

Package ‘PopGenKit’

February 19, 2015

Type Package

Title Useful functions for (batch) file conversion and data resampling in microsatellite datasets

Version 1.0

Date 2011-07-21

Author Sebastien Rioux Paquette

Maintainer Sebastien Rioux Paquette

<sebastien.riouxpaquette@gmail.com>

Description There are two main purposes to this package. The first is to allow batch conversion of Genepop (Rousset 2008) input files for use with Arlequin (Excoffier and Lischer 2010), which has a simple GUI to analyze batch files. Two commonly used simulation software, BottleSim (Kuo & Janzen 2003) and Easypop (Balloux 2001) produce Genepop output files that can be analyzed this way. There are also functions to convert to and from BottleSim format, to quickly produce allele frequency tables or to convert a file directly for use in ordination analyses (e.g. principal component analysis). This package also includes functions to calculate allele rarefaction curves, confidence intervals on heterozygosity and allelic richness with resampling strategies (bootstrap and jackknife).

License GPL-2

LazyLoad yes

Repository CRAN

Date/Publication 2012-10-29 08:57:27

NeedsCompilation no

R topics documented:

PopGenKit-package	2
batchconvert	3
batchconvertjack	4

batchjacklociArp	5
batchjacklociArpRANDOM	6
batchrecodeBi	7
bootstrapHet	8
bootstrapHet_byloc	10
combineBsim	11
convert	12
convertjack	13
freqPCA	14
inputBottleSim	15
jacklociArp	16
jacklociArpRANDOM	17
jackmsat	18
jackmsatpop	19
outputBottleSim	20
PICalc	21
popgen	22
recodeBi	23

Index	25
--------------	-----------

PopGenKit-package	<i>Useful functions for (batch) file conversion and data resampling in microsatellite datasets</i>
-------------------	--

Description

There are two main purposes to this package. The first is to allow batch conversion of Genepop (Rousset 2008) input files for use with Arlequin (Excoffier and Lischer 2010), which has a simple GUI to analyze batch files. Two commonly used simulation software, BottleSim (Kuo & Janzen 2003) and EasyPop (Balloux 2001) produce Genepop output files that can be analyzed this way. There are also functions to convert to and from BottleSim format, to quickly produce allele frequency tables or to convert a file directly for use in ordination analyses (e.g. principal component analysis). This package also includes functions to calculate allele rarefaction curves, confidence intervals on heterozygosity and allelic richness with resampling strategies (bootstrap and jackknife).

Details

Package:	PopGenKit
Type:	Package
Version:	1.0
Date:	2011-07-21
License:	What license is it under?
LazyLoad:	yes

List of functions: [convert](#), [batchconvert](#) [convertjack](#), [batchconvertjack](#) [jacklociArp](#), [batchjacklociArp](#) [jacklociArpRANDOM](#), [batchjacklociArpRANDOM](#) [recodeBi](#), [batchrecodeBi](#) [bootstrapHet](#), [bootstrapHet_byloc](#) [freqPCA](#) [popgen](#) [jackmsatpop](#) [PICalc](#) [inputBottleSim](#), [outputBottleSim](#), [combineBsim](#)

Author(s)

Maintainer: Sebastien Rioux Paquette <sebastien.riouxpaquette@gmail.com>

References

Balloux, F. (2001). EASYPOP (version 1.7): A computer program for population genetics simulations. *J. Hered.* 92: 301–302.

Excoffier, L. and H.E.L. Lischer (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Res.* 10: 564-567.

Kuo C.H., and F.J.Janzen (2003). BottleSim: a bottleneck simulation program for long-lived species with overlapping generations. *Mol. Ecol. Notes*, 3, 669-673.

Rousset, F. (2008). Genepop'007: a complete reimplementaion of the Genepop software for Windows and Linux. *Mol. Ecol. Res.* 8: 103-106.

batchconvert

Convert all Genepop files in dir to Arlequin format

Description

This function converts all Genepop files (extension .gen) in the working directory to Arlequin files (extension .arp). It also creates a batch Arlequin file (extension .arb) to allow batch analysis in Arlequin.

Usage

```
batchconvert(ndigit = 3)
```

Arguments

`ndigit` The number of digits per allele in the input file. Can be 2 or 3 (default 3).

Details

Relies on the function [convert](#) to perform the batch conversion.

Input files must end with the extension '.gen'. Make sure to follow standard Genepop format for input files. The IDs of all individuals should be immediately followed by a comma. Population labels should be Pop, POP, or pop. There should not be invisible characters (tabs or spaces) at the end of lines, especially lines on which population labels are found. These lines should have only three characters (Pop). See example file 'glyptemys3.gen' if needed.

Value

All arp files and the single arb file are saved in the working directory

Author(s)

Sebastien Rioux Paquette

References

Excoffier, L. and H.E.L. Lischer (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Res.* 10: 564-567.

