Package 'StAMPP'

October 12, 2022

Type Package

Title Statistical Analysis of Mixed Ploidy Populations

Depends R (>= 3.2.0), pegas

Imports parallel, doParallel, foreach, adegenet, methods, utils

Version 1.6.3

Date 2021-07-30

Author LW Pembleton

Maintainer LW Pembleton <lwpembleton@gmail.com>

Description Allows users to calculate pairwise Nei's Genetic Distances (Nei 1972), pairwise Fixation
Indexes (Fst) (Weir & Cockerham 1984) and also Genomic Relationship matrixes follow-
ing Yang et al. (2010) in mixed and single
ploidy populations. Bootstrapping across loci is implemented during Fst calculation to gener-
ate confidence intervals and p-values
around pairwise Fst values. StAMPP utilises SNP geno-
type data of any ploidy level (with the ability to handle missing data) and is coded to
utilise multithreading where available to allow efficient analy-
sis of large datasets. StAMPP is able to handle genotype data from genlight objects
allowing integration with other packages such adegenet.
Please refer to LW Pembleton, NOI Cogan & JW Forster, 2013, Molecular Ecology Re-
sources, 13(5), 946-952. <doi:10.1111/1755-0998.12129> for the appropriate cita-
tion and user manual. Thank you in advance.

URL https://github.com/lpembleton/StAMPP

License GPL-3

RoxygenNote 7.1.1

Encoding UTF-8

NeedsCompilation no

Repository CRAN

Date/Publication 2021-08-08 04:20:05 UTC

R **topics documented:**

---

potato   *Example genotype input format*

Description

A data frame containing Solcap potato genotype data in tetraploid and diploid format as an small example of the input format required by StAMPP

Usage

```
data(potato)
```

Format

A data frame with 30 rows and 48 variables:

Sample Sample names

Pop Population name

Ploidy Ploidy level

Format Format of genotype data

solcap_snp_c1_1 genotype data

solcap_snp_c1_1000 genotype data

solcap_snp_c1_10000 genotype data

solcap_snp_c1_10001 genotype data

solcap_snp_c1_10011 genotype data

solcap_snp_c1_10012 genotype data

solcap_snp_c1_10031 genotype data

solcap_snp_c1_10042 genotype data

solcap_snp_c1_10050 genotype data

solcap_snp_c1_10054 genotype data

solcap_snp_c1_10109 genotype data

solcap_snp_c1_10130 genotype data

solcap_snp_c1_10157 genotype data

solcap_snp_c1_10202 genotype data

solcap_snp_c1_10252 genotype data

solcap_snp_c1_10253 genotype data

solcap_snp_c1_10255 genotype data

solcap_snp_c1_1029 genotype data

solcap_snp_c1_10295 genotype data

solcap_snp_c1_10297 genotype data

solcap_snp_c1_10351 genotype data

solcap_snp_c1_10384 genotype data

solcap_snp_c1_10397 genotype data

solcap_snp_c1_10457 genotype data

solcap_snp_c1_10491 genotype data

solcap_snp_c1_10492 genotype data

solcap_snp_c1_10494 genotype data

solcap_snp_c1_10579 genotype data

solcap_snp_c1_10646 genotype data

solcap_snp_c1_10669 genotype data

solcap_snp_c1_10715 genotype data

solcap_snp_c1_10737 genotype data

solcap_snp_c1_10743 genotype data

solcap_snp_c1_10762 genotype data

solcap_snp_c1_10855 genotype data

solcap_snp_c1_10873 genotype data

solcap_snp_c1_10879 genotype data

solcap_snp_c1_10900 genotype data

solcap_snp_c1_10932 genotype data

solcap_snp_c1_1094 genotype data

solcap_snp_c1_11137 genotype data

solcap_snp_c1_11144 genotype data

solcap_snp_c1_11196 genotype data

solcap_snp_c1_11206 genotype data

Source

The example genotype data is a subset of data from the publically avaliable Solcap potato dataset which was re-scored in GenomeStudio in diploid and tetraploid formats

---

potato.mini	*Smaller example genotype input format*

Description

A data frame containing Solcap potato genotype data in tetraploid and diploid format as an small example of the input format required by StAMPP

Usage

```
data(potato.mini)
```

Format

A data frame with 6 rows and 48 variables:

