Package 'Geneland'

February 19, 2015

Title Detection of structure from multilocus genetic data.

Author Gilles Guillot, Filipe Santos, Arnaud Estoup

Description Stochastic simulation and MCMC inference of structure from genetic data.

Maintainer Gilles Guillot <gigu@dtu.dk>

URL http://www2.imm.dtu.dk/~gigu/Geneland/

License GPL

Suggests mapproj,maps

Version 4.0.5

Date Mon Jan 12 18:56:19 2015

Depends RandomFields,fields,tcltk

NeedsCompilation yes

Repository CRAN

Date/Publication 2015-01-12 22:51:20

R topics documented:

`Geneland-package` *Simulation and inference for subdivided populations*

Description

Detect population structure (i.e sub-populations), making use of genetic (and optionally geographic) information.

Details

The main purpose of the program is to perform Bayesian inference of all the parameters involved through Markov Chain Monte-Carlo simulation. This is achievied by the function `MCMC`. Function `PostProcessChain` read some output files of `MCMC` and computes some statistics suitable to print maps of inferred populations.

See `Storage format` section in `MCMC help page`.

The following functions are provided by the package:

`simFmodel`: simulation from the prior of the spatial F-model

`simdata`: Simulation of georeferenced genotypes under an IBD + barrier model

`show.simdata`: Graphical display of data simulated by simdata

`MCMC`: Full Bayesian Markov Chain Monte Carlo inference of parameters in the spatial F-model

`PostProcessChain`: Post-procesing of MCMC output for maps of posterior probability of populations subdomains

`PlotTessellation`: Graphical display of inferred sub-domains

The following functions are very basic and are only intended to be an aid for those not familiar with R. Most probably you may want to use directly the output files of `MCMC` and `PostProcessChain` to print your own figures.

`PlotDrift`: Graphical display of drift factors along MCMC run

`PlotFreqA`: Graphical display of allele frequencies in the ancestral population along MCMC run

`PlotFreq`: Graphical display of allele frequencies in the present time population along MCMC run

`Plotnpop`: Graphical display of number of populations along MCMC run

Package:	Geneland
Type:	Package
License:	GPL

Author(s)

Arnaud Estoup, Gilles Guillot, Filipe Santos

http://www2.imm.dtu.dk/~gigu/Geneland/

See also

http://www2.imm.dtu.dk/~gigu/Geneland/

References

- G. Guillot, Estoup, A., Mortier, F. Cosson, J.F. A spatial statistical model for landscape genetics. Genetics, 170, 1261-1280, 2005.
- G. Guillot, Mortier, F., Estoup, A. Geneland : A program for landscape genetics. Molecular Ecology Notes, 5, 712-715, 2005.
- Gilles Guillot, Filipe Santos and Arnaud Estoup, Analysing georeferenced population genetics data with Geneland: a new algorithm to deal with null alleles and a friendly graphical user interface Bioinformatics 2008 24(11):1406-1407.
- G. Guillot. Inference of structure in subdivided populations at low levels of genetic differentiation. The correlated allele frequencies model revisited. Bioinformatics, 24:2222-2228, 2008
- G. Guillot and F. Santos A computer program to simulate multilocus genotype data with spatially auto-correlated allele frequencies. Molecular Ecology Resources, 2009
- G. Guillot, R. Leblois, A. Coulon, A. Frantz Statistical methods in spatial genetics, Molecular Ecology, 2009.

`coordinates` *Example of coordinate file*

Description

Example of coordinate file

Author(s)

G. Guillot

EstimateFreqNA *Estimate frequencies of null alleles.*

Description

Estimate frequencies of null alleles at each locus.

Usage

```
EstimateFreqNA(path.mcmc)
```

Arguments

path.mcmc	path to directory containing stored MCMC computations

Value

A vector of length nloc (the number of loci) whose entries are estimated frequencies of null alleles.

Author(s)

Gilles Guillot

References

Gilles Guillot, Filipe Santos and Arnaud Estoup, Analysing georeferenced population genetics data with Geneland: a new algorithm to deal with null alleles and a friendly graphical user interface Bioinformatics 2008 24(11):1406-1407.

