

# Package ‘FinePop2’

February 28, 2020

**Type** Package

**Title** Fine-Scale Population Analysis (Rewrite for Gene-Trait-Environment Interaction Analysis)

**Version** 0.4

**Date** 2020-02-20

**Author** Reiichiro Nakamichi, Shuichi Kitada, Hirohisa Kishino

**Maintainer** Reiichiro Nakamichi <nakamichi@affrc.go.jp>

**Description** Statistical tool set for population genetics. The package provides following functions: 1) estimators of genetic differentiation (FST), 2) regression analysis of environmental effects on genetic differentiation using generalized least squares (GLS) method, 3) interfaces to read and manipulate 'GENEPOP' format data files). For more information, see Kitada, Nakamichi and Kishino (2020) <doi:10.1101/2020.01.30.927186>.

**License** GPL (>= 2.0)

**Depends** R (>= 3.4.0)

**LazyLoad** yes

**NeedsCompilation** no

**Encoding** UTF-8

**Repository** CRAN

**Date/Publication** 2020-02-28 10:20:02 UTC

## R topics documented:

FinePop2-package . . . . .	2
clip.GENEPOP . . . . .	2
globalFST . . . . .	4
GLS . . . . .	5
herring . . . . .	6
jsmackerel . . . . .	7
locus_specificFST . . . . .	8
pop_pairwiseFST . . . . .	9
pop_specificFST . . . . .	11
read.GENEPOP . . . . .	12

**Index****14**


---

FinePop2-package	<i>Fine-Scale Population Analysis (Rewrite for Gene-Trait-Environment Interaction Analysis)</i>
------------------	---

---

**Description**

Statistical tool set for population genetics. The package provides following functions: 1) estimators of genetic differentiation (FST), 2) regression analysis of environmental effects on genetic differentiation using generalized least squares (GLS) method, 3) interfaces to read and manipulate 'GENEPOP' format data files). For more information, see Kitada, Nakamichi and Kishino (2020) <doi:10.1101/2020.01.30.927186>.

**Author(s)**

Reiichiro Nakamichi, Shuichi Kitada, Hirohisa Kishino  
 Maintainer: Reiichiro Nakamichi <nakamichi@affrc.go.jp>

**References**

Kitada S, Nakamichi R, Kishino H (2017) The empirical Bayes estimators of fine-scale population structure in high gene flow species. *Mol. Ecol. Resources*, 17, 1210-1222.  
 Nei M, Chesser RK (1983) Estimation of fixation indices and gene diversity. *Annals of Human Genetics*, 47, 253-259.  
 Rousset F (2008) Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Mol. Ecol. Resources*, 8, 103-106.  
 Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, 38, 1358-1370.  
 Weir BS, Goudet J (2017) A Unified Characterization of Population Structure and Relatedness. *Genetics*, 206, 2085-2103.

---

clip.GENEPOP	<i>Remove designated markers from a GENEPOP file.</i>
--------------	---

---

**Description**

This function reads a GENEPOP file (Rousset 2008), remove designated markers, and write a GENEPOP file of clipped data. The user can directly designate the names of the markers to be removed. The user also can set the filtering threshold of major allele frequency.

**Usage**

```
clip.GENEPOP(infile, outfile, remove.list = NULL, major.af = NULL)
```

**Arguments**

<code>infile</code>	A character value specifying the name of the GENEPOP file to be clipped.
<code>outfile</code>	A character value specifying the name of the clipped GENEPOP file.
<code>remove.list</code>	A character value or vector specifying the names of the markers to be removed. The names must be included in the target GENEPOP file.
<code>major.af</code>	A numeric value specifying the threshold of major allele frequency for marker removal. Markers with major allele frequencies higher than this value will be removed. This value must be between 0 and 1.

