of sample sizes for each locus.
`object`	Information whether inputdata are objects or should be read in as files.

`value`	String defining method to calculate allele sharing or similarity estimates. Can be set as "Bxy", "Sxy", "Mxy", "Li", "lxy", "rxy", "loiselle", "wang.fin", "wang", "ritland", "morans.fin" or "morans" allele.sharing.
`Fis`	*logical*. Should F_{is} values be calculated for each population?
`iteration`	Number of bootstrap iterations in F_{is} calculations.
`pairs`	Number of pairs calculated from reference populations for randomized full siblings, half siblings and non related individuals.
`file.output`	*logical*. Should a cluster dendogram, histograms and .txt files be sent as standard output in your working directory. In some cases (inflating NA values) it may be necessary that this value has to be set as FALSE due to problems in calculating clusters on pairwise NA values.
`p.correct`	*logical*. Should Yates correction from `prop.test(...)` be used in χ^2 statistics when calculating p-values on differences between empirical and randomized relatedness in populations.
`dis.data`	The kind of data to be used as distance measure. Can be "relative" - relative x and y coordinates should be given in `tab.dist` or "decimal" for geographic decimal degrees.
`NA.rm`	logical. If set as TRUE samples with NA in any position are removed from the calculation. If set as FALSE you may get an error message telling you to remove some individuals to run through the procedure. Always be aware that if your calculations are successful although you have NA values in your populations your may be biased by missing data.
`genotype.ref`	logical. If set as TRUE random non related populations are generated from genotypes of the reference population. If set as false allele frequencies are used for reference population generation. If ref.pop is given as list of allele frequencies genotype.ref = FALSE is forced.

Details

Pairwise relatedness is calculated from inputdata. Be sure to fit exactly the inputformat. Missing values are omitted when flagged as `NA`. If no additional reference populations are defined, inputdata omitting population information are used to calculate references. If no good reference populations are available you need to take care of bias in calculations. In any case you should consult for example Oliehoek et al. 2006 to get an idea of bias in relatedness calculations.
Geographic distances between individual pairs are calculated when `tab.dist = ...` . Distances calculated from x-y coordinates by simple Pythagorean mathematics can be applied to any metrical positions in sampling. Geographic coordinates from e.g. GPS need to be transformed to decimal GPS coordinates. Be sure to have positions for each individual or remove missing values from inputdata.
Each calculation will have its unique bar-code and is named with the date and population name. Calculations are performed for each population in the inputdata.

Value

Function returns files in a folder named with a bar-code and date of analysis as follows if file.output is set as TRUE:

`Empirical.relatedness.Population.txt`
: Matrix of relatedness values for each population.

`Geographic.distances.Population.txt`
: Matrix of geographic distances for each population.

`Relate.mean.Populationout.name.txt`
: Depends on selected estimators and mode of analysis. Either a summary of correlation of relatedness with geographic distance for each population or a summary of tests for relatedness within populations compared to reference populations is written to the file.

`Random.Halfsib.distances.overall.txt`
: Matrix of relatedness values calculated from randomized reference population for half siblings.

`Random.NonRelated.distances.overall.txt`
: Matrix of relatedness values calculated from randomized reference population for non related individuals.

`Random.Fullsib.distances.overall.txt`
: Matrix of relatedness values calculated from randomized reference population for full siblings.

`Cluster.Populationout.name.pdf`
: Containing an UPGMA cluster dendogram of relatedness values and a histogram of relatedness values per locus and for loci overall.

`Total-Regression.Population.pdf`
: Containing regression plot and linear fit for geographic distance and genetic relatedness.

`Summary.Populationout.name.txt`
: Summary of analysis of F statistics and allele/genotype frequencies.

Function returns via `return` following objects as one list:

`dem.results[[1]]`
: Settings of the calculation are passed to this list object.

`dem.results[[2]]`
: Mean relatedness for empirical population over all loci.

`dem.results[[3]]`
: Summarized relatedness statistics with thresholds and randomized populations from the dataset.

`dem.results[[4]]`
: Statistical analysis of the number of siblings found for each population.

`dem.results[[5]]`
: Thresholds for relatedness if "Bxy" or "Mxy" are selected as estimators

`dem.results[[6]]`
: F_{is} values and statistics for each population if `Fis==TRUE`

`dem.results[[7]]`
: Summary of linear regression of distance data are provided.

Author(s)

Philipp Kraemer, <philipp.kraemer@uni-oldenburg.de>

References

Armstrong, W. (2012) fts: R interface to tslib (a time series library in c++). by R package version 0.7.7.
Blouin, M., Parsons, M., Lacaille, V. and Lotz, S. (1996) Use of microsatellite loci to classify individuals by relatedness. Molecular Ecology, 5, 393-401.
Hardy, O.J. and Vekemans, X. (1999) Isolation by distance in a contiuous population: reconciliation between spatial autocorrelation analysis and population genetics models. Heredity, 83, 145-154.
Li, C.C., Weeks, D.E. and Chakravarti, A. (1993) Similarity of DNA fingerprints due to chance and relatedness. Human Heredity, 43, 45-52.
Li, C.C. and Horvitz, D.G. (1953) Some methods of estimating the inbreeding coefficient. American Journal of Human Genetics, 5, 107-17.
Loiselle, B.A., Sork, V.L., Nason, J. and Graham, C. (1995) Spatial genetic structure of a tropical understory shrub, Psychotria officinalis (Rubiaceae). American Journal of Botany, 82, 1420-1425.
Lynch, M. (1988) Estimation of relatedness by DNA fingerprinting. Molecular Biology and Evolution, 5(5), 584-599.
Lynch, M. and Ritland, K. (1999) Estimation of pairwise relatedness with molecular markers. Genetics, 152, 1753-1766.
Oliehoek, P. A. et al. (2006) Estimating relatedness between individuals in general populations with a focus on their use in conservation programs. Genetics, 173, 483-496.
Queller, D.C. and Goodnight, K.F. (1989) Estimating relatedness using genetic markers. Evolution, 43, 258-275.
Ritland, K. (1999) Estimators for pairwise relatedness and individual inbreeding coefficients. Genetics Research, 67, 175-185.
Wang, J. (2002) An estimator for pairwise relatedness using molecular markers. Genetics, 160, 1203-1215.

