# Package 'fhetboot'

January 5, 2017

**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

**Suggests** knitr, rmarkdown

**VignetteBuilder** knitr

## R topics documented:

allele.counts . . . . . 2
boot.means . . . . . 2
boot.out . . . . . 3
boot.out.list . . . . . 4
calc.actual.fst . . . . . 4
calc.allele.freq . . . . . 5
calc.exp.het . . . . . 5
calc.fst . . . . . 6
ci.df . . . . . 6
ci.means . . . . . 7
fhetboot . . . . . 7
find.outliers . . . . . 8
fst.boot . . . . . 8
fst.boot.means . . . . . 9
fst.boot.onecol . . . . . 10
fst.options.print . . . . . 10
fsts . . . . . 11
fsts.wc . . . . . 11
fsts.wcc . . . . . 12
gpop . . . . . 13
my.read.genepop . . . . . 13

p.boot . . . . . 14
plotting.cis . . . . . 14
remove.spaces . . . . . 15
wc.corr.fst . . . . . 16
wc.fst . . . . . 16

**Index** **18**

---

`allele.counts` *This counts the number of alleles at a locus.*

---

## Description

This counts the number of times each allele occurs at a locus from a list of genotypes (the sum of all the counts is 2*number of individuals).

## Usage

```
allele.counts(genotypes)
```

## Arguments

| | |
|---|---|
| `genotypes` | A list of genotypes. |

## Value

| | |
|---|---|
| `AlleleCounts` | The number of times each allele is recorded at the locus. |

---

`boot.means` *Example dataframe of mean Fst and heterozygosity from the bootstrapping for the bins.*

---

## Description

Example dataframe of mean Fst and Ht from bootstrapping. This file contains a dataframe with 22 columns and 5 rows. The columns are mean Ht, mean Fst, the number of loci in the bin, and the lower and upper bounds for each bin.

## Usage

```
boot.means
```

## Format

data.frame

## Source

boot.means<-fst.boot.means(boot.out)

## References

See Flanagan & Jones

---

`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 a list of data.frames, each containing the bins used in the bootstrapping module. The final list is a list of list. Each of the lists is a list of the upper and lower confidence intervals.

## 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, fst.choice="WCC")
```

## Arguments

| | |
|---|---|
| `df` | Provide the genepop dataframe (from my.read.genepop). |
| `fst.choice` | Specify which type of fst calculation should be used. See fst.options.print for the choices. |

## 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 2 rows. The rownames are the Ht x-value for plotting. The columns are the upper and lower confidence intervals.

### Usage

```
ci.df
```

### Format

data.frame

### Source

ci.df<-data.frame(do.call(cbind(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 the CIs from fst.boot (boot.out[[3]]). |

## Value

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

---

`fhetboot` *This is a wrapper to run the bootstrapping and plot the confidence intervals and significant loci.*

---

## Description

This calcualtes global Fsts from a genepop dataframe and then does: p-value calculations plots the Heterozygosity-Fst relationship with smoothed CIs outputs the loci lying outside the confidence intervals. Returns a data frame containing Locus ID, Ht, Fst, P-value, a Benjamini-Hochberg-corrected P-value, and a true/false value of whether it's an outlier.

## Usage

```
fhetboot(gpop, fst.choice, alpha,nreps)
```

## Arguments

| | |
|---|---|
| `gpop` | Provide the genepop dataframe (from my.read.genepop). |
| `fst.choice` | Specify which type of fst calculation should be used. See fst.options.print for the choices. |
| `alpha` | The alpha value for the confidence intervals and the p-value adjustment calculations (default is 0.05). |
| `nreps` | The number of bootstrap replicates to use. The default is 10. |

### Value

| | |
|---|---|
| fsts | This returns a dataframe with Locus, Ht, Fst, P-value, correcte P-value, and True/False of whether it's an outlier. |

---

## 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, boot.out, ci.df = NULL, file.name = NULL)
```

### Arguments

| | |
|---|---|
| df | Provide the dataframe with Ht and Fst values. |
| boot.out | Bootstrap output. You must provide this. |
| ci.df | List of confidence intervals. You may provide this in addition to bootstrap output to save a small amount of time. |
| file.name | You may provide a file name to output the outliers to a csv file. Otherwise, the function will only return the outliers. |

### Value

| | |
|---|---|
| out | A list of the 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 the width of smooth.rate. The Fst value which separates the top 100*(ci/2) and bottom 100*(ci/2) percent in each bin are the upper and lower CIs. This function can be slow. We recommend running it 10 times to generate confidence intervals for analysis.

### Usage

```
fst.boot(df,fst.choice="WCC",ci=0.05,num.breaks=25)
```

### Arguments

| | |
|---|---|
| df | A dataframe containing the genepop information, where the first column is the population ID. |
| fst.choice | A character defining which fst calculation is to be used. The three options are: Nei's Fst (nei,Nei,NEI,N) Weir and Cockerham 1993's beta (WeirCockerham,weircockerham,wc,WC) Corrected Weir and Cockerham 1993's beta from Beaumont and Nichols 1996 (WeirCockerhamCorrected, weircockerhamcorrected,corrected,wcc,WCC) Default is Nei's. |
| ci | A value for the confidence intervals alpha (default is 0.05). |
| num.breaks | The number of breaks used to create bins (default is 25) |

### Value

| | |
|---|---|
| Fsts | The bootstrapped Fst and Ht values |
| Bins | A dataframe containing the bins start and stop Ht values. |
| fst.CI | A list of dataframes containing the lower and upper confidence intervals' Ht values. |

---

## fst.boot.means *Calculates mean values within the bins.*

---

### Description

This calculates mean heterozygosity and Fst values for each bin used in bootstrapping.