FormatGenotypes *Formatting file of genotype data.*

Description

Takes genotype data as a matrix with one line per individual and two columns per locus, with alleles coded by integers (number of replications for micro-satellites data). Build a new matrix with alleles codes as consecutive integers. If a locus has 7 alleles they will be coded as 1,2,...7. Since version 1.0.1, this function does not have to be called by users. It is called through MCMC.

Usage

```
FormatGenotypes(genotypes,ploidy)
```

Arguments

genotypes	A matrix with one line per individual and two columns per locus, with alleles coded by integers
ploidy	1 or 2

Value

A list with elements:

genotypes	a matrix with one line per individual and one or two columns per locus with alleles coded by integers
allele.numbers	a vector giving the number of possible alleles per locus

Author(s)

Gilles Guillot

Fstat *Computes F statistics on a population genetics dataset given as R object*

Description

Computes F statistics according to Weir and Cockerham's estimators. Missing values are allowed and accounting for in the computation of Fst. The presence of missing values involves a downward bias in the computation of Fis. This function should not be used on haploid data.

Usage

```
Fstat(genotypes,npop,pop.mbrship,ploidy)
```

Arguments

genotypes	Diploid codominant genotype data. A matrix with one line per individual and 2 columns per locus
npop	total number of popluation present in the dataset
pop.mbrship	Vector of integers giving the population membership for each individual
ploidy	Integer: 1 or 2 (default is 2) under development. Do not use for haploid data.

Value

A list with components

Fis	A vector of estimations of within-population Fis
Fst	A matrix of estimations the pairwise population Fst. (Only the upper triangular part is returned).

Author(s)

Arnaud Estoup for original code in Turbo Pascal. Translation in Fortran and interface with R by Gilles Guillot

References

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

`Fstat.output` *Computes F statistics on the output of an inference by MCMC simulation*

Description

Computes F statistics according to Weir and Cockerham's estimators. Missing values are allowed but as of today, the NA code is only treated as an extra allele which might bias the result. This function should not be used on haploid data.

Usage

```
Fstat.output(coordinates,genotypes,ploidy,burnin,path.mcmc)
```

Arguments

`coordinates`	Matrix with one line per individual and two columns
`genotypes`	Genotypes of individuals. A matrix with one line per individual and 2 columns per locus
`burnin`	Integer: number of saved iterations to discard.
`ploidy`	Integer: 1 or 2 (default is 2). Do not use for haploid data.
`path.mcmc`	Path to output files directory

Value

A list with components

`Pairwise.Fis`	A matrix of real numbers estimating the pairwise Fis
`Pairwise.Fst`	A matrix of real numbers estimating the pairwise Fst

Author(s)

Arnaud Estoup for original code in Turbo Pascal.

Translation in Fortran and interface with R by Gilles Guillot

References

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

Geneland.GUI *Graphical interface for package Geneland*

Description

Launch a menu driven interface to package Geneland

Usage

```
Geneland.GUI(lib.loc)
```

Arguments

lib.loc	A character string giving the path to the directory where Geneland is installed

Author(s)

Filipe Santos

genotypes *Example of coordinate file*

Description

Example of coordinate file

gl2gp *Geneland to Genepop conversion*

Description

Takes the matrices of coordinates and genotypes in the Geneland format and writes it as an ascci file in the Genepop format

Usage

```
gl2gp(coordinates, genotypes, file)
```

Arguments

coordinates	Matrix of coordinates
genotypes	MAtrix of genotypes
file	Character string giving the path to the file where the data in Genepop format should be written

Value

An ascci file is written. This file may need to be process in order to convert eol character from unix style to dos style.

Author(s)

Gilles Guillot and Tanguy de l'Argentaye

HZ *Estimate parameters of the hybrid zone model by MCMC simulation*

Description

Estimate parameters of a hybrid zone model by MCMC simulation. The function does not currently accept more than one genotype matrix.

Usage

```
HZ(coordinates,
   geno.dip.codom=NULL,
   geno.dip.dom=NULL,
   geno.hap=NULL,
   dist.IC=NULL,
   allele.freq=NULL,
   ncluster=NULL,
   cluster.indiv=NULL,
   path.mcmc.noadm=NULL,
   a.init=NULL,
   b.init=NULL,
   c.init=1,
   a.max=10,
   b.max=NULL,
   c.max=1,
   estimate.a=TRUE,
   estimate.b=TRUE,
   estimate.c=FALSE,
   common.param=TRUE,
   nit,
   thinning,
   path.mcmc.adm=NULL)
```

Arguments

coordinates	Spatial coordinates of individuals. A matrix with 2
geno.dip.codom	Genotypes for diploid data with codominant markers. A matrix with one line per individual and two columns per locus.