See Also

inputformat Emp.calc stat.pops F.stat

Examples

```
## Data set is used to calculate Blouins allele sharing index on
## population data. Pairs are set to 10 for convenience.
## For statistical reason for your final results you may want to
## use more pairs to model relatedness (1000 pairs are recommended).

data(demerelpop)

dem.results <- Demerelate(demerelpop[,1:6], value="Mxy",
               file.output=FALSE, object=TRUE, pairs=10)


## Demerelate can be executed with several different values
## should consult the references to decided which estimator may
## be useful in your case.
```

```
## Be careful some estimators may be biased in situations with
## no reference populations or violatin of Hardy-Weinberg
## Equilibrium.
```

`Emp.calc` *Function to calculate pairwise relatedness within populations based on allele sharing.*

Description

Allele sharing and relatedness estimators can be calculated for one population as a mean over several loci.

Usage

```
Emp.calc(tab.pop.pop, value="NA", ref.pop="NA")
```

Arguments

`tab.pop.pop`	Object with information for one population and loci. Object needs to be formatted according to inputformat.
`value`	String defining method to calculate allele sharing or similarity estimates. Can be set as "Bxy", "Sxy", "Mxy", "Li", "lxy", "rxy", "loiselle", "wang.fin", "wang", "ritland", "morans.fin" or "morans" allele.sharing.
`ref.pop`	Reference population needs to be specified for several calculations.

Details

The function calculates pairwise relatedness for all individuals in the dataframe. The output is a matrix of similarities by relatedness values in the population.

Value

`empirical.list`	Object containing a dataframe of mean individual pairwise relatedness. If value="NA" all available estimators are calculated in one call of function `Emp.calc`.

Author(s)

Philipp Kraemer, <philipp.kraemer@uni-oldenburg.de>

References

Armstrong, W. (2012) fts: R interface to tslib (a time series library in c++). by R package version 0.7.7.

Blouin, M., Parsons, M., Lacaille, V. and Lotz, S. (1996) Use of microsatellite loci to classify individuals by relatedness. Molecular Ecology, 5, 393-401.

Hardy, O.J. and Vekemans, X. (1999) Isolation by distance in a contiuous population: reconciliation between spatial autocorrelation analysis and population genetics models. Heredity, 83, 145-154.

Li, C.C., Weeks, D.E. and Chakravarti, A. (1993) Similarity of DNA fingerprints due to chance and relatedness. Human Heredity, 43, 45-52.

Li, C.C. and Horvitz, D.G. (1953) Some methods of estimating the inbreeding coefficient. American Journal of Human Genetics, 5, 107-17.

Loiselle, B.A., Sork, V.L., Nason, J. and Graham, C. (1995) Spatial genetic structure of a tropical understory shrub, Psychotria officinalis (Rubiaceae). American Journal of Botany, 82, 1420-1425.

Lynch, M. (1988) Estimation of relatedness by DNA fingerprinting. Molecular Biology and Evolution, 5(5), 584-599.

Lynch, M. and Ritland, K. (1999) Estimation of pairwise relatedness with molecular markers. Genetics, 152, 1753-1766.

Oliehoek, P. A. et al. (2006) Estimating relatedness between individuals in general populations with a focus on their use in conservation programs. Genetics, 173, 483-496.

Queller, D.C. and Goodnight, K.F. (1989) Estimating relatedness using genetic markers. Evolution, 43, 258-275.

Ritland, K. (1999) Estimators for pairwise relatedness and individual inbreeding coefficients. Genetics Research, 67, 175-185.

Wang, J. (2002) An estimator for pairwise relatedness using molecular markers. Genetics, 160, 1203-1215.

See Also

allele.sharing Bxy Sxy Mxy Li rxy lxy loiselle wang.fin wang ritland morans.fin

Examples

```
## demerelpop needs to be an object of class data.frame formatted
## according to inputformat. For each pairwise comparison in
## in the data.frame empirical estimates of relatedness are
## calculated omitting statistics.

 data(demerelpop)

 # As example Mxy is calculated for one population of demerelpop

 demerelpop.sp <- split(demerelpop,demerelpop[,2])

 empirical.result <- Emp.calc(demerelpop.sp[[1]], value="Mxy",
                              ref.pop="NA")
```

`Example.Data` *Example dataset for package Demerelate to calculate inter individual pairwise genetic relatedness.*

Description

The dataset gives randomized loci information of populations with differently related individuals. *Pop-FS-HS* consists of 10 full siblings (45 pairs) combined of 10 half siblings (45 pairs). *Pop-FS-Non* consists of 10 full sibilings (45 pairs) and 10 random individuals (45 pairs). *Pop-Non* is a population of 20 random drawn individuals (180 pairs). Datasets are based on information of 8 diploid loci with total number in alleles indicated by column names, i.e. number of alleles = 5, 10, 15, 20, 25, 30, 35, 40.
The dataset `demerelref` gives randomized loci information of a population with 1000 individuals. Dataset is based on information of 8 diploid loci with total number in alleles indicated by column names, i.e. number of alleles = 5, 10, 15, 20, 25, 30, 35, 40.
The dataset `demereldist` gives randomized relative positions for each individual in the dataset.

Usage

```
data(demerelpop)
data(demerelref)
data(demereldist)
```

Details

Standard inputformat is given for all applications in a similar way. Table is formatted as dataframe with headers set as TRUE. Headernames are not necessary but recommended.
First column contains sample IDs (mode=character), which must be unique. However, no proof-reading is implemented yet, but double named individuals may lead to errors or strange results. Column two contains population information (mode=factor). Pairwise relatedness is only calculated within populations. If you want to compare pairwise individual comparisons of several populations you need to build a new fictive population. You should do that with caution, since hidden population structures may produce strange results. Column three and four containing two alleles of a diploid marker for the sample defined in column one. Each marker should be appended in pairwise columns from then on.
Allele size should be sorted in columns with the small allele in the odd numbered column and the bigger in even numbered column for convenience. However, `Demerelate` will take care of not sorted allele sizes and handle it correctly. The same inputformat is used for reference populations.
All populations of inputdata are used as reference if no reference is defined by `reference.pop`. Additionally, the distance data are given (`tab.dist`) by the same inputformat. In column three and four you will need to define the x and y coordinate of each individual either by relative x-y coordinates or by decimal degrees of geographic coordinates.

Example input. Every dataframe in Demerelate should be organized in this way.

Sample-ID	Population	locus.1.a	locus.1.b	locus.2.a	locus.2.b	...
Ind.Norway.01	Norway	001	002	001	002	...

Ind.Norway.02	Norway	001	003	002	005	...
Ind.Norway.03	Norway	001	004	003	004	...
Ind.Norway.04	Norway	002	005	006	008	...
Ind.Germany.01	Germany	001	001	001	006	...
Ind.Germany.02	Germany	002	002	001	007	...
Ind.Germany.03	Germany	001	006	001	004	...
Ind.Germany.04	Germany	003	004	001	002	...