Rousset, F. (2008). Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Mol. Ecol. Res.* 8: 103-106.

See Also

[convert](#)

batchconvertjack	<i>Convert all Genepop files in dir to Arlequin format while removing any subset of loci</i>
------------------	--

Description

This function converts all Genepop files (extension .gen) in the working directory to Arlequin files (extension .arp). It also creates a batch Arlequin file (extension .arb) to allow batch analysis in Arlequin. The user specifies a subset of loci to be removed.

Usage

```
batchconvertjack(ndigit = 3, rm.loci)
```

Arguments

ndigit	The number of digits per allele in the input file. Can be 2 or 3 (default is 3).
rm.loci	A vector indicating which locus or loci to remove from the final output file. Loci should be indicated as numbers according to the order in which they are listed in the input file (e.g. c(1,3) would remove the first and third loci).

Details

Relies on the function [convertjack](#) to do file conversion.

Input file must end with the extension '.gen'. Make sure to follow standard Genepop format for input file. The IDs of all individuals should be immediately followed by a comma. Population labels should be Pop, POP, or pop. There should not be invisible characters (tabs or spaces) at the end of lines, especially lines on which population labels are found. These lines should have only three characters (Pop). See example file 'glyptemys3.gen' if needed.

The option to remove loci from the output files can be extremely useful in simulation studies, in order to assess the influence of certain markers or estimate confidence intervals with jackknifing procedures. Among others, EasyPop generates Genepop files that could be processed with this function.

Value

All arp files and the single arb file are saved in the working directory

Author(s)

Sebastien Rioux Paquette

References

- Balloux, F. (2001). EASYPOP (version 1.7): A computer program for population genetics simulations. *J. Hered.* 92: 301–302.
- Excoffier, L. and H.E.L. Lischer (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Res.* 10: 564-567.
- Rousset, F. (2008). Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Mol. Ecol. Res.* 8: 103-106.

batchjacklociArp	<i>Remove a subset of loci from all Arlequin project files (extension .arp) in dir</i>
------------------	--

Description

This function removes a specified subset of loci from all Arlequin (extension .arp) files in the working directory and creates new input .arp files. It also creates a Arlequin batch file (.arb) to allow batch analysis.

Usage

```
batchjacklociArp(rm.loci)
```

Arguments

rm.loci	A vector indicating which locus or loci to remove from the final output file. Loci should be indicated as numbers according to the order in which they are listed in the input file (e.g. c(1,3) would remove the first and third loci).
---------	--

Details

Relies on the function [jacklociArp](#), .

Some .arp files might not be imported correctly depending on which software was used to generate them. If this is the case, start from the Genepop files and convert them first with [batchconvert](#) or use [batchconvertjack](#).

Value

New .arp files (with the suffix 'jackloci' in the filename) and the single batch file (.arb) will be saved in the working directory.

Author(s)

Sebastien Rioux Paquette

References

Excoffier, L. and H.E.L. Lischer (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Res.* 10: 564-567.

Rousset, F. (2008). Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Mol. Ecol. Res.* 8: 103-106.

See Also

[jacklociArp](#), [batchconvert](#), [batchconvertjack](#)

batchjacklociArpRANDOM

Remove a random subset of loci from all Arlequin project files (extension .arp) in dir

Description

This function retains a user-specified number of loci from all Arlequin (extension .arp) files in the working directory and creates new input .arp files. The identity of retained markers is chosen randomly for each file. The function also creates a Arlequin batch file (.arb) to allow batch analysis.

Usage

```
batchjacklociArpRANDOM(n.retain)
```

Arguments

n.retain	Number of loci to keep in final file (must be between 1 and the number of loci in input file)
----------	---

Details

This function relies on [jacklociArpRANDOM](#) to jackknife markers in each file.

This function is useful to investigate the effect of the number of sampled loci on population genetics parameter estimates, especially in simulations like those that can be performed with the software EasyPop.

Some .arp files might not be imported correctly depending on which software was used to generate them. If this is the case, start from the Genepop files and convert them first with [batchconvert](#).

Value

New .arp files (with a suffix indicating the number of markers in the filename) and the single batch file (.arb) will be saved in the working directory.

Author(s)

Sebastien Rioux Paquette

References

Balloux, F. (2001). EASYPOP (version 1.7): A computer program for population genetics simulations. *J. Hered.* 92: 301–302.

Excoffier, L. and H.E.L. Lischer (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Res.* 10: 564-567.

Rousset, F. (2008). Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Mol. Ecol. Res.* 8: 103-106.