Sample Sample names

Pop Population name

Ploidy Ploidy level

Format Format of genotype data

solcap_snp_c1_1 genotype data

solcap_snp_c1_1000 genotype data

solcap_snp_c1_10000 genotype data

solcap_snp_c1_10001 genotype data

solcap_snp_c1_10011 genotype data

solcap_snp_c1_10012 genotype data

solcap_snp_c1_10031 genotype data

solcap_snp_c1_10042 genotype data

solcap_snp_c1_10050 genotype data

solcap_snp_c1_10054 genotype data

solcap_snp_c1_10109 genotype data

solcap_snp_c1_10130 genotype data

solcap_snp_c1_10157 genotype data

solcap_snp_c1_10202 genotype data

solcap_snp_c1_10252 genotype data

solcap_snp_c1_10253 genotype data

solcap_snp_c1_10255 genotype data

solcap_snp_c1_1029 genotype data

solcap_snp_c1_10295 genotype data

solcap_snp_c1_10297 genotype data

solcap_snp_c1_10351 genotype data

solcap_snp_c1_10384 genotype data

solcap_snp_c1_10397 genotype data

solcap_snp_c1_10457 genotype data

solcap_snp_c1_10491 genotype data

solcap_snp_c1_10492 genotype data

solcap_snp_c1_10494 genotype data

solcap_snp_c1_10579 genotype data

solcap_snp_c1_10646 genotype data

solcap_snp_c1_10669 genotype data

solcap_snp_c1_10715 genotype data

solcap_snp_c1_10737 genotype data

solcap_snp_c1_10743 genotype data

solcap_snp_c1_10762 genotype data

solcap_snp_c1_10855 genotype data

solcap_snp_c1_10873 genotype data

solcap_snp_c1_10879 genotype data

solcap_snp_c1_10900 genotype data

solcap_snp_c1_10932 genotype data

solcap_snp_c1_1094 genotype data

solcap_snp_c1_11137 genotype data

solcap_snp_c1_11144 genotype data

solcap_snp_c1_11196 genotype data

solcap_snp_c1_11206 genotype data

Source

The example genotype data is a subset of data from the publically avaliable Solcap potato dataset which was re-scored in GenomeStudio in diploid and tetraploid formats

---

stampp2genlight   *Convert StAMPP genotype data to genlight object*

Description

Converts a StAMPP formated allele frequency data frame generated from the stamppConvert function to a genlight object for use in other packages

Usage

```
stampp2genlight(geno, pop = TRUE)
```

Arguments

geno	a data frame containing allele frequency data generated from stamppConvert
pop	logical. True if population IDs are present in the StAMPP genotype data, False if population IDs are absent.

Details

StAMPP only exports to genlight objects as they are able to handle mixed ploidy datasets unlike genpop and genloci objects. The genlight object allows the intergration between StAMPP and other common R packages such as ADEGENET

Value

A object of class genlight which contains genotype data, individual IDs, population IDs (if present) and ploidy levels

Author(s)

Luke Pembleton <luke.pembleton at agriculture.vic.gov.au>

Examples

```
# import genotype data and convert to allele frequecies
data(potato.mini, package="StAMPP")
potato.freq <- stamppConvert(potato.mini, "r")
# Convert the StAMPP formatted allele frequency data frame to a genlight object
potato.genlight <- stampp2genlight(potato.freq, TRUE)
```

---

stamppAmova	*Analysis of Molecular Variance*

Description

Calculates an AMOVA based on the genetic distance matrix from stamppNeisD() using the amova() function from the package PEGAS for exploring within and between population variation

Usage

```
stamppAmova(dist.mat, geno, perm = 100)
```

Arguments

dist.mat	the matrix of genetic distances between individuals generated from stamppNeisD()
geno	a data frame containing allele frequency data generated from stamppConvert, or a genlight object containing genotype data, individual IDs, population IDs and ploidy levels
perm	the number of permutations for the tests of hypotheses

Details

Uses the formula distance ~ populations, to calculate an AMOVA for population differentiation and within & between population variation. This function uses the amova function from the PEGAS package.

Value

An object of class "amova" which is a list containing a table of sum of square deviations (SSD), mean square deviations (MSD) and the number of degrees of freedom as well as the variance components

Author(s)

Luke Pembleton <luke.pembleton at agriculture.vic.gov.au>

References

Paradis E (2010) pegas: an R package for population genetics with an integrated-modular approach. Bioinformatics 26, 419-420. <doi:10.1093/bioinformatics/btp696>

Examples

```
# import genotype data and convert to allele frequecies
data(potato.mini, package="StAMPP")
potato.freq <- stamppConvert(potato.mini, "r")
# Calculate genetic distance between individuals
potato.D.ind <- stamppNeisD(potato.freq, FALSE)
# Calculate AMOVA
stamppAmova(potato.D.ind, potato.freq, 100)
```

---

stamppConvert   *Import and Convert*

Description

Imports biallelic AB formated or allele A frequency genotype data. If the data is in imported in biallelic AB format this function also converts it to allele frequencies

Usage

```
stamppConvert(genotype.file, type = "csv")
```

Arguments

genotype.file   the genotype input file. This should be a R matrix object or a file path for a csv file containing the genotype data in either bialleleic AB format or allele 'A' frequency format, or a genlight object containing genotype data

type   the type of file the genotype data is being imported from; "csv" = comma seperated file, "r" = data frame in the R workspace, "genlight" = genlight object.