Author(s)

Philipp Kraemer, <philipp.kraemer@uni-oldenburg.de>

See Also

Demerelate Emp.calc Loci.test

Examples

```
# Please consult examples from Demerelate
```

`F.stat` *Head function to calculate F statistics.*

Description

Calculation of F_{is} values for a single population and several loci. Calculations are based on statistics from Weir and Cockerham 1984 and Nei 1983.

Usage

```
F.stat(tab.pop, object = TRUE, iteration = 1000,
       directory.name = "NA", out.name = "NA")
```

Arguments

`tab.pop`	A file with a dataframe containing genetic information with format inputformat.
`object`	logical - whether data are object or file.
`iteration`	Number of bootstrap replicates
`directory.name`	Directory name where files should be exported. Directory must be present when running the function. If set as "NA" results are only printed on screen.
`out.name`	Name of file for exporting summary of statistics.

Details

This function is executed by `Demerelate` when `Fis` set as TRUE. However, F_{is} values can be calculated on single populations without the need of doing a whole new relatedness analysis. This may be useful if F_{is} values were not calculated in the first place. Be sure that the path already is present prior to analysis. However, if set as NA results are printed on screen or exported to an object. Inputdata will be splitted by population names and calculations will be made for each single population. Note that mean F_{is} values calculated by Weir and Cockerham's method are either given as arithmetic mean of single loci F_{is} and as weighted average over loci, which should be unbiased to a first aproximation according to Weir and Cockerham 1984. Since there are many ways to calculate mean F_{is} and all may serve in different situations as a 'good' estimate you should spent some thoughts on this issue.

Value

Function returns a list for each population containing the following information:

`output.fis[[1]]`
: Empirical F_{is} values according to Nei 1983.

`output.fis[[2]]`
: Empirical F_{is} values according to Weir and Cockerham 1984.

`output.fis[[3]]`
: p values for significance for F_{is} values according to Nei 1983.

`output.fis[[4]]`
: p values for significance for F_{is} values according to Weir and Cockerham 1984.

Additionally, a file will be generated containing all these information if file.output is set as TRUE or alternatively when directory.name and out.name are given.

`SummaryPopulationout.name.txt`
: Combined output with different F_{is} metrices and allele\genotype frequencies

Author(s)

Philipp Kraemer, <philipp.kraemer@uni-oldenburg.de>

References

Weir, B.S. and Cockerham, C.C. (1984) Estimating F-Statistics for the analysis of population structure. Evolution, 38, 1358-1370.
Nei, M. (1977) F-statistics and analysis of gene diversity in subdivided populations. Annals of Human Genetics, 41, 225-233.
Nei, M. and Chesser R.K. (1983) Estimation of fixation indices and gene diversities. Annals of Human Genetics, 47, 253-259.

See Also

`Fis.calc`

Examples

```
  ## Fis statistics are calculated on demerelpop. Weir and Cockerham
  ## and Nei estimates are provided and either written to an object
  ## or to a file.
  ## For end results iteration of 1000 is recommended.


data(demerelpop)
fstat.results <- F.stat(demerelpop, iteration = 10,
                        directory.name = "NA",
                        out.name = "NA")
```

`Fis` *Calculates allele and genotype frequencies*

Description

Internal function to prepare allele and genotype frequencies for F statistics. Rows with `NA` values are removed from the calculation.

Usage

```
Fis(tab.pop, allele.column)
```

Arguments

`tab.pop`	Object with information for one or several populations and loci. Format needs to be in accordance with inputformat.
`allele.column`	Loci information from column `allele.column` and `allele.column`+1 are used for the calculations.

Details

Be sure to remove all missing data in your data file. Prior to calculation of F_{is} values. Every sample containing `NA` values in any column will be removed. It may be better to remove complete loci columns to get good results. In any case you should check this.

Value

`tab.freq`	Table of allele frequencies
`tab.freq.gen`	Table of genotype frequencies
`Homo`	Sum of homozygotes
`pop.size`	Population size
`Fis`	F_{is} values calculated according to Nei 1983
`Fis.weir`	F_{is} values calculated according to Weir and Cockerham 1984

Author(s)

Philipp Kraemer, <philipp.kraemer@uni-oldenburg.de>

References

Weir, B.S. and Cockerham, C.C. (1984) Estimating F-Statistics for the analysis of population structure. Evolution, 38, 1358-1370.
Nei, M. (1977) F-statistics and analysis of gene diversity in subdivided populations. Annals of Human Genetics, 41, 225-233.
Nei, M. and Chesser R.K. (1983) Estimation of fixation indices and gene diversities. Annals of Human Genetics, 47, 253-259.

See Also

`weir` `Fis.calc` `F.stat`

Examples

```
## internal function of F.stat not intended for direct use
```

`Fis.calc` *Calculation of Fis empirical and bootstrapped values*

Description

Internal function of `Demerelate` and `F.stat` to calculate F_{is} values from empirical datasets. Performing randomization statistics and preparing output.

Usage

```
Fis.calc(tab.pop, iteration, number.loci, object,
         directory.name, out.name)
```

Arguments

`tab.pop`	Data.frame following format of inputformat only column three and four are used for calculations.
`iteration`	Number of replicates for bootstrap statistics.
`number.loci`	Number of loci in `tab.pop`.
`object`	Whether `tab.pop` is an object or file.
`directory.name`	Name of the directory results should be send to.
`out.name`	Filename of the output.

Value

Function returns a list containing the following information:

`output.fis[[1]]`
: Empirical F_{is} values according to Nei 1983.

`output.fis[[2]]`
: Empirical F_{is} values according to Weir and Cockerham 1984.

`output.fis[[3]]`
: p values for significance for F_{is} values according to Nei 1983.

`output.fis[[4]]`
: p values for significance for F_{is} values according to Weir and Cockerham 1984.

Additionally, a file will be generated containing all these information if `file.output` is set as TRUE or alternatively when directory.name and out.name are given.

`SummaryPopulationout.name.txt`
: Combined output with different F_{is} metrices and allele\genotype frequencies

Author(s)

Philipp Kraemer, <philipp.kraemer@uni-oldenburg.de>

References

Weir, B.S. and Cockerham, C.C. (1984) Estimating F-Statistics for the analysis of population structure. Evolution, 38, 1358-1370.
Nei, M. (1977) F-statistics and analysis of gene diversity in subdivided populations. Annals of Human Genetics, 41, 225-233.
Nei, M. and Chesser R.K. (1983) Estimation of fixation indices and gene diversities. Annals of Human Genetics, 47, 253-259.