See Also

[jacklociArpRANDOM](#), [batchconvert](#)

batchrcodeBi	<i>Transform all multi-allelic Arlequin datafiles in dir to bi-allelic datafiles</i>
--------------	--

Description

This function transforms datasets with multi-allelic markers (e.g. microsatellites) to bi-allelic datasets, where all alleles but one (usually the most frequent) are pooled together to form a second allelic class. All files should comprise the same markers because the same alleles (e.g. simulation replicates from the same data) will be retained for all files.

Usage

```
batchrcodeBi(major.alleles)
```

Arguments

`major.alleles` A vector listing which allele to retain for each locus. Length of this vector should be equal to the number of markers in datafile.

Details

Because several population genetics parameter estimates are affected by the level of polymorphism of loci (e.g. F_{st}), it may be useful to recode multi-allelic markers to a bi-allelic state. For instance, this could allow the comparison of F_{st} values for microsatellites and single nucleotide polymorphisms (SNPs). McDonald (1994) introduced this as F_{2st} , and it has been used elsewhere (e.g. Allendorf and Seeb 2000).

The only .arp files in the working directory should be those to recode.

The function `popgen` provides a quick way to obtain allele frequencies in order to determine which alleles are the most frequent, within populations or overall.

Some .arp files might not be imported correctly depending on which software was used to generate them. If this is the case, start from the Genepop files and convert them first with `batchconvert`.

Value

All Arlequin files coded in bi-allelic format and the Arlequin batch file are saved in the working directory.

Author(s)

Sebastien Rioux Paquette

References

Allendorf, F.W. and L.W. Seeb (2000) Concordance of genetic divergence among sockeye salmon populations at allozyme, nuclear DNA, and mitochondrial DNA markers. *Evolution* 54: 640-651.

Excoffier, L. and H.E.L. Lischer (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Res.* 10: 564-567.

McDonald, J.H. (1994) Detecting natural selection by comparing geographic variation in protein and DNA polymorphisms. Pages 88-100 in Golding B, ed. *Non-neutral evolution: theories and molecular data*. New York: Chapman and Hall.

Rousset, F. (2008). Genepop'007: a complete reimplementations of the Genepop software for Windows and Linux. *Mol. Ecol. Res.* 8: 103-106.

See Also

`recodeBi`

bootstrapHet

Calculate confidence intervals on observed and expected heterozygosity by bootstrapping individuals in each population

Description

This function uses a Arlequin project file (.arp) as input. For each population, it resamples individuals with replacement (i.e. bootstrap resampling) and calculates the corresponding confidence intervals on H_o (observed heterozygosity) and H_e (expected heterozygosity).

Usage

```
bootstrapHet(datafile, ndigit = 3, nrepet = 1000)
```

Arguments

datafile	Name of the Arlequin file to use as input, with the extension (e.g. 'glyptemys3.arp')
ndigit	Number of digits per allele in input file. Can be 2 or 3 (default is 3)
nrepet	Number of bootstrap replicates to perform. Default is 1000. Can be slow for large datafiles.

Details

Some .arp files might not be imported correctly depending on which software was used to generate them. If this is the case, start from the Genepop file and convert it first with [convert](#).

It is possible to perform bootstrapping on markers instead of individuals. To do so, use the function [bootstrapHet_byloc](#) instead.

Value

Results are printed to a file saved in the working directory, with the suffix 'bootHet' in the filename. First part of the table reports real H_o and H_e values for each locus in each population. Bottom part reports bootstrap results for each population: mean calculated H_o and H_e as well as the lower and upper bounds of their 95 % confidence intervals.

Author(s)

Sebastien Rioux Paquette

References

Excoffier, L. and H.E.L. Lischer (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Res.* 10: 564-567.

Rousset, F. (2008). Genepop'007: a complete reimplementations of the Genepop software for Windows and Linux. *Mol. Ecol. Res.* 8: 103-106.

See Also

[bootstrapHet_byloc](#), [convert](#)

bootstrapHet_byloc	<i>Calculate confidence intervals on observed and expected heterozygosity in each population by bootstrapping markers</i>
--------------------	---

Description

This function uses a Arlequin project file (.arp) as input. It resamples loci with replacement (i.e. bootstrap resampling) and calculates the corresponding confidence intervals on H_o (observed heterozygosity) and H_e (expected heterozygosity) in each population.

Usage

```
bootstrapHet_byloc(datafile, ndigit = 3, nrepet = 1000)
```

Arguments

datafile	Name of the Arlequin file to use as input, with the extension (e.g. 'glyptemys3.arp')
ndigit	Number of digits per allele in input file. Can be 2 or 3 (default is 3)
nrepet	Number of bootstrap replicates to perform. Default is 1000. Can be slow for large datafiles.

Details

Some .arp files might not be imported correctly depending on which software was used to generate them. If this is the case, start from the Genepop file and convert it first with [convert](#).

It is possible to perform bootstrapping on individuals instead of loci. To do so, use the function [bootstrapHet](#) instead.