**Author(s)**

Reiichiro Nakamichi

**References**

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

**Examples**

```
# Example of GENEPOP file
data(jsmackerel)
jsm.genepop.file <- tempfile()
cat(jsmackerel$MS.genepop, file=jsm.genepop.file, sep="\n")

# Remove markers designated by their names
clipped_by_name.jsm.genepop.file <- tempfile()
clip.GENEPOP(infile=jsm.genepop.file,
             outfile=clipped_by_name.jsm.genepop.file,
             remove.list=c("Sni21","Sni26"))

# Remove markers with high major allele frequencies (in this example, > 0.5)
clipped_by_af.jsm.genepop.file <- tempfile()
clip.GENEPOP(infile=jsm.genepop.file,
             outfile=clipped_by_af.jsm.genepop.file,
             major.af=0.5)

# Remove markers both by their names and by major allele frequencies
clipped_by_both.jsm.genepop.file <- tempfile()
clip.GENEPOP(infile=jsm.genepop.file,
             outfile=clipped_by_both.jsm.genepop.file,
             remove.list=c("Sni21","Sni26"), major.af=0.5)

# See four text files in temporary directory.
# jsm.genepop.file           : original data of five markers
# clipped_by_name.jsm.genepop.file : clipped data by marker names
# clipped_by_af.jsm.genepop.file   : clipped data by allele frequency
# clipped_by_both.jsm.genepop.file : clipped data by both names and frequency
```

---

globalFST

*Genome-wide global FST (Weir & Cockerham 1984).*

---

### Description

This function estimates genom-wide global FST based on Weir and Cockerham's theta (Weir & Cockerham 1984) from a GENEPOP data object (Rousset 2008). Missing genotype values in the GENEPOP file ("0000" or "000000") are simply ignored.

### Usage

```
globalFST(popdata)
```

### Arguments

popdata	Population data object created by read.GENEPOP function from a GENEPOP file.
---------	--

### Value

fst	Estimated genome-wide global FST
se	Standard error of estimated FST

### Author(s)

Reiichiro Nakamichi, Shuichi Kitada, Hirohisa Kishino

### References

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

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, 38, 1358-1370.

### See Also

[read.GENEPOP](#)

### Examples

```
# Example of GENEPOP file
data(jsmackerel)
jms.ms.genepop.file <- tempfile()
jms.popname.file <- tempfile()
cat(jsmackerel$MS.genepop, file=jms.ms.genepop.file, sep="\n")
cat(jsmackerel$popname, file=jms.popname.file, sep=" ")
```

```

# Data load
# Prepare your GENEPOP file and population name file in the working directory.
# Replace "jsm.ms.genepop.file" and "jsm.popname.file" by your file names.
popdata <- read.GENEPOP(genepop=jsm.ms.genepop.file, popname=jsm.popname.file)

# theta estimation
result.globalFST <- globalFST(popdata)
print(result.globalFST)

```

---

GLS *Generalized least squares for regression analysis considering auto correlation.*

---

### Description

This function provides a multiple regression analysis considering auto correlation of response variable using generalized least squares method (Aitken 1934). It supports `lm` like format of model. A typical model has the form `response ~ terms`. Terms specification supports only first + second form. Cross term specification of `first * second` form is not supported.

### Usage

```
GLS(model, data, omega = NULL)
```

### Arguments

<code>model</code>	Symbolic description of the model to be fitted.
<code>data</code>	Data frame containing variables in the model.
<code>omega</code>	A numeric matrix of auto correlation of response variable.

### Value

<code>coefficients</code>	Estimated coefficient, standard error, Z value and p value of each factor.
<code>variance</code>	Variance-covariance matrix of estimated coefficients.
<code>logL</code>	Log likelihood of fitted model.

### Author(s)

Reiichiro Nakamichi, Shuichi Kitada, Hirohisa Kishino

### References

Aitken AC (1934) On Least-squares and Linear Combinations of Observations. *Proceedings of the Royal Society of Edinburgh*, 55, 42-48.