See Also

`weir` `F.stat` `Fis`

Examples

```
## internal function of F.stat not intended for direct usage
```

geo.dist *Calculates geographic distances.*

Description

Internal function to prepare geographic distances for the combined analysis with genetic relatedness.

Usage

```
geo.dist(pop1, pop2, onlypairs = FALSE, value)
```

Arguments

pop1	Specific type of dataframe as in inputformat. Population one used for calculations. Individuals passed to rows of resulting matrix. Inputformat should be standard with x and y coordinate mode for Demerelate.
pop2	Specific type of dataframe as in inputformat. Population two used for calculations. Individuals passed to columns of resulting matrix. Inputformat should be standard with x and y coordinate mode for Demerelate.
onlypairs	If set as TRUE geographic distances is calculated only for diagonal comparisons in matrix. If set as FALSE lower triangel is calculated.
value	String defining method to calculate geographic distances. Can be set as "relative" or "decimal".

Details

Two different methods of distance calculations are implemented in Demerelate. If using "relative", distances will be calculated from x-y coordinates using normal pythagoratic mathematics. When working with geographic positions value needs to be set to "decimal". x and y coordinates need to be given as geographic positions in decimal degrees (demereldist).

Value

matrix.share	Object containing geographic distances.

Author(s)

Philipp Kraemer, <philipp.kraemer@uni-oldenburg.de>

See Also

Demerelate allele.sharing inputformat

Examples

```
## internal function not intended for direct use
```

`glm.prep` *Preparation for multinomial logistic regression.*

Description

Function combines randomized full and half siblings with empirical values and non related random individuals to one multinomial logistic regression to calculate relatedness thresholds.

Usage

```
glm.prep(offfull.list, offhalf.list, offnon.list)
```

Arguments

`offfull.list`	Mean relatedness of populations randomized as full siblings over all loci from the reference population.
`offhalf.list`	Mean relatedness of populations randomized as half siblings over all loci from the reference population.
`offnon.list`	Mean relatedness of populations of randomized non relatives over all loci from the reference population.

Details

The function uses the package mlogit from Croissant 2011 to combine all the information from reference populations in one multilogistic regression model using `mlogit(...)`. Thresholds are calculated and used for downstream calculations.

Value

`sumlmr`	Summary of mlogit model
`half`	Thresholds for full and half siblings

Author(s)

Philipp Kraemer, <philipp.kraemer@uni-oldenburg.de>

References

Croissant, Y. 2011 mlogit: multinomial logit model R package version 0.2-2.

See Also

offspring

Examples

```
## internal function not intended for direct use
```

---

input.txt *Reads different types of tables and returns an object.*

Description

Internal function to prepare inputdata for Demerelate.

Usage

```
input.txt(tab.txt, mod)
```

Arguments

tab.txt	Can be either a filename which will be passed to an object or an object in the R workspace.
mod	Can be either "dist", "pop" or "ref.pop". Different informations are printed

Value

tab	Object containing dataframe as inputdata.

Author(s)

Philipp Kraemer, <philipp.kraemer@uni-oldenburg.de>

See Also

Demerelate inputformat

Examples

```
## internal function not intended for direct use
```

`Lin.reg.distance` *Linear regression of empirical genetic relatedness with geographic distance.*

Description

Function calculates mantel statistics and exports plots and matrices for regression of empirical genetic relatedness with geographic distance.

Usage

```
Lin.reg.distance(dist.m, emp.dist, pairs, tab.pop.pop,
                 offhalf.list, offfull.list,
                 relate.off.non.Mxy.mean, file.output,
                 directory.name, out.name, inputdata,
                 object, value, iteration)
```

Arguments

`dist.m`	Object provided by `geo.dist`.
`emp.dist`	Empirical relatedness of population.
`pairs`	Number of bootstrap replicates to calculate confidence interval for linear regression.
`tab.pop.pop`	Population information to model linear fit.
`offhalf.list`	`Object` - matrix of mean pairwise relatedness of each randomized comparison of half siblings.
`offfull.list`	`Object` - matrix of mean pairwise relatedness of each randomized comparison of full siblings.
`relate.off.non.Mxy.mean`	`Object` - matrix of mean pairwise relatedness of each randomized comparison of non relatives.
`file.output`	*logical* - Should a cluster dendogram, histograms and .txt files be sent as standard output in your working directory. In some cases (inflating NA values) it may be necessary that this value has to be set as FALSE due to problems in calculating clusters on pairwise NA values.
`directory.name`	Name of the directory results are sent to.
`out.name`	Filename of the output.
`inputdata`	R object or external file to be read internally with standard Demerelate inputformat. Dataframe will be split by population information and calculations will run separately. If no reference population information is specified (reference.pop = "NA") all information on loci are used as reference by omitting population information.
`object`	Information whether inputdata are an object or should be read in as file.

`value` String defining method to calculate allele sharing or similarity estimates. Can be set as "Bxy", "Sxy", "Mxy", "Li", "lxy", "rxy", "loiselle", "wang.fin", "wang", "ritland", "morans.fin" or "morans" allele.sharing.

`iteration` Number of bootstrap iterations in F_{is} calculations.

Details

Mantel statistics is calculated by mantel(...) from vegan package for datasets. Similarities are transformed to distances by D=1-S for each estimator. For visualization a scatterplot is exported with relatedness thresholds calculated for each dataset to get an idea of the relatedness between genetic and geographic distance in addition to mantel statistics.

Value

`Total-Regression.Population.pdf`
Containing regression plot and linear fit for geographic distance and genetic relatedness. Randomly generated populations for Full siblings, Halfsiblings and Non related individuals are used as optical reference in the plot. The mean relatedness of each population is plotted as line - blue dotted line for full siblings, red dotted line for half siblings and black dotted line for non related individuals.

`Relate.mean.Populationout.name.txt`
A summary of correlation of relatedness with geographic distance.

Summary statistics from mantel correlation of geogrphical distance and mean relatedness over loci are exported as list object.