Value

Results are printed to a file saved in the working directory, with the suffix 'bootHet_byloc' in the filename. First part of the table reports real H_o and H_e values for each locus in each population. Bottom part reports bootstrap results for each population: mean calculated H_o and H_e as well as the lower and upper bounds of their 95 % confidence intervals.

Author(s)

Sebastien Rioux Paquette

References

Excoffier, L. and H.E.L. Lischer (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Res.* 10: 564-567.

Rousset, F. (2008). Genepop'007: a complete reimplement of the Genepop software for Windows and Linux. *Mol. Ecol. Res.* 8: 103-106.

See Also

[bootstrapHet](#),

combineBsim	<i>Combines real and simulated multilocus data (from the software BottleSim) in new Genepop files (one file per simulation replicate).</i>
-------------	--

Description

One might be interested in comparing data from simulated bottlenecked populations performed in the software BottleSim (Kuo and Janzen 2003) with the original data that was used as a starting point for these simulations (e.g. calculate post-bottleneck F_{st} values). This function will take a genotype output file from BottleSim that has previously been recoded according to real allele sizes (using [outputBottleSim](#)) and append each of the replicates in that file to the original Genepop file. Each new file created comprises a single simulation replicate.

Usage

```
combineBsim(introduced, source, ndigit = 3, subsample = 0)
```

Arguments

introduced	Name of the simulated genotypes file, generated with the function outputBottleSim .
source	Name of the original data that was used to generate the BottleSim input file with the function inputBottleSim .
ndigit	Number of digits per allele. Can be 2 or 3 (default is 3).
subsample	Number of individuals to retain in each replicate. Individuals are chosen randomly in each replicate. Zero (default value) means all individuals will be kept.

Details

The names 'introduced' and 'source' come from the study for which this function was created. In that study, the authors were interested in evaluating the extent of differentiation (F_{st}) that would be expected between an introduced population and its source(s), under different scenarios (Heeg et al.)

Once all combined Genepop files have been created (as many as there are BottleSim replicates in the simulated genotype file), it is possible to convert them all to Arlequin format for fast, batch analysis of genetic parameters, using [batchconvert](#).

Value

A Genepop file will be created for each replicate found in 'introduced'. Each of these files will comprise all data found in 'source' and a single replicate from 'introduced'.

Author(s)

Sebastien Rioux Paquette

References

- Excoffier, L. and H.E.L. Lischer (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Res.* 10: 564-567.
- Kuo C.H., and F.J.Janzen (2003). BottleSim: a bottleneck simulation program for long-lived species with overlapping generations. *Mol. Ecol. Notes*, 3, 669-673.
- Rousset, F. (2008). Genepop'007: a complete reimplementaion of the Genepop software for Windows and Linux. *Mol. Ecol. Res.* 8: 103-106.

See Also

[inputBottleSim](#), [outputBottleSim](#), [batchconvert](#)

convert

Convert a Genepop file to Arlequin format

Description

This function converts a datafile from Genepop format (extension .gen) to Arlequin project file format (extension .arp).

Usage

```
convert(datafile, ndigit = 3)
```

Arguments

`datafile` Name of the file to convert, with the extension (e.g. 'glyptemys3.gen').

`ndigit` The number of digits per allele in the input file. Can be 2 or 3 (default is 3).

Details

Input file must end with the extension '.gen'. Make sure to follow standard Genepop format for input file. The IDs of all individuals should be immediately followed by a comma. Population labels should be Pop, POP, or pop. There should not be invisible characters (tabs or spaces) at the end of lines, especially lines on which population labels are found. These lines should have only three characters (Pop). See example file 'glyptemys3.gen' if needed.

For converting multiple files, use the function [batchconvert](#) instead.

Value

A file with the same file name as the input file, but with the .arp extension (Arlequin), will be saved in the working directory.

Author(s)

Sebastien Rioux Paquette

References

Excoffier, L. and H.E.L. Lischer (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Res.* 10: 564-567.

Rousset, F. (2008). Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Mol. Ecol. Res.* 8: 103-106.

See Also

[batchconvert](#)

convertjack	<i>Convert a Genepop file to Arlequin format while allowing the removal of any subset of loci from the output</i>
-------------	---

Description

This function converts a datafile from Genepop format (extension .gen) to Arlequin project file format (extension .arp). The user can specify which loci to remove so that only the others will be included in the output file.

Usage

```
convertjack(datafile, ndigit = 3, rm.loci)
```

Arguments

datafile	Name of the file to convert, with the extension (e.g. 'glyptemys3.gen').
ndigit	The number of digits per allele in the input file. Can be 2 or 3 (default is 3).
rm.loci	A vector indicating which locus or loci to remove from the final output file. Loci should be indicated as numbers according to the order in which they are listed in the input file (e.g. c(1,3) would remove the first and third loci).