`geno.dip.dom` Genotypes for diploid data with dominant markers. A matrix with one line per individual and one column per locus. Presence/absence of a band should be coded as 0/1 (0 for absence / 1 for presence). Dominant and codominant markers can be analyzed jointly by passing variables to arguments geno.dip.codom and geno.dip.dom. Haploid data and diploid dominant data can not be analyzed jointly in the current version.

`geno.hap` Genotypes of haploid data. A matrix with one line per individual and one column per locus. Dominant diploid data and haploid data can be analyzed jointly (e.g. to analyse microsatelite data or SNP data together with mtDNA. Haploid data and diploid dominant data can not be analyzed jointly in the current version.

`dist.IC` A matrix with `nindiv` lines and `ncluster` column. The parameter `ncluster` being the number of clusters, most often 2 when there is a single hybrid zone. If `dist.IC` is missing, the user has to provide instead the path to a directory storing results from a no-admixture MCMC run.

`allele.freq` An array with `ncluster` x `nloc` x `nalmax` If missing, it will be estimated from the output from the MCMC run to estimate clusters.

`ncluster` Number of clusters. If missing, the user has to provide instead the path to a directory storing results from a no-admixture MCMC run

`cluster.indiv` Cluster membership of individuals. A numeric vector with integer values (maximum values being the total number of clusters)

`path.mcmc.noadm`
Path to output files directory of the previous Geneland no-admixture run. It seems that the path should be given in the Unix style even under Windows (use \/ instead of \). This path *has to* end with a slash (\/) (e.g. path.mcmc="/home/me/Geneland-noadmixture/")

`a.init` A numerical value to use as fixed or initial value for the `a` parameter

`b.init` A numerical value to use as fixed or initial value for the `b` parameter

`c.init` A numerical value to use as fixed or initial value for the `c` parameter

`a.max` Maximum value allowed along MCMC simulation for parameter `a` (default is 1)

`b.max` Maximum value allowed along MCMC simulation for parameter `b` (default is a small fraction of the study area diameter)

`c.max` Maximum value allowed along MCMC simulation for parameter `c` (default is 1)

`estimate.a` Logical. If TRUE, parameter `a` is estimated, if FALSE it is left at the initial value.

`estimate.b` Logical. If TRUE, parameter `b` is estimated, if FALSE it is left at the initial value.

`estimate.c` Logical. If TRUE, parameter `c` is estimated, if FALSE it is left at the initial value.

`common.param` If TRUE, parameters `a`, `b` and `c` are common to all clusters. If FALSE, the program attemps to estimate cluster-specific values.

`nit` Number of MCMC iteration.

`thinning` Number of MCMC iterations between two writing steps (if `thinning=1`, all states are saved whereas if e.g. `thinning=10` only each 10 iteration is saved)

`path.mcmc.adm` Path to output files directory for the admixture model. It seems that the path should be given in the Unix style even under Windows (use \/ instead of \). This path *has to* end with a slash (\/) (e.g. path.mcmc="/home/me/Geneland-admixture/")

Value

No object is returned. All outputs are stored in ascii file located in the `path.mcmc` directory

Author(s)