Author(s)

Philipp Kraemer, <philipp.kraemer@uni-oldenburg.de>

References

Oksanen, J. et al. (2013) vegan: Community Ecology Package. R package version 2.0-8

See Also

`geo.dist` `allele.sharing` `Demerelate`

Examples

```
## internal function not intended for direct use
```

`Loci.test` *Analysis on differences in mean relatedness based on number of loci used for calculations*

Description

Random samples are drawn from populations specified in `tab.pop`. Pairwise relatedness is calculated by either "Bxy", "Sxy", "Mxy", "Li", "lxy", "rxy", "loiselle", "wang.fin", "wang", "ritland", "morans.fin" or "morans" allele.sharing. Pairwise relatedness is calculated for different number of loci beginning with only one up to the maximum number of loci in the dataset. Reference populations can be set for the calculations if `ref.pop="NA"`. `tab.pop` is used as an estimate of overall allele frequencies. Distance between pairwise relatedness estimates are calculated using `dist()` with euclidean calculation.
Graphically, a .pdf file is exported and the object `Random.loci.differences` contains each pairwise difference for usage in further statistics.

Usage

```
Loci.test(tab.pop, ref.pop = "NA", object = FALSE,
          value = NA, bt = 1000,
          file.output = FALSE)
```

Arguments

`tab.pop`	Specific dataframe of type inputformat. All populations which should be analyzed.
`ref.pop`	Specific dataframe of type inputformat. Population information which should be used as a reference for either threshold or allele frequency calculation.
`object`	*logical* - is tab.pop object.
`value`	String defining method to calculate allele sharing or similarity estimates. Can be set as "Bxy", "Sxy", "Mxy", "Li", "lxy", "rxy", "loiselle", "wang.fin", "wang", "ritland", "morans.fin" or "morans".
`bt`	Number of bootstrap replicates.
`file.output`	*logical* - should a .pdf file be sent as standard output in your working directory.

Value

`Loci.test.Sys.Date().pdf`
: Plot of linear regression of used loci and calculated mean estimate of relatedness from each bootstrapped replicate defined by `bt`.

`Random.loci.differences`
: List of differences in relatedness derived from bootstrap replicates. The x'th list element contains a matrix of means from replicates each over x random loci.

Author(s)

Philipp Kraemer, <philipp.kraemer@uni-oldenburg.de>

References

Armstrong, W. (2012) fts: R interface to tslib (a time series library in c++). by R package version 0.7.7.
Blouin, M., Parsons, M., Lacaille, V. and Lotz, S. (1996) Use of microsatellite loci to classify individuals by relatedness. Molecular Ecology, 5, 393-401.
Hardy, O.J. and Vekemans, X. (1999) Isolation by distance in a contiuous population: reconciliation between spatial autocorrelation analysis and population genetics models. Heredity, 83, 145-154.
Li, C.C., Weeks, D.E. and Chakravarti, A. (1993) Similarity of DNA fingerprints due to chance and relatedness. Human Heredity, 43, 45-52.
Li, C.C. and Horvitz, D.G. (1953) Some methods of estimating the inbreeding coefficient. American Journal of Human Genetics, 5, 107-17.
Loiselle, B.A., Sork, V.L., Nason, J. and Graham, C. (1995) Spatial genetic structure of a tropical understory shrub, Psychotria officinalis (Rubiaceae). American Journal of Botany, 82, 1420-1425.
Lynch, M. (1988) Estimation of relatedness by DNA fingerprinting. Molecular Biology and Evolution, 5(5), 584-599.
Lynch, M. and Ritland, K. (1999) Estimation of pairwise relatedness with molecular markers. Genetics, 152, 1753-1766.
Oliehoek, P. A. et al. (2006) Estimating relatedness between individuals in general populations with a focus on their use in conservation programs. Genetics, 173, 483-496.
Queller, D.C. and Goodnight, K.F. (1989) Estimating relatedness using genetic markers. Evolution, 43, 258-275.
Ritland, K. (1999) Estimators for pairwise relatedness and individual inbreeding coefficients. Genetics Research, 67, 175-185.
Wang, J. (2002) An estimator for pairwise relatedness using molecular markers. Genetics, 160, 1203-1215.

See Also

inputformat allele.sharing

Examples

```
## Loci.test to analyse Loci distribution in example data

data(demerelpop)
demerelpop.sp <- split(demerelpop,demerelpop[,2])

Loci.results <- Loci.test(demerelpop.sp[[1]][,1:8],
                ref.pop = "NA", object = TRUE,
                value = "rxy", bt = 10)
```

`offspring` *Mendelian offspring generator*

Description

Random generation of offspring from two parental individuals for one locus.

Usage

```
offspring(parent1, parent2, allele.column, pairs)
```

Arguments

`parent1`	data.frame of parent one.
`parent2`	data.frame of parent two.
`allele.column`	Loci information in column `allele.column` and `allele.column+1`.
`pairs`	Number of offspring generated from parent1 and parent2.

Details

Mendelian generator of random offspring.

Value

`offspr`	Dataframe of offpsring generated with length pairs.

Author(s)

Philipp Kraemer, <philipp.kraemer@uni-oldenburg.de>

Examples

```
## internal function not intended for direct usage
```

`relate.calc` *Coordinates internally reference and empirical datasets and statistics*

Description

Internal function of Demerelate to combine different populations of randomized offspring and empirical populations to calculate thresholds and statistics for later use as reference for each empirical population.

Usage

```
relate.calc(tab.pop, pairs, file.output, value,
            directory.name, pm)
```

Arguments

`tab.pop`	Dataframe following format of inputformat used as reference for randomizations.
`pairs`	Number of pairwise comparison for each randomization.
`file.output`	*logical*. Should a cluster dendogram, histograms and .txt files be sent as standard output in your working directory. In some cases (inflating NA values) it may be necessary that this value has to be set as FALSE due to problems in calculating clusters on pairwise NA values.
`value`	String defining method to calculate allele sharing or similarity estimates. Can be set as "Bxy", "Sxy", "Mxy", "Li", "lxy", "rxy", "loiselle", "wang.fin", "wang", "ritland", "morans.fin" or "morans".
`directory.name`	Name of the directory results send to.
`pm`	R object or external file to be read internally. Custom reference populations will be loaded for the analysis. The object should be a list of allele frequencies. For each locus a vector with allele frequencies p and allele names as vector names needs to be combined to a list. The last list object is a vector of sample sizes for each locus.

Details

The function internally calls any procedure of randomization and offspring generation. Finally, it coordinates the preparation of multinomial logistic regression for threshold calculation.

Value

`relate.return[[1]]`

`Object` - matrix of mean pairwise relatedness of all loci of each comparison in tab.pop.

`relate.return[[2]]`
: `Object` - matrix of mean pairwise relatedness of each randomized comparison of full siblings.

`Random.Fullsib.relatedness.overall.txt`
: `file` - matrix of mean pairwise relatedness of each randomized comparison of full siblings.