Details

Input file must end with the extension '.gen'. Make sure to follow standard Genepop format for input file. The IDs of all individuals should be immediately followed by a comma. Population labels should be Pop, POP, or pop. There should not be invisible characters (tabs or spaces) at the end of lines, especially lines on which population labels are found. These lines should have only three characters (Pop). See example file 'glyptemys3.gen' if needed.

For converting multiple files, use the function [batchconvertjack](#) instead.

Value

A file with the same file name as the input file, but with the .arp extension (Arlequin) will be saved in the working directory.

Author(s)

Sebastien Rioux Paquette

References

Excoffier, L. and H.E.L. Lischer (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Res.* 10: 564-567.

Rousset, F. (2008). Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Mol. Ecol. Res.* 8: 103-106.

See Also

[batchconvertjack](#)

freqPCA

Create a table of allele frequencies from Arlequin file (.arp) that can be used directly for ordination analyses (e.g. Principal Component Analysis)

Description

This function will take a Arlequin input file (.arp) of multilocus genotypes and create a table of allele frequencies for multivariate analyses. For each locus, one allele is omitted, as it would be a redundant character for PCA analyses.

Usage

```
freqPCA(datafile, ndigit = 3, default.pop = T, vecpops)
```

Arguments

datafile	Name of the input Arlequin file, with extension (e.g. 'glyptemys3.arp').
ndigit	Number of digits per allele in input file. Can be 2 or 3. (Default is 3).
default.pop	If TRUE, population names will be taken from input file. If FALSE, they will be taken from an additional vector of population names. (Default is TRUE).
vecpops	Optional vector of population names, needed if default.pop is set to FALSE.

Details

Some .arp files might not be imported correctly depending on which software was used to generate them. If this is the case, start from the Genepop file and convert it first with [convert](#).

Value

Returns a table of allele frequencies formatted for PCA analysis. Also saves the table in a file placed in the working directory.

Author(s)

Sebastien Rioux Paquette

References

Excoffier, L. and H.E.L. Lischer (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Res.* 10: 564-567.

Rousset, F. (2008). Genepop'007: a complete reimplement of the Genepop software for Windows and Linux. *Mol. Ecol. Res.* 8: 103-106.

inputBottleSim	<i>Create a BottleSim input file from a Genepop (.gen) file.</i>
----------------	--

Description

This function creates a genotype input file for use with the population genetics simulation software BottleSim (Kuo and Janzen 2003). It also saves all required allele information to allow retro-conversion of BottleSim simulation results back to original alleles with the function [outputBottleSim](#).

Usage

```
inputBottleSim(datafile, ndigit = 3)
```

Arguments

datafile	Name of the input Genepop file with the extension (e.g. 'glyptemys3.gen')
ndigit	Number of digits per allele in input file. Can be 2 or 3 (default 3).

Details

Two files will be created after using this function; one is the BottleSim input file, the other is called 'alleles'. This file is required to retro-convert BottleSim results back to the original format with [outputBottleSim](#). Do not modify or erase this file.

Make sure to follow standard Genepop format for input file. The IDs of all individuals should be immediately followed by a comma. Population labels should be Pop, POP, or pop. There should not be invisible characters (tabs or spaces) at the end of lines, especially lines on which population labels are found. These lines should have only three characters (Pop). See example file 'glyptemys3.gen' if needed.

Value

The BottleSim file and the 'alleles' file will be saved in the working directory

Author(s)

Sebastien Rioux Paquette

References

Kuo C.H., and F.J.Janzen (2003). BottleSim: a bottleneck simulation program for long-lived species with overlapping generations. *Mol. Ecol. Notes*, 3, 669-673.

Rousset, F. (2008). Genepop'007: a complete reimplementaion of the Genepop software for Windows and Linux. *Mol. Ecol. Res.* 8: 103-106.

See Also

[outputBottleSim](#), [combineBsim](#)

jacklociArp

Remove a subset of loci from a Arlequin project file (extension .arp)

Description

This function will create a new .arp file without the loci specified by the user.

Usage

```
jacklociArp(datafile, rm.loci)
```

Arguments

datafile	Name of the file to jackknife, with the extension (e.g. 'glyptemys3.arp').
rm.loci	A vector indicating which locus or loci to remove from the final output file. Loci should be indicated as numbers according to the order in which they are listed in the input file (e.g. c(1,3) would remove the first and third loci).

Details

Some .arp files might not be imported correctly depending on which software was used to generate them. If this is the case, start from the Genepop file and convert it first with [convert](#) or use [convertjack](#).

For converting multiple files, use the function [batchjacklociArp](#) instead.

Value

A new .arp file will be saved in the working directory, with the suffix '_jackloci' before the extension in the filename.

Author(s)

Sebastien Rioux Paquette

References

Excoffier, L. and H.E.L. Lischer (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Res.* 10: 564-567.