### See Also

[lm](#)

**Examples**

```

# Example data of Atlantic herring
data(herring)
ah.genepop.file <- tempfile()
ah.popname.file <- tempfile()
cat(herring$genepop, file=ah.genepop.file, sep="\n")
cat(herring$popname, file=ah.popname.file, sep=" ")

# Data load
popdata <- read.GENEPOP(ah.genepop.file, ah.popname.file)

# Pop-specific FST and correlation among populations
fst.posp <- pop_specificFST(popdata, cov=TRUE)
cov.fst.posp <- fst.posp$cov
sd.fst.posp <- sqrt(diag(cov.fst.posp))
cov2.fst.posp <- apply(cov.fst.posp, 2, function(x){x / sd.fst.posp})
cor.fst.posp <- apply(cov2.fst.posp, 1, function(x){x / sd.fst.posp})

# Pop-pairwise FST and population structure
fst.poppair <- pop_pairwiseFST(popdata)
fst.md <- cmdscale(fst.poppair)

# GLS analysis of FST and environmental factors
test.data <- data.frame(fst=fst.md[,1], herring$environment)
GLS(fst~., scale(test.data), omega=cor.fst.posp)

```

---

herring

*An example dataset of Atlantic herring.*


---

**Description**

An example of a genetic data for Atlantic herring population (Limborg et al. 2012). It contains genotypic information of 281 SNPs from 18 subpopulations of 607 individuals. GENEPOP format (Rousset 2008) text file is available. Subpopulation names, environmental factors (longitude, latitude, temperature and salinity) at each subpopulation are attached.

**Usage**

```
data("herring")
```

**Format**

\$ genepop : Genotypic information of 281 SNPs in GENEPOP format text data.  
 \$ popname : Names of subpopulations.  
 \$ environment : Table of temperature and salinity at each subpopulation.

## References

Limborg MT, Helyar SJ, de Bruyn M et al. (2012) Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (*Clupea harengus*). *Molecular Ecology*, 21, 3686-3703.

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

## See Also

[read.GENEPOP](#), [GLS](#)

## Examples

```
data(herring)
ah.genepop.file <- tempfile()
ah.popname.file <- tempfile()
cat(herring$genepop, file=ah.genepop.file, sep="\n")
cat(herring$popname, file=ah.popname.file, sep=" ")

# See two text files in temporary directory.
# ah.genepop.file : GENEPOP format file of 281SNPs in 18 subpopulations
# ah.popname.file : plain text file of subpopulation names

print(herring$environment)
```

---

jsmackerel

*An example dataset of Japanese Spanish mackerel in GENEPOP and frequency format.*

---

## Description

An example of a genetic data for a Japanese Spanish mackerel population (Nakajima et al. 2014). It contains genotypic information of 5 microsatellite markers from 8 subpopulations of 715 individuals. GENEPOP format (Rousset 2008) text files are available. Name list of subpopulations also is attached.

## Usage

```
data("jsmackerel")
```

## Format

\$ MS.genepop: Genotypic information of 5 microsatellites in GENEPOP format text data.  
\$ popname: Names of subpopulations.

## References

Nakajima K et al. (2014) Genetic effects of marine stock enhancement: a case study based on the highly piscivorous Japanese Spanish mackerel. *Canadian Journal of Fisheries and Aquatic Sciences*, 71, 301-314.

Kitada S, Kitakado T, Kishino H (2007) Empirical Bayes inference of pairwise FST and its distribution in the genome. *Genetics*, 177, 861-873.

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

## See Also

[read.GENEPOP](#)

## Examples

```
data(jsmackerel)
jsm.ms.genepop.file <- tempfile()
jsm.popname.file <- tempfile()
cat(jsmackerel$MS.genepop, file=jsm.ms.genepop.file, sep="\n")
cat(jsmackerel$popname, file=jsm.popname.file, sep=" ")

# See two text files in temporary directory.
# jsm.ms.genepop.file : GENEPOP format file of microsatellite data
# jsm.popname.file   : plain text file of subpopulation names
```

---

locus\_specificFST      *Locus-specific global FST (Kitada et al. 2007, 2017).*

---

## Description

This function estimates locus-specific global FST among subpopulations using empirical Bayes method (Kitada et al. 2007, 2017) from a GENEPOP data object (Rousset 2008). Missing genotype values in the GENEPOP file ("0000" or "000000") are simply ignored.

## Usage

```
locus_specificFST(popdata)
```

## Arguments

popdata            Population data object created by read.GENEPOP function from a GENEPOP file.

## Value

Estimated locus-specific global FST.