`relate.return[[3]]`
: `Object` - matrix of mean pairwise relatedness of each randomized comparison of half siblings.

`Random.Halfsib.relatedness.overall.txt`
: `file` - matrix of mean pairwise relatedness of each randomized comparison of half siblings.

`relate.return[[4]]`
: `Object` - matrix of mean pairwise relatedness of each randomized comparison of non relatives.

`Random.NonRelated.relatedness.overall.txt`
: `file` - matrix of mean pairwise relatedness of each randomized comparison of non relatives.

`relate.return[[5]]`
: `Object` - calculated thresholds for relatedness from reference populations.

Author(s)

Philipp Kraemer, <philipp.kraemer@uni-oldenburg.de>

References

Oliehoek, P. A. et al. (2006) Estimating relatedness between individuals in general populations with a focus on their use in conservation programs. Genetics, 173, 483-496.

See Also

`Emp.calc` `offspring` `glm.prep`

Examples

```
## internal function not intended for usage without Demerelate
```

`Similarity.Index` *Pairwise genetic similarity calculated based on different relatedness estimators or allele shraing indices*

Description

Pairwise estimates are based on one of several inidices:

Allele/genotype sharing indices, which count the number of shared alleles in different ways:

1. B_{xy} (number of alleles shared) as described in Li and Horvitz 1953.
2. S_{xy} (proportion of shared alleles) as described in Lynch 1988.
3. M_{xy} (genotype sharing) as described in Blouin et al. 1996.

l Relatedness estimators, which account for the similarity in allele composition of two individuals by chance (identity by state; IBS) based on reference allele frequencies:

1. The estimator r_{xy} based on Queller and Goodnight 1989 adapted to pairwise comparisons as described in Oliehoek et al. 2006.
2. l_{xy} is calculated based on Lynch and Ritland 1999.
3. The estimator $ritland$ is based on Ritland 1996.
4. The estimator $wang.fin$ is based on Wang 2002 for a finite sample.
5. The estimator $morans.fin$ is based on Hardy and Vekemans 1999 omitting correction for sample size.

Relatedness estimators, which additionally unbias for sample size effects:

1. Li is based on the equations from Li et al. 1993.
2. The estimator $loiselle$ is based on Loiselle et al. 1995.
3. The estimator $wang$ is based on Wang 2002 including bias correction for sample size.
4. The estimator $morans$ is based on Hardy and Vekemans 1999 with correction for sample size bias.

Single locus similarities are either simply averaged over loci for each pairwise comparison (B_{xy}, S_{xy}, M_{xy}), weighted for each locus before averaging (r_{xy}, l_{xy}, $ritland$, Li) or the multilocus estimate is weigthed by the average of weigths for each pairwise comparison over loci ($wang.fin$, $morans.fin$, $loiselle$, $wang$, $morans$).

Usage

```
Bxy(row, data, pop1, pop2, allele.column, ref.pop = NA)
Sxy(row, data, pop1, pop2, allele.column, ref.pop = NA)
Mxy(row, data, pop1, pop2, allele.column, ref.pop = NA)
rxy(row, data, pop1, pop2, allele.column, ref.pop = NA)
Li(row, data, pop1, pop2, allele.column, ref.pop = NA)
ritland(row, data, pop1, pop2, allele.column, ref.pop = NA)

lxy(row, data, pop1, pop2, allele.column, ref.pop = NA)
lxy.w(row, data, pop1, pop2, allele.column, ref.pop = NA)

loiselle(row, data, pop1, pop2, allele.column, ref.pop = NA)

wang(row, data, pop1, pop2, allele.column, ref.pop = NA)
wang.fin.w(allele.column, ref.pop = NA)
wang.w(allele.column, ref.pop = NA)
wang.compose(Ps, as)

morans.fin(row, data, pop1, pop2, allele.column, ref.pop = NA)
morans.w(pop1, pop2, allele.column, ref.pop = NA)
```

Arguments

row	A numeric value which sets the row of data used for calculations
data	A dataframe with all pairwise combinations of pop1 and pop2, column one set the name, column two the row number of the individual in pop1 and column three set the row number of individual two in pop2.
pop1	Specific dataframe of type inputformat. Population one used for calculations. Inputformat needs to be standard three digits mode for Demerelate.
pop2	Specific dataframe of type inputformat. Population two used for calculations. Inputformat needs to be standard three digits mode for Demerelate.
allele.column	Numeric value - It equals the number of of the first column in the dataframe containing allele information. Order of loci in both populations needs to be exactly equal.
ref.pop	Reference population used for relatedness calculations.
Ps	Vector of observed similarity classes P as a part of the twogen and fourgen similarity as described in Wang 2002
as	Vector of parameters to unbias the observed similarity class of P based on observed gene frequencies. Parameters are equvalent to those denoted as b-g and u in Wang 2002

Details

B_{xy}
The similarity is calculated based on the average number of alleles shared between two individuals. If there are at least three alleles in the locus - A, B and C, two diploid individuals may have four different states of similarity. If all alleles are the same in individuals Bxy=1, if two are the same between individuals (either AB vs AB or AA vs AB) Bxy=0.5, if only one allele is the same (AB vs AC) Bxy=0.25 and if no allele is shared Bxy=0 (Li and Horvitz 1953).

S_{xy}
The similarity is calculated based on the average number of allele positions, which share the same allele in both individuals. If there are at least three alleles in the locus - A, B and C, two diploid individuals may have four different states of similarity. If all alleles are the same in individuals - Sxy=1. If both individuals are heterozygous and both alleles are present in both individuals (AB vs AB) Sxy=1. If one individual is homozygous for a shared allele (eg. AA vs AB) Sxy=0.75. If only one allele is the same in both individuals (AB vs AC) Sxy=0.5 and if no allele is shared Sxy=0 (Lynch 1988).