Rousset, F. (2008). Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Mol. Ecol. Res.* 8: 103-106.

See Also

[batchjacklociArp](#), [convert](#), [convertjack](#)

jacklociArpRANDOM	<i>Remove a random subset of loci from a Arlequin project file (extension .arp)</i>
-------------------	---

Description

This function will create a new .arp file with only a random subset of the loci in the original dataset.

Usage

```
jacklociArpRANDOM(datafile, n.retain)
```

Arguments

datafile	Name of the file to jackknife, with the extension (e.g. 'glyptemys3.arp').
n.retain	Number of loci to keep in final file (must be between 1 and the number of loci in input file)

Details

This function is useful to investigate the effect of the number of sampled loci on population genetics parameter estimates.

Some .arp files might not be imported correctly depending on which software was used to generate them. If this is the case, start from the Genepop file and convert it first with [convert](#).

For converting multiple files, use the function [batchjacklociArpRANDOM](#) instead.

Value

A new .arp file will be saved in the working directory, with a suffix in the filename indicating the number of loci that were kept.

Author(s)

Sebastien Rioux Paquette

References

Excoffier, L. and H.E.L. Lischer (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Res.* 10: 564-567.

Rousset, F. (2008). Genepop'007: a complete reimplementaion of the Genepop software for Windows and Linux. *Mol. Ecol. Res.* 8: 103-106.

See Also

[batchjacklociArpRANDOM](#), [convert](#)

jackmsat

An internal function used by [jackmsatpop](#). Should not be used.

Description

Please see [jackmsatpop](#).

Usage

```
jackmsat(table, interval = 1, nrepet = 100, richness = F)
```

Arguments

table
interval
nrepet
richness

Value

See [jackmsatpop](#)

Author(s)

Sebastien Rioux Paquette

See Also

[jackmsatpop](#)

jackmsatpop	<i>From a Genepop input file (.gen), it calculates allele rarefaction curves within each population for a given sampling interval.</i>
-------------	--

Description

This function calculates (by jackknife resampling) the number of sampled alleles for a given constant increase in sample size within each population. The results can be used to trace allele rarefaction curves to assess whether it is likely that all genetic diversity was sampled and if population sample sizes are sufficient to do so. It is not a statistical test, but it is a useful graphical method to explore variation in the dataset.

Usage

```
jackmsatpop(datafile, ndigit = 3, interval = 1, nrepet = 100, richness = F)
```

Arguments

datafile	Name of the input Genepop datafile, with the extension (e.g. 'glyptemys3.gen')
ndigit	Number of digits per allele in the input file. Can be 2 or 3 (default is 3).
interval	This value indicates the number of individuals that should be added at each step. For instance, if interval = 10, the number of alleles will be estimated for 10, 20, 30 individuals until the true sample size is reached. Unless sample sizes are very large, an interval of 1 should be the most appropriate.
nrepet	Number of jackknife replicates to perform for each sampling interval.
richness	Default to FALSE. If set to TRUE, the function will do something completely different. Instead of estimating the number of alleles for each sampling interval, it will calculate allelic richness for a specific sample size. In this case, the sample size will be the value provided for 'interval'.

Details

Make sure to follow standard Genepop format for input file. The IDs of all individuals should be immediately followed by a comma. Population labels should be Pop, POP, or pop. There should not be invisible characters (tabs or spaces) at the end of lines, especially lines on which population labels are found. These lines should have only three characters (Pop). See example file 'glyptemys3.gen' if needed.

When richness is set to default, the function will save a result file for each population. This result file contains a table of average number of alleles discovered, where columns indicate different markers and rows different sampling intervals. A second table indicates standard deviation of the number of alleles.

When richness is TRUE, a single file is saved reporting number of alleles and standard deviation in tables of the form population x marker.

Anytime the actual sample size for any locus is smaller than the number of individuals the function is trying to resample within a population, the value '0' will be reported in the results table.

Value

All results file are saved in the working directory.

Author(s)

Sebastien Rioux Paquette

References

Rousset, F. (2008). Genepop'007: a complete reimplementaion of the Genepop software for Windows and Linux. *Mol. Ecol. Res.* 8: 103-106.

outputBottleSim	<i>Converts alleles from BottleSim simulation results to real allele sizes as found in the input file generated with the function inputBottleSim</i>
-----------------	--

Description

The population genetics simulation software BottleSim (Kuo and Janzen 2003) uses an unconventional input file format. The function [inputBottleSim](#) creates automatically such input files. After BottleSim simulations are performed, a Genepop file containing genotypes of all replicates (each replicate is labelled as a different 'population') is produced, but alleles are coded in a different manner that makes comparison with the starting real data impossible (for instance, one may be interested in calculating Fst values between simulated populations and the real data). This function converts alleles back to their original sizes. It also allows the user to retain only a random subset of individuals for each replicate. Real and simulated data can subsequently be combined for analysis with the function [combineBsim](#).

