Package 'fhetboot'

June 24, 2016

Version 1.0

Title fhetboot: Fst-heterozygosity bootstrapping

Description
A program to generate bootstrapped confidence intervals for the Fst-heterozygosity distribution

Author Sarah P. Flanagan and Adam G. Jones

Maintainer Sarah P. Flanagan <spflanagan.phd@gmail.com>

License GPL-2

R topics documented:

`boot.out` *Example bootstrap output from numerical simulations*

Description

Example bootstrap output from numerical simulations It was generated by using a numerical analysis with Nm = 10, 75 demes, and 5 population samples taken. No selection was imposed. Ten bootstrap replicates were run on the example.genepop dataframe. This is a data.frame of lists. The first list is Fsts, which is a list of dataframes, each dataframe containing two columns: Ht and Fst. The second list is the the 95

Usage

```
boot.out
```

Format

data.frame

Source

From bootstrapping 10 reps over the dataframe gpop.

References

See Flanagan & Jones

`boot.out.list` *Example list of CI matrices from bootstrap output from numerical simulations*

Description

Example list of CI matrices from bootstrap output from numerical simulations The data were generated using a numerical analysis with Nm = 10, 75 demes, and 5 population samples taken. No selection was imposed. Ten bootstrap replicates were run on gpop. This is a lists of matrices containing the 10 sets of values from the 95 percent confidence intervals.

Usage

```
boot.out.list
```

Format

list

Source

From bootstrapping 10 reps over the dataframe gpop.

References

See Flanagan & Jones

---

calc.actual.fst *This calcualtes global Fsts from a genepop dataframe.*

Description

This calcualtes global Fsts from a genepop dataframe. This does not include bootstrapping.

Usage

```
calc.actual.fst(df)
```

Arguments

df	Provide the genepop dataframe (from my.read.genepop).

Value

fsts	This returns a dataframe with Locus, Ht, and Fst characters.

---

calc.allele.freq *This calculates allele frequencies.*

Description

This calculates allele frequencies from a list of genotypes.

Usage

```
calc.allele.freq(genotypes)
```

Arguments

genotypes	A list of genotypes.

Value

obs.af	A list of observed allele frequencies in the genotypes list.

calc.exp.het *This calculates expected heterozygosities.*

Description

This calculates expected heterozygosities from a list of allele frequencies.

Usage

```
calc.exp.het(af)
```

Arguments

af	is a list of allele frequencies.

Value

ht	The expected heterozygosity under Hardy-Weinberg expectations. This is a single numerical value.

calc.fst *This calculates Fst.*

Description

This calculates Fst. The caluclation is done as (Ht-Hs)/Ht, where Ht is the expected heterozygosity for all populations and Hs is the expected heterozygosity for each population. This calculation is used in bootstrapping functions.

Usage

```
calc.fst(df, i)
```

Arguments

df	A dataframe containing the genepop information, where the first column is the population ID.
i	Column number containing genotype information.

Value

ht	The expected heterozygosity under Hardy-Weinberg expectations. This is a single numerical value.
fst	The calculated Fst value for this locus.

ci.df — *Example dataframe of confidence intervals from bootstrapping.*

Description

Example dataframe of confidence intervals from bootstrapping. This file contains a dataframe with 22 columns and 4 rows. The rownames are the Ht x-value for plotting. The columns are the upper and lower 95 and 99 percent confidence intervals in the following order: low95, upp95, low99, upp99.

Usage

```
ci.df
```

Format

data.frame

Source

ci.df<-data.frame(low95=ci.means(boot.out[[2]])[[1]],upp95=ci.means(boot.out[[2]])[[2]], low99=ci.means(boot.out[[3]])[[

References

See Flanagan & Jones

ci.means — *This calculates the average confidence intervals from multiple bootstrap outputs.*

Description

This calculates the mean upper and lower confidence intervals from a list of bootstrap CI matrices.

Usage

```
ci.means(boot.out.list)
```

Arguments

boot.out.list	A list of matrices. Each matrix is either the 95 or 99 percent CIs from fst.boot.

Value

avg.cil	A list of the average lower CI values
avg.ciu	A list of the average upper CI values

`find.outliers` *This identifies all of the SNPs outside of the confidence intervals in the dataset.*

Description

This identifies all of the SNPs outside of the confidence intervals in the dataset.

Usage

```
find.outliers(df, ci.df = NULL, boot.out = NULL, file.name = NULL)
```

Arguments

`df`	Provide the dataframe with Ht and Fst values.
`boot.out`	Bootstrap output. You must either provide this or a list of confidence interval values.
`ci.df`	List of confidence intervals. You must either provide this or bootstrap output.
`file.name`	You may provide a file name to output the outliers to a csv file. Otherwise, the function will only return the 95 and 99 percent outliers.