M_{xy}
Sharing rate is calculated according to shared allele positions i.e. 0, 1 or 2 shared allele positions for diploids. A sharing rate of 0 is calculated if no alleles are shared, a rate of 0.5 if only one allele position is equal in individuals (AC vs AB or AA vs AB) and a rate of 1 if individuals match in both allele positions (AA vs AA or AB vs AB) (Blouin et al. 1996).

r_{xy}
The estimator r_{xy} based on Queller and Goodnight 1989 adapted to pairwise comparisons as described in Oliehoek et al. 2006 is calculated as follows:

$$r_{xy,l} = \frac{0.5(I_{ac}+I_{ad}+I_{bc}+I_{bd})-p_a-p_b}{1+I_{ab}-p_a-p_b}$$
$$I_{a-d} = \text{allele identities of individual I in locus l}$$
$$p_{a-b} = \text{frequencies of allele a or b in reference populations}$$

The equation is consulted twice for each individual pairing via a RE-RAT procedure. This means individuals are switched by calculating r_{xy} and r_{yx} and averaging for the final pairwise estimate.

l_{xy}
In Lynch and Ritland 1999 l_{xy} is referred to as r_{xy} arcoding to equation (5a). For multilocus estimates weights over loci are calculated according to equation (6a and 7a) (Lynch and Ritland 1999).

$$l_{xy,l} = \frac{p_a*(S_{bc}+S_{bd})+p_b*(S_{ac}+S_{ad})-4p_ap_b}{(1+S_{ab})*(p_a+p_b)-4p_ap_b}$$
$$w_{xy,l} = \frac{(1+S_{ab})*(p_a+p_b)-4p_ap_b}{2p_ap_b}$$
$$S_{a-d} = \text{allele identities of individual S in locus l}$$
$$p_{a-b} = \text{frequencies of allele a or b in reference populations}$$

Each pairwise estimate is weighted with $w_{xy,l}$ for each locus and calculated via RE-RAT as de-

scribed for `rxy`. The final pairwise estimate is divided by the sum of weights for loci W_{xy}, which is averaged analogous to RE-RAT procedure of the pairwise relatedness estimate.

Li

The estimator is calculated according to Li et al. 1993 equation 9 corrected for the average similarity for unrelated individuals based on reference allele frequencies as $U = 2a_2 + a_3$. The inital similarity S_{xy} is calculated based on Lynch 1988.

$$Li_{xy,l} = \frac{S-U}{1-U}$$
$S =$ similarity after Lynch 1988
$U = S(unrelated) = 2a_2 + a_3$ after Li et al. 1993

ritland

Ritlands original estimator from Ritland 1996 (equation 5) is calculated according to Lynch and Ritland 1999 by multiplying the final estimate with 2. As the sum of all p is equal to 1 the equation is simplyfied to:

$$ritland_{xy,l} = \frac{2}{n-1}[(\sum \frac{S_i}{p_i}) - 1]$$
$S_i =$ allele identity in both individuals
$p_i =$ reference frequency of allele i
$n =$ number of different alleles in locus l

The single locus estimate is averaged over loci. The basic similarity is calculated for allele i as 0.25 if i is present in both individuals, 0.5 if i is present in both and one individual is homozogious for i and 1 if both individuals are homozygous for i. Note that this similarity measure is equivalent to Blouin's approach M_{xy} (Blouin et al. 1996) if summed over all alleles for each pairwise comparison.

loiselle

The estimator first described in Loiselle et al. 1995 is implemented as described by Hardy and Vekemans 2015. The frequency of each allele in individuals (i.e. 0.5 or 1 for diploids) in a pairwise comparison are combined and corrected for the allele frequency in the reference population. The product of corrected shared allele frequencies is additionally corrected for sample size bias and combined over loci via weighting with the polymorphic index ($\sum p_i * (1 - p_i)$).

$$loiselle_{xy,l} = \sum \frac{\sum (p_{ila} - pla) * (p_{jla} - p_{la}) + \sum (p_{la} * (1 - p_{la}))}{n_l - 1} / \sum \sum p_{la} * (1 - p_{la})$$
$p_{ila} =$ frequency of allele a in individual i of locus l
$n =$ number of different alleles in locus l

wang

Equations described by Wang 2002 are implemented as follows:

A binary vector P of classes of observed similarities is calculated for each pairwise comparison.

One of the 4 categories is set as 1, all remaining P are set as 0. P_1 is 1 if both animals are homozygous for the same allele or both are heterozygous with the same allele combination. P_2 is 1 if one individual is homozygous and the other is heterozygous sharing only 1 allele. P_3 is 1 if only 1 allele is shared between individuals with 1 copy per individual. All other comibnations fall into the category four P_4. Note that the categories follow the general classification as described for S_{xy} described by Lynch 1988.

The estimator descirbed by Wang 2002 is relatively complex and can be discussed only superficial here. The probability of each category i.e. the joint probability of genotypes can be estimated using the two-gene (Θ) and four-gene (Δ) coefficient based on the sum of powers of allele frequencies ($a_m = \sum p_i^m$) summing up each probability of each category of P. Formulas used for calculations can be found in Wang 2002 Equation (9), (10) and (10). Combining these estimates into r=$\frac{\Theta}{2} + \Delta$ yields the estimate by Wang (2002). By using the estimator wang each expected sum of powers of allele frequency is corrected for sample size N ($\bar{a}_2$, $\bar{a}_3$, $\bar{a}_4$) compare eqation (12), (13) and (14) in Wang 2002.

Finally, to get an multilocus estimate each single locus estimate is corrected for the average similarity value for unrelated individuals $u = 2a_2 + a_3$ as described in Li et al. 1993. Each single term of the estimate Θ and Δ, namely $b - g$ and each P are unbiased by u and the combined estimate is weigthed with the sum of u over loci.

$wang.fin$
This estimate is based on the same equations as wang. However, for the use of finite samples the bias correction of sample size (N) is omitted. Instead of $\bar{a}_2$, $\bar{a}_3$, $\bar{a}_4$ ((12), (13), and (14) in Wang 2002) the pure sum of powers of allele frequencies ($a_m = \sum p_i^m$) are used.

$morans$
The estimator morans refers to morans I, which is widely used as estimator for spatial autocorrelation. It is described in Hardy and Vekemans 1999 as estimator for genetic relatedness. The approach is similar to loiselle (Loiselle et al. 1995) or r_{xy} (Queller and Goodnight 1989) by correcting each individual allele frequency by the allele frequency of the reference sample for each shared allele of each pairwise comparison. Each shared allele is then weighted by the variance of individual allele frequencies and a term for sample size bias. In order to calculate the estimator over loci the sum of pairwise estimates over loci is weighted by the sum of weights over loci for each comparison.

$$morans_{xy,l} = \frac{\sum (p_{ila} - p_{la}) * (p_{jla} - p_{la})}{\sum Var(p_{ila} + 1/(n_l - 1))}$$

p_{ila} = frequency of allele a in individual i of locus l
n = number of different alleles in locus l

$morans.fin$
The estimator morans.fin refers to the same calculations as morans but omitting weights for sample size bias. Thus it should only be applied for finite samples (Hardy and Vekemans 1999).

Value

The value of similarity is returned according to the arguments defining the individuals compared.