Usage

```
outputBottleSim(datafile, subsample = 0)
```

Arguments

datafile	Name of the output file generated by BottleSim (with the extension) containing multilocus genotypes for all simulation replicates.
subsample	Number of individuals to retain in each replicate. Individuals are chosen randomly in each replicate. Zero (default value) means all individuals will be kept.

Details

To use this function, the input file for BottleSim simulations first needs to be created using the function [inputBottleSim](#). This will also create a file called 'alleles' in which all required information about allele numbers and sizes is stored. Once BottleSim simulations are done and a genotype file is obtained, it is necessary to put the 'alleles' file in the same directory as the BottleSim genotype output file before using this function.

Value

A new Genepop file will be saved in the working directory, with the suffix 'OKAllelesBsim'.

Author(s)

Sebastien Rioux Paquette

References

Kuo C.H., and F.J.Janzen (2003). BottleSim: a bottleneck simulation program for long-lived species with overlapping generations. Mol. Ecol. Notes 3, 669-673.

Rousset, F. (2008). Genepop'007: a complete reimplementaion of the Genepop software for Windows and Linux. Mol. Ecol. Res. 8: 103-106.

See Also

[inputBottleSim](#), [combineBsim](#)

PICalc	<i>Compute the polymorphism information content (PIC) of each locus in a Arlequin project file (.arp)</i>
--------	---

Description

This function computes the PIC (Botstein et al. 1980) of each locus.

Usage

```
PICalc(datafile, ndigit = 3)
```

Arguments

datafile	Name of the input file, with the extension (e.g. 'glyptemys3.arp')
ndigit	Number of digit per allele in the input file. Default is 3.

Details

Some .arp files might not be imported correctly depending on which software was used to generate them. If this is the case, start from the Genepop file and convert it first with [convert](#).

Value

Returns a table of PIC values to the R console.

Author(s)

Sebastien Rioux Paquette

References

- Botstein, D., R.L. White, M. Skolnick, and A.W. Davis (1980). Construction of a Genetic Linkage Map in Man Using Restriction Fragment Length Polymorphisms. *Am. J. Hum. Genet.* 32: 314-331.
- Excoffier, L. and H.E.L. Lischer (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Res.* 10: 564-567.
- Rousset, F. (2008). Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Mol. Ecol. Res.* 8: 103-106.

See Also

[convert](#)

popgen	<i>Compute allele frequencies, population genetic parameters, and pairwise Fst values from a Arlequin (.arp) input file</i>
--------	---

Description

This function does not do anything that other population genetic software could not do, but provides a quick way to obtain allele frequencies in a table format (overall and within each population), and it can calculate allelic richness, number of private alleles, expected and observed heterozygosity (He and Ho), and population pairwise Fst values, for each locus and across all markers.

Usage

```
popgen(datafile, ndigit = 3, freq.overall = T, freq.by.pop = T, genetic.stats = T, pairwise.fst = T)
```

Arguments

datafile	Name of the input Arlequin file, with extension (e.g. 'glyptemys3.arp').
ndigit	Number of digits per allele in input file. Can be 2 or 3. (Default is 3).
freq.overall	Obtain a table of overall allele frequencies?
freq.by.pop	Obtain a table of allele frequencies in each population?
genetic.stats	Obtain a table listing number of alleles, allelic richness, private alleles, Ho and He?
pairwise.fst	Obtain a table of population pairwise Fst values, for each marker and overall?

Details

Some .arp files might not be imported correctly depending on which software was used to generate them. If this is the case, start from the Genepop file and convert it first with [convert](#).

All output files will be saved to the working directory.

Allelic richness is calculated by jackknife resampling (1000 replicates), and the sample size for each locus is determined as the smallest number of individuals sampled among all populations for that specific locus. Thus, the sample size may vary between markers.

Pairwise Fst values are calculated according to Wright's formula (Wright 1931; 1951). This option is provided as a data exploration tool. Numerous software allow the computation of Fst indices that integrate sampling effects, such as Weir and Cockerham's theta (1984), and these indices should be preferably reported. See Holsinger and Weir (2009) for more details.

Value

Results will be saved in tab-delimited text files in the working directory.

Author(s)

Sebastien Rioux Paquette

References

Excoffier, L. and H.E.L. Lischer (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Res.* 10: 564-567.

Holsinger, K.E. and B.S. Weir (2009) Genetics in geographically structured populations: defining, estimating and interpreting Fst. *Nat. Rev. Genet.* 10:639-650.

Rousset, F. (2008). Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Mol. Ecol. Res.* 8: 103-106.

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

Wright, S. (1931) Evolution in Mendelian populations. *Genetics* 16: 97-159.