Value

An object of class data.frame which contains allele frequency data for use in other StAMPP functions

Author(s)

Luke Pembleton <luke.pembleton at agriculture.vic.gov.au>

Examples

```
# Import example data into the R workspace
data(potato.mini, package="StAMPP")
# Convert to allele frequencies
potato.freq <- stamppConvert(potato.mini, "r")
```

---

stamppFst   *Fst Computation*

Description

This function calculates pairwise Fst values along with confidence intervals and p-values between populations according to the method proposed by Wright(1949) and updated by Weir and Cockerham (1984)

Usage

```
stamppFst(geno, nboots = 100, percent = 95, nclusters = 1)
```

Arguments

geno	a data frame containing allele frequency data generated from stamppConvert, or a genlight object containing genotype data, individual IDs, population IDs and ploidy levels
nboots	number of bootstraps to perform across loci to generate confidence intervals and p-values
percent	the percentile to calculate the confidence interval around
nclusters	number of proccesor treads or cores to use during calculations.

Details

If possible, using multiple processing threads or cores is recommended to assist in calculating Fst values over a large number of bootstraps.

Value

An object list with the components: `Fsts` a matrix of pairwise Fst values between populations `Pvalues` a matrix of p-values for each of the pairwise Fst values containined in the 'Fsts' matrix `Bootstraps` a dataframe of each Fst value generated during Bootstrapping and the associated confidence intervals If nboots<2, no bootstrapping is performed and therefore only a matrix of Fst values is returned.

Author(s)

Luke Pembleton <luke.pembleton at agriculture.vic.gov.au>

References

Wright S (1949) The Genetical Structure of Populations. Annals of Human Genetics 15, 323-354. <doi:10.1111/j.1469-1809.1949.tb02451.x> Weir BS, Cockerham CC (1984) Estimating F Statistics for the ANalysis of Population Structure. Evolution 38, 1358-1370. <doi:10.2307/2408641>

Examples

```
# import genotype data and convert to allele frequecies
data(potato.mini, package="StAMPP")
potato.freq <- stamppConvert(potato.mini, "r")
# Calculate pairwise Fst values between each population
potato.fst <- stamppFst(potato.freq, 100, 95, 1)
```

---

stamppGmatrix   *Genomic Relationship Calculation*

Description

This function calculates a genomic relationship matrix following the method decribed by Yang et al (2010)

Usage

```
stamppGmatrix(geno)
```

Arguments

geno   a data frame containing allele frequency data generated from stamppConvert, or
                  a genlight object containing genotype data, individual IDs, population IDs and
                  ploidy levels

Value

An object of class matrix which contains the genomic relationship values between each individual

Author(s)

Luke Pembleton <luke.pembleton at agriculture.vic.gov.au>

References

Yang J, Benyamin B, McEvoy BP, et al (2010) Common SNPs explain a large proportion of the heritability for human height. Nat Genet 42, 565-569. <doi:10.1038/ng.608>

Examples

```
# import genotype data and convert to allele frequecies
data(potato.mini, package="StAMPP")
potato.freq <- stamppConvert(potato.mini, "r")
# Calculate genomic relationship values between each individual
potato.fst <- stamppGmatrix(potato.freq)
```

---

stamppNeisD   *Genetic Distance Calculation*

Description

This function calculates Nei's genetic distance (Nei 1972) between populations or individuals

Usage

```
stamppNeisD(geno, pop = TRUE)
```

Arguments

geno	a data frame containing allele frequency data generated from stamppConvert, or a genlight object containing genotype data, individual IDs, population IDs and ploidy levels
pop	logical. True if genetic distance should be calculated between populations, false if it should be calculated between individual

Value

A object of class matrix which contains the genetic distance between each population or individual

Author(s)

Luke Pembleton <luke.pembleton at agriculture.vic.gov.au>

References

Nei M (1972) Genetic Distance between Populations. The American Naturalist 106, 283-292.

Examples

```
# import genotype data and convert to allele frequecies
data(potato.mini, package="StAMPP")
potato.freq <- stamppConvert(potato.mini, "r")
# Calculate genetic distance between individuals
potato.D.ind <- stamppNeisD(potato.freq, FALSE)
# Calculate genetic distance between populations
potato.D.pop <- stamppNeisD(potato.freq, TRUE)
```

---

stamppPhylip *Export to Phylip Format*

Description

Converts the genetic distance matrix generated with stamppNeisD into Phylip format and exports it as a text file

Usage

```
stamppPhylip(distance.mat, file = "")
```

Arguments

distance.mat   the matrix containing the genetic distances generated from stamppNeisD to be converted into Phylip format

file   the file path and name to save the Phylip format matrix as

Details

The exported Phylip formated text file can be easily imported into sofware packages such as DAR-Win (Perrier & Jacquemound-Collet 2006) to be used to generate neighbour joining trees

Author(s)

Luke Pembleton <luke.pembleton at agriculture.vic.gov.au>

References

Perrier X, Jacquemound-Collet JP (2006) DARWin - Dissimilarity Analysis and Representation for Windows. Agricultural Research for Development

Examples

```
# import genotype data and convert to allele frequecies
data(potato.mini, package="StAMPP")
potato.freq <- stamppConvert(potato.mini, "r")
# Calculate genetic distance between populations
potato.D.pop <- stamppNeisD(potato.freq, TRUE)
# Export the genetic distance matrix in Phylip format
## Not run: stamppPhylip(potato.D.pop, file="potato_distance.txt")
```

Index