Author(s)

Philipp Kraemer, <philipp.kraemer@uni-oldenburg.de>

References

Armstrong, W. (2012) fts: R interface to tslib (a time series library in c++). by R package version 0.7.7.
Blouin, M., Parsons, M., Lacaille, V. and Lotz, S. (1996) Use of microsatellite loci to classify individuals by relatedness. Molecular Ecology, 5, 393-401.
oissant, Y. 2011 mlogit: multinomial logit model. R package version 0.2-2.
Hardy, O.J. and Vekemans, X. (1999) Isolation by distance in a contiuous population: reconciliation between spatial autocorrelation analysis and population genetics models. Heredity, 83, 145-154.
Li, C.C., Weeks, D.E. and Chakravarti, A. (1993) Similarity of DNA fingerprints due to chance and relatedness. Human Heredity, 43, 45-52.
Li, C.C. and Horvitz, D.G. (1953) Some methods of estimating the inbreeding coefficient. American Journal of Human Genetics, 5, 107-17.
Loiselle, B.A., Sork, V.L., Nason, J. and Graham, C. (1995) Spatial genetic structure of a tropical understory shrub, Psychotria officinalis (Rubiaceae). American Journal of Botany, 82, 1420-1425.
Lynch, M. (1988) Estimation of relatedness by DNA fingerprinting. Molecular Biology and Evolution, 5(5), 584-599.
Lynch, M. and Ritland, K. (1999) Estimation of pairwise relatedness with molecular markers. Genetics, 152, 1753-1766.
Maechler, M. and many others. (2012) Utilities from Seminar fuer Statistik ETH Zurich. R package version 1.0-20.
Nei, M. (1977) F-statistics and analysis of gene diversity in subdivided populations. Annals of Human Genetics, 41, 225-233.
Nei, M. and Chesser R.K. (1983) Estimation of fixation indices and gene diversities. Annals of Human Genetics, 47, 253-259.
Oliehoek, P. A. et al. (2006) Estimating relatedness between individuals in general populations with a focus on their use in conservation programs. Genetics, 173, 483-496.
Oksanen, J. et al. (2013) vegan: Community Ecology Package. R package version 2.0-8.
Queller, D.C. and Goodnight, K.F. (1989) Estimating relatedness using genetic markers. Evolution, 43, 258-275.
Ritland, K. (1999) Estimators for pairwise relatedness and individual inbreeding coefficients. Genetics Research, 67, 175-185.
Wang, J. (2002) An estimator for pairwise relatedness using molecular markers. Genetics, 160, 1203-1215.
Weir, B.S. and Cockerham, C.C. (1984) Estimating F-Statistics for the analysis of population structure. Evolution, 38, 1358-1370.

See Also

allele.sharing Demerelate Emp.calc

Examples

```
## internal function not intended for direct use
```

stat.pops *Calculation of Fis empirical and bootstrapped values*

Description

Internal function of Demerelate to use prepared thresholds of relatedness to calculate state of relatedness for empirical data. Additionally, it combines information in different plots and prepares each single output.

Usage

```
stat.pops(Thresholds, tab.pop.pop, pairs, p.correct,
          directory.name, out.name, file.output,
          inputdata, object, value, iteration,
          pm, genotype.ref)
```

Arguments

Thresholds	Thresholds of relatedness to be used for statistics on empirical data.
tab.pop.pop	Dataframe following format of inputformat only column three and four are used for calculations.
pairs	Number of randomized pairings used for reference populations.
p.correct	*logical* - should Yates correcture be used for χ^2 statistics.
directory.name	Name of the directory results send to.
out.name	Filename of the output.
file.output	*logical*. Should a cluster dendogram, histograms and .txt files be sent as standard output in your working directory. In some cases (inflating NA values) it may be necessary that this value has to be set as FALSE due to problems in calculating clusters on pairwise NA values.
inputdata	R object or external file to be read internally with standard Demerelate inputformat. Dataframe will be split by population information and calculations will run separately. If no reference population information is specified (ref.pop = "NA") all information on loci are used as reference by omitting population information.
object	Information whether inputdata are an object or should be read in as file.
value	String defining method to calculate allele sharing or similarity estimates.
iteration	Number of bootstrap iterations in F_{is} calculations.
pm	R object or external file to be read internally. Custom reference populations will be loaded for the analysis. The object should be a list of allele frequencies. For each locus a vector with allele frequencies p and allele names as vector names needs to be combined to a list. The last list object is a vector of sample sizes for each locus.

`genotype.ref` logical. If set as TRUE random non related populations are generated from genotypes of the reference population. If set as false allele frequencies are used for reference population generation.

Details

Values for logical operators and data are given by Demerelate.

Value

`ClusterPopulationNameOutName.pdf`
Combined information on dataset by different types of plots.

`Relate.meanPopulationNameOutName.txt`
Combined information on dataset regarding calculated thresholds and number of different types of relatives in population. Summary of χ^2 statistics.

`out.stat[[1]]` Empirical relatedness in populations.

`out.stat[[2]]` Summary statistics for χ^2 statistics.

`error.individuals.txt`
If too many NAs are found in the data the program may stop here and tells you what individuals caused the troubles in calculations.

Author(s)

Philipp Kraemer, <philipp.kraemer@uni-oldenburg.de>

References

Oliehoek, P. A. et al. (2006) Estimating relatedness between individuals in general populations with a focus on their use in conservation programs. Genetics, 173, 483-496.

See Also

`weir` `F.stat`

Examples

```
## internal function not intended for usage without Demerelate
```

weir *Calculation of F_is based on Weir and Cockerham 1984*

Description

Function used internally by `Demerelate` and `F.stat` to calculate F_{is} values based on Weir and Cockerham (1984) statistics.

Usage

```
weir(tab.pop, tab.freq, popsize)
```

Arguments

`tab.pop`	Dataframe following format of inputformat only column three and four are used for calculations.
`tab.freq`	Vector of allele frequencies with name of each allele in column name.
`popsize`	Population size.

Details

The function provides the alternative methods to Nei's F_{is} calculation. F_{is} values are calculated according to Weir and Cockerham 1984. Consult references for details

Value

`fis.weir`	F_{is} value calculated by method of Weir and Cockerham 1984

Author(s)

Philipp Kraemer, <philipp.kraemer@uni-oldenburg.de>

References

Weir, B.S. and Cockerham, C.C. (1984) Estimating F-Statistics for the analysis of population structure. Evolution, 38, 1358-1370.

See Also

`Fis.calc` `F.stat` `Fis`

Examples

```
## internal function of F.stat not intended for direct usage
```

Index