Wright, S. (1951) The genetical structure of populations. *Annals of Eugenics* 15: 323-354.

recodeBi

Transform a multi-allelic Arlequin datafile to a bi-allelic datafile

Description

This function transforms a dataset with multi-allelic markers (e.g. microsatellites) to a bi-allelic dataset, where all alleles but one (usually the most frequent) are pooled together to form a second allelic class.

Usage

```
recodeBi(datafile, major.alleles)
```

Arguments

<code>datafile</code>	Name of the file to recode, with the extension (e.g. 'glyptemys3.arp').
<code>major.alleles</code>	A vector listing which allele to retain for each locus. Length of this vector should be equal to the number of markers in datafile.

Details

Because several population genetics parameter estimates are affected by the level of polymorphism of loci (e.g. *Fst*), it may be useful to recode multi-allelic markers to a bi-allelic state. For instance, this could allow the comparison of *Fst* values for microsatellites and single nucleotide polymorphisms (SNPs). McDonald (1994) introduced this as *F2st*, and it has been used elsewhere (e.g. Allendorf and Seeb 2000).

All alleles aside from the retained one at each locus are pooled together and designated as '300'. Thus, the allele retained at each locus should be different than 300.

The function [popgen](#) provides a quick way to obtain allele frequencies in order to determine which alleles are the most frequent, within populations or overall.

To transform multiple files, use [batchrecodeBi](#) instead. Some .arp files might not be imported correctly depending on which software was used to generate them. If this is the case, start from the Genepop file and convert it first with [convert](#).

Value

The Arlequin file coded in bi-allelic format is saved in the working directory.

Author(s)

Sebastien Rioux Paquette

References

- Allendorf, F.W. and L.W. Seeb (2000) Concordance of genetic divergence among sockeye salmon populations at allozyme, nuclear DNA, and mitochondrial DNA markers. *Evolution* 54: 640-651.
- Excoffier, L. and H.E.L. Lischer (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Res.* 10: 564-567.
- McDonald, J.H. (1994) Detecting natural selection by comparing geographic variation in protein and DNA polymorphisms. Pages 88-100 in Golding B, ed. *Non-neutral evolution: theories and molecular data*. New York: Chapman and Hall.
- Rousset, F. (2008). Genepop'007: a complete reimplementations of the Genepop software for Windows and Linux. *Mol. Ecol. Res.* 8: 103-106.

See Also

[batchrecodeBi](#), [convert](#), [popgen](#)

Index

*Topic `\textasciitildekwd1`

- batchconvert, 3
- batchconvertjack, 4
- batchjacklociArp, 5
- batchjacklociArpRANDOM, 6
- batchrecodeBi, 7
- bootstrapHet, 8
- bootstrapHet_byloc, 10
- combineBsim, 11
- convert, 12
- convertjack, 13
- freqPCA, 14
- inputBottleSim, 15
- jacklociArp, 16
- jacklociArpRANDOM, 17
- jackmsat, 18
- jackmsatpop, 19
- outputBottleSim, 20
- PICalc, 21
- popgen, 22
- recodeBi, 23

*Topic `\textasciitildekwd2`

- batchconvert, 3
- batchconvertjack, 4
- batchjacklociArp, 5
- batchjacklociArpRANDOM, 6
- batchrecodeBi, 7
- bootstrapHet, 8
- bootstrapHet_byloc, 10
- combineBsim, 11
- convert, 12
- convertjack, 13
- freqPCA, 14
- inputBottleSim, 15
- jacklociArp, 16
- jacklociArpRANDOM, 17
- jackmsat, 18
- jackmsatpop, 19
- outputBottleSim, 20

- PICalc, 21
- popgen, 22
- recodeBi, 23

- batchconvert, 3, 3, 5–8, 11–13
- batchconvertjack, 3, 4, 5, 6, 13, 14
- batchjacklociArp, 3, 5, 16, 17
- batchjacklociArpRANDOM, 3, 6, 17, 18
- batchrecodeBi, 3, 7, 24
- bootstrapHet, 3, 8, 10, 11
- bootstrapHet_byloc, 3, 9, 10

- combineBsim, 3, 11, 16, 20, 21
- convert, 3, 4, 9, 10, 12, 14, 16–18, 21–24
- convertjack, 3, 4, 13, 16, 17

- freqPCA, 3, 14

- inputBottleSim, 3, 11, 12, 15, 20, 21

- jacklociArp, 3, 5, 6, 16
- jacklociArpRANDOM, 3, 6, 7, 17
- jackmsat, 18
- jackmsatpop, 3, 18, 19

- outputBottleSim, 3, 11, 12, 15, 16, 20

- PICalc, 3, 21
- popgen, 3, 8, 22, 24
- PopGenKit (PopGenKit-package), 2
- PopGenKit-package, 2

- recodeBi, 3, 8, 23