**Author(s)**

Reiichiro Nakamichi, Shuichi Kitada, Hirohisa Kishino

**References**

Kitada S, Kitakado T, Kishino H (2007) Empirical Bayes inference of pairwise FST and its distribution in the genome. *Genetics*, 177, 861-873.

Kitada S, Nakamichi R, Kishino H (2017) The empirical Bayes estimators of fine-scale population structure in high gene flow species. *Mol. Ecol. Resources*, DOI: 10.1111/1755-0998.12663

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

**See Also**

[read.GENEPOP](#), [read.frequency](#)

**Examples**

```
# Example of GENEPOP file
data(jsmackerel)
jsm.ms.genepop.file <- tempfile()
jsm.popname.file <- tempfile()
cat(jsmackerel$MS.genepop, file=jsm.ms.genepop.file, sep="\n")
cat(jsmackerel$popname, file=jsm.popname.file, sep=" ")

# Data load
# Prepare your GENEPOP file and population name file in the working directory.
# Replace "jsm.ms.genepop.file" and "jsm.popname.file" by your file names.
popdata <- read.GENEPOP(genepop=jsm.ms.genepop.file, popname=jsm.popname.file)

# FST estimation
locspFST <- locus_specificFST(popdata)
print(locspFST)
```

---

pop\_pairwiseFST      *Genome-wide population-paiwise FST (Nei & Chesser 1983).*

---

**Description**

This function estimates genome-wide population-paiwise FST among subpopulations based on Nei and Chesser's corrected GST (Nei&Chesser 1983) from a GENEPOP data object (Rousset 2008). Missing genotype values in the GENEPOP file ("0000" or "000000") are simply ignored.

**Usage**

```
pop_pairwiseFST(popdata)
```

**Arguments**

popdata            Population data object created by read.GENEPOP function from a GENEPOP file.

**Value**

Estimated genome-wide population-pairwise FST.

**Author(s)**

Reiichiro Nakamichi, Shuichi Kitada, Hirohisa Kishino

**References**

Nei M, Chesser RK (1983) Estimation of fixation indices and gene diversity. *Annals of Human Genetics*, 47, 253-259.

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

**See Also**

[read.GENEPOP](#), [as.dist](#), [as.dendrogram](#), [hclust](#), [cmdscale](#)

**Examples**

```
# Example of GENEPOP file
data(jsmackerel)
jsm.ms.genepop.file <- tempfile()
jsm.popname.file <- tempfile()
cat(jsmackerel$MS.genepop, file=jsm.ms.genepop.file, sep="\n")
cat(jsmackerel$popname, file=jsm.popname.file, sep=" ")

# Data load
# Prepare your GENEPOP file and population name file in the working directory.
# Replace "jsm.ms.genepop.file" and "jsm.popname.file" by your file names.
popdata <- read.GENEPOP(genepop=jsm.ms.genepop.file, popname=jsm.popname.file)

# GST estimation
result.poppairFST <- pop_pairwiseFST(popdata)
poppairFST.d <- as.dist(result.poppairFST)
print(poppairFST.d)

# dendrogram
poppairFST.hc <- hclust(poppairFST.d,method="average")
plot(as.dendrogram(poppairFST.hc), xlab="",ylab="",main="", las=1)

# MDS plot
mds <- cmdscale(poppairFST.d)
plot(mds, type="n", xlab="",ylab="")
text(mds[,1],mds[,2], popdata$pop_names)
```

---

pop_specificFST	<i>Genome-wide population-specific FST (Weir &amp; Goudet 2017).</i>
-----------------	--

---

### Description

This function estimates genome-wide population-specific FST based on Weir and Goudet's Method (Weir&Goudet 2017) from a GENEPOP data object (Rousset 2008). Missing genotype values in the GENEPOP file ("0000" or "000000") are simply ignored.

### Usage

```
pop_specificFST(popdata, cov = FALSE)
```

### Arguments

popdata	Population data object created by read.GENEPOP function from a GENEPOP file.
cov	A Logical argument indicating whether variance-covariance matrix of estimated FST should be calculated.

### Value

fst	Estimated genome-wide population-specific FST and standard error.
cov	Variance-covariance matrix of estimated FST

### Author(s)

Reiichiro Nakamichi, Shuichi Kitada, Hirohisa Kishino

### References

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

Weir BS, Goudet J (2017) A Unified Characterization of Population Structure and Relatedness. *Genetics*, 206, 2085-2103.