Value

`out95`	A list of the 95 percent outlier loci
`out99`	A list of the 99 percent outlier loci

`fst.boot` *This is the major bootstrapping function to calculate confidence intervals.*

Description

This randomly samples all of the loci, with replacement (so if you have 200 loci, it will choose 200 loci to calculate Fst for, but some may be sampled multiply) It makes use of fst.boot.onerow. To calculate the confidence intervals, this function bins the Fst values based on heterozygosity values. The bins are overlapping and each bin is 0.1 Ht wide. The Fst value which separates the top 2.5 and bottom 2.5 percent in each bin are the upper and lower 95 percent CIs. The Fst value which separates the top 0.5 and bottom 0.5 percent in each bin are the upper and lower 99 percent CIs. This function can be slow. We recommend running it 10 times to generate confidence intervals for analysis.

Usage

```
fst.boot(df)
```

Arguments

df	A dataframe containing the genepop information, where the first column is the population ID.

Value

Fsts	The bootstrapped Fst and Ht values
CI95	A matrix with the lower 95 CI in column 1 and the upper 95 percent CI in column 2.
CI99	A matrix with the lower 99 CI in column 1 and the upper 99 percent CI in column 2.

---

fst.boot.onecol *This bootstraps across all individuals to calculate a bootstrapped Fst for a randomly-sampled locus.*

Description

This calculates Fst using calc.fst. It randomly selects a column containing genotype information for all individuals. It then calculates Fst and Ht for that locus.

Usage

```
fst.boot.onecol(df)
```

Arguments

df	A dataframe containing the genepop information, where the first column is the population ID.

Value

ht.fst	A vector containin Ht and Fst

---

fsts *Example fst calculations from a genepop file.*

Description

Example fst calculations from a genepop file. The original data were generated by using a numerical analysis with Nm = 10, 75 demes, and 5 population samples taken. No selection was imposed. The fsts were calculated using calc.actual.fst(gpop) This file contains a dataframe with 2000 columns and 3 rows. The first column is the Locus ID, the second column is the Ht for that locus, and the third column is the Fst for that locus.

Usage

```
fsts
```

Format

data.frame

Source

Generated by numerical analysis

References

See Flanagan & Jones

---

gpop *Example genepop file from numerical simulations*

Description

Example genepop file from numerical simulations. It was generated by using a numerical analysis with Nm = 10, 75 demes, and 5 population samples taken. No selection was imposed. This file contains a dataframe with 2002 columns and 250 rows. The first two columns are the population name and the individual name. The remaining columns are genotypes for each locus (one column per locus). Each row is an individual.

Usage

```
gpop
```

Format

data.frame

Source

Generated by numerical analysis

References

See Flanagan & Jones

my.read.genepop *This reads a genepop file into R*

Description

This reads a genepop file into R. It was adapted from a similar functionin adegenet.

Usage

```
my.read.genepop(file, ncode = 2L, quiet = FALSE)
```

Arguments

file	is the filename of the genpop file.
quiet	If quiet = FALSE updates will be printed. If quiet = T status updates will not be printed.
ncode	Do not change this argument.

Value

res	A dataframe with the Population ID in column 1, the Individual ID in column 2, and the genotypes in columns following that. There is one row per individual.

References

http://adegenet.r-forge.r-project.org/

plotting.cis *This plots a dataframe of fsts with bootstrapped confidence intervals.*

Description

This plots a dataframe of fsts with bootstrapped confidence intervals.

Usage

```
plotting.cis(df, boot.out, ci.list, Ht.name, Fst.name, ci.col, pt.pch,file.name,make.file)
```

Arguments

df	A dataframe of Fst and Ht values. It must have at least two columns, one named "Ht" and one named "Fst". Or you must pass the column names to the function
boot.out	Bootstrap output. You must either provide this or a list of confidence interval values.
ci.list	List of confidence intervals. You must either provide this or bootstrap output.
Ht.name	Provide the name of the column with the heterozygosity values, unless the column is named "Ht".
Fst.name	Provide the name of the column with the Fst values, unless the column is named "Fst".
ci.col	You can input the colors of the confidence intervals to be plotted. First is the 95 percent CI, second is the 99 percent CI. Defaults are "red" and "gold".
pt.pch	You can change the point shape here. Default is 1 (open circles)
file.name	You can provide the filename. If not provided, default is "OutlierLoci" in the current directory.
make.file	A boolean value (TRUE or FALSE). If TRUE, a file will be created with the plot. If FALSE, the plot will be made in R only (and can be further annotated).

---

remove.spaces *This removes spaces from a character vector*

Description

This removes spaces from a character vector. It was adapted from a similar functionin adegenet.

Usage

```
remove.spaces(charvec)
```

Arguments

charvec	is a vector of characters containing spaces to be removed.

Value

charvec	A vector of characters without spaces

References

http://adegenet.r-forge.r-project.org/

Index