### Usage

```
fst.boot.means(boot.out)
```

### Arguments

| | |
|---|---|
| boot.out | The first item in the output lists from fst.boot (aka boot.out[[1]]. |

### Value

| | |
|---|---|
| bmu | A dataframe containing four columns: heterozygosity Fst the number of loci in the bin the lower Ht value for the bin and the upper Ht value for the bin. |

`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, fst.choice)
```

## Arguments

| | |
|---|---|
| `df` | A dataframe containing the genepop information, where the first column is the population ID. |
| `fst.choice` | A character defining which fst calculation is to be used. The three options are: Nei's Fst (nei,Nei,NEI,N) Weir and Cockerham 1993's beta (WeirCockerham,weircockerham,wc,WC) Corrected Weir and Cockerham 1993's beta from Beaumont and Nichols 1996 (WeirCockerhamCorrected, weircockerhamcorrected,corrected,wcc,WCC) |

## Value

| | |
|---|---|
| `ht.fst` | A vector containin Ht and Fst |

`fst.options.print` *This prints the options for choosing an Fst calculation.*

## Description

This prints the options for choosing an Fst calculation.

## Usage

```
fst.options.print()
```

## 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

## fsts.wc — *Example fst calculations from a genepop file.*

### Description

Example fst calculations from a genepop file. The Fsts were calculated using the Weir and Cockerham (1993) calculation. 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.wc
```

### Format

data.frame

## Source

Generated by numerical analysis

## References

See Flanagan & Jones

---

`fsts.wcc` *Example fst calculations from a genepop file.*

---

## Description

Example fst calculations from a genepop file. The Fsts were calculated using the sample-size corrected Weir and Cockerham (1993) calculation used in FDIST2. 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.wcc
```

## 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/

---

## p.boot *Calculates mean values within the bins.*

---

### Description

This calculates mean heterozygosity and Fst values for each bin used in bootstrapping.

### Usage

```
p.boot(actual.fsts, boot.out,boot.means=NULL)
```

### Arguments

| | |
|---|---|
| actual.fsts | The first item in the output lists from fst.boot. |
| boot.out | The output from a bootstrapping run. Either supply this or boot.means. |
| boot.means | The output from fst.boot.means. Either supply this or bootstrapping output. |

### Value

| | |
|---|---|
| pvals | A numeric containing uncorrected p-values for each locus. The names attribute are the locus names. |

---

## 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.df=NULL,sig.list=NULL,Ht.name="Ht",Fst.name="Fst",
ci.col="red", pt.pch=1,file.name=NULL,sig.col=ci.col,make.file=TRUE)
```

### 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.df | Data frame of confidence intervals. You must either provide this or bootstrap output. |
| sig.list | List of significant locus names (this acts as a way to highlight particular loci). This is optional and colors some of the points using the same shape as pt.pch and the color of sig.col (default sig.color is same as ci.col). |
| 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) |
| sig.col | The color of the significant loci, if that option is taken. The default is the same color as the confidence interval. |
| 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/

## wc.corr.fst *This calculates Beaumont & Nichols's Fst.*

### Description

This calculates Beaumont & Nichols (1996)'s Fst. This is just a sample-size corrected version of the Weir & Cockerham (1993)'s beta. The caluclation is done as beta=(q2-q3)/(1-q3), where: q3=2Y/(N(N-1)), q2=x0/N, N=number of populations x0=sum(sum((cij*cij)-nj)/(nj(nj-1))) where cij is the number of allele i (allele count i) in group j and nj is the sample size of population j Y=sum((sum(cij*cik)/(nj*nk))) where cij is allele count i in population j and cik is allele count i in population k and nj is the sample size in pop j and nk is the sample size in pop k.

This calculation is used in bootstrapping functions.

### Usage

```
wc.corr.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 | 1-q3. This is a single numerical value. |
| fst | The calculated Fst value ((q2-q3)/(1-q3))for this locus. |

## wc.fst *This calculates Weir & Cockerham's Fst.*

### Description

This calculates Weir & Cockerham (1993)'s Fst. The caluclation is done as beta=(F0-F1)/(1-F1), where: F0=(M*X-N)/((M-1)*N), F1=(Y-X)/(N*(N-1)), N=number of populations M=average number of individuals per population X=sum(sum(pij^2) Y=sum((sum(pij)^2) where pij is the frequency of allele i in group j This calculation is used in bootstrapping functions.

### Usage

```
wc.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 | 1-F1. This is a single numerical value. |
| fst | The calculated Fst value ((F0-F1)/(1-F1))for this locus. |

# Index

allele.counts, 2

boot.means, 2
boot.out, 3
boot.out.list, 4

calc.actual.fst, 4
calc.allele.freq, 5
calc.exp.het, 5
calc.fst, 6
ci.df, 6
ci.means, 7

fhetboot, 7
find.outliers, 8
fst.boot, 8
fst.boot.means, 9
fst.boot.onecol, 10
fst.options.print, 10
fsts, 11
fsts.wc, 11
fsts.wcc, 12

gpop, 13

my.read.genepop, 13

p.boot, 14
plotting.cis, 14

remove.spaces, 15

wc.corr.fst, 16
wc.fst, 16