### See Also

[read.GENEPOP](#)

**Examples**

```
# Example of GENEPOP file
data(jsmackerel)
jsm.ms.genepop.file <- tempfile()
jsm.popname.file <- tempfile()
cat(jsmackerel$MS.genepop, file=jsm.ms.genepop.file, sep="\n")
cat(jsmackerel$popname, file=jsm.popname.file, sep=" ")

# Data load
# Prepare your GENEPOP file and population name file in the working directory.
# Replace "jsm.ms.genepop.file" and "jsm.popname.file" by your file names.
popdata <- read.GENEPOP(genepop=jsm.ms.genepop.file, popname=jsm.popname.file)

# FST estimation
result.pospFST <- pop_specificFST(popdata)
print(result.pospFST$fst)
```

---

read.GENEPOP	<i>Create a genotype data object of populations from a GENEPOP format file.</i>
--------------	---

---

**Description**

This function reads a GENEPOP format file (Rousset 2008) and parse it into an R data object. This data object provides a summary of genotype/haplotype of each sample, allele frequency in each population, and marker status. This data object is used in downstream analysis of this package.

**Usage**

```
read.GENEPOP(genepop, popname = NULL)
```

**Arguments**

genepop	A character value specifying the name of the GENEPOP file to be analyzed.
popname	A character value specifying the name of the plain text file containing the names of subpopulations to be analyzed. This text file must not contain other than subpopulation names. The names must be separated by spaces, tabs or line breaks. If this argument is omitted, serial numbers will be assigned as subpopulation names.

**Value**

num_pop	Number of subpopulations.
pop_sizes	Number of samples in each subpopulation.
pop_names	Names of subpopulations.
ind_names	Names of samples in each subpopulation.
num_loci	Number of loci.

loci_names	Names of loci.
num_allele	Number of alleles at each locus.
allele_list	A list of alleles at each locus.
ind_count	Observed count of genotyped samples in each subpopulation at each locus.
allele_count	Observed count of genotyped alleles in each subpopulation at each locus.
allele_freq	Observed allele frequencies in each subpopulation at each locus.
genotype	Genotypes of each sample at each locus in haploid designation.
call_rate_loci	Call rate of each locus (rate of genotyped samples at each locus).
call_rate_ind	Call rate of each sample (rate of genotyped markers for each sample).
He	Expected heterozygosity in each subpopulation.
Ho	Observed heterozygosity in each subpopulation.

**Author(s)**

Reiichiro Nakamichi

**References**

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

**Examples**

```
# Example of GENEPOP file
data(jsmackerel)
jsm.ms.genepop.file <- tempfile()
jsm.popname.file <- tempfile()
cat(jsmackerel$MS.genepop, file=jsm.ms.genepop.file, sep="\n")
cat(jsmackerel$popname, file=jsm.popname.file, sep=" ")

# Read GENEPOP file with subpopulation names.
# Prepare your GENEPOP file and population name file in the working directory.
# Replace "jsm.ms.genepop.file" and "jsm.popname.file" by your file names.
popdata <- read.GENEPOP(genepop=jsm.ms.genepop.file, popname=jsm.popname.file)

# Read GENEPOP file without subpopulation names.
popdata.noname <- read.GENEPOP(genepop=jsm.ms.genepop.file)
```

# Index

\*Topic **datasets**

herring, [6](#)

jsmackerel, [7](#)

\*Topic **package**

FinePop2-package, [2](#)

as.dendrogram, [10](#)

as.dist, [10](#)

clip.GENEPop, [2](#)

cmdscale, [10](#)

FinePop2-package, [2](#)

globalFST, [4](#)

GLS, [5](#), [7](#)

hclust, [10](#)

herring, [6](#)

jsmackerel, [7](#)

lm, [5](#)

locus\_specificFST, [8](#)

pop\_pairwiseFST, [9](#)

pop\_specificFST, [11](#)

read.frequency, [9](#)

read.GENEPop, [4](#), [7–11](#), [12](#)