G. Guillot

References

B. Guedj and G. Guillot, A Bayesian model for infering hybrid zones.

`MCMC` *Markov Chain Monte-Carlo inference of clusters from genotpype data*

Description

Markov Chain Monte-Carlo inference of clusters from genotpype data

Usage

```
          MCMC(
 ## input data
                coordinates=NULL, # spatial coordinates
                geno.dip.codom=NULL, # diploid codominant markers
                                     # one line per indiv.
                                     # two column per marker
                geno.dip.dom=NULL, # diploid dominant markers
                                   # one line per indiv.
                                   # one column per marker
                geno.hap=NULL, # haploid
                               # one line per indiv.
                               # one column per marker
                qtc, # quantitative continuous variables
                qtd, # quantitative discrete variables
                ql, # qualitative variables
                ## path to output directory
                path.mcmc,
                ## hyper-prior parameters
                rate.max,delta.coord=0,shape1=2,shape2=20,
                npopmin=1,npopinit,npopmax,
                ## dimensions
                nb.nuclei.max,
```

```
## mcmc computations options
nit,
thinning=1,
freq.model="Uncorrelated",
varnpop=TRUE,
spatial=TRUE,
jcf=TRUE,
filter.null.alleles=TRUE,
prop.update.cell=0.1,
## writing mcmc output files options
write.rate.Poisson.process=FALSE,
write.number.nuclei=TRUE,
write.number.pop=TRUE,
write.coord.nuclei=TRUE,
write.color.nuclei=TRUE,
write.freq=TRUE,
write.ancestral.freq=TRUE,
write.drifts=TRUE,
write.logposterior=TRUE,
write.loglikelihood=TRUE,
write.true.coord=TRUE,
write.size.pop=FALSE,
write.mean.quanti=TRUE,
write.sd.quanti=TRUE,
write.betaqtc=FALSE,
miss.loc=NULL)
```

Arguments

coordinates	Spatial coordinates of individuals. A matrix with 2 columns and one line per individual.
geno.dip.codom	Genotypes for diploid data with codominant markers. A matrix with one line per individual and two columns per locus. Note that the object has to be of type matrix not table. This can be forced by function as.matrix.
geno.dip.dom	Genotypes for diploid data with dominant markers. A matrix with one line per individual and one column per locus. Presence/absence of a band should be coded as 0/1 (0 for absence / 1 for presence). Dominant and codominant markers can be analyzed jointly by passing variables to arguments geno.dip.codom and geno.dip.dom. Haploid data and diploid dominant data can not be analyzed jointly in the current version. Note that the object has to be of type matrix not table. This can be forced by function as.matrix.
geno.hap	Genotypes of haploid data. A matrix with one line per individual and one column per locus. Dominant diploid data and haploid data can be analyzed jointly (e.g. to analyse microsatelite data or SNP data together with mtDNA. Haploid data and diploid dominant data can not be analyzed jointly in the current version. Note that the object has to be of type matrix not table. This can be forced by function as.matrix.

qtc	A matrix of continuous quantitative phenotypic variables. One line per individual and one column per phenotypic variable. Note that the object has to be of type matrix not table. This can be forced by function as.matrix.
qtd	A matrix of discrete quantitative phenotypic variables. NOT IMPLEMENTED YET
ql	A matrix of categorical phenotypic variables. NOT IMPLEMENTED YET
path.mcmc	Path to output files directory. It seems that the path should be given in the Unix style even under Windows (use \/ instead of \). This path *has to* end with a slash (\/) (e.g. path.mcmc="/home/me/Geneland-stuffs/")
rate.max	Maximum rate of Poisson process (real number >0). Setting rate.max equal to the number of individuals in the dataset has proved to be efficient in many cases.
delta.coord	Parameter prescribing the amount of uncertainty attached to spatial coordinates. If delta.coord=0 spatial coordinates are considered as true coordinates, if delta.coord>0 it is assumed that observed coordinates are true coordinates blurred by an additive noise uniform on a square of side delta.coord centered on 0.
shape1	First parameter in the Beta(shape1,shape2) prior distribution of the drift parameters in the Correlated model.
shape2	Second parameter in the Beta(shape1,shape2) prior distribution of the drift parameters in the Correlated model.
npopmin	Minimum number of populations (integer >=1)
npopinit	Initial number of populations (integer sucht that npopmin =< npopinit =< npopmax)
npopmax	Maximum number of populations (integer >= npopinit). There is no obvious rule to select npopmax, it should be set to a value larger than any value that you can reasonably expect for your data.
nb.nuclei.max	Integer: Maximum number of nuclei in the Poisson-Voronoi tessellation. A good guess consists in setting this value equal to 3*rate.max. Lower values can also be used in order to speed up computations. The relevance of the value set can be checked by inspection of the MCMC run. The number of tiles should not go too close to nb.nuclei.max. If it does, you should re-run your chain with a larger value for nb.nuclei.max. In case of use of the option SPATIAL=FALSE, nb.nuclei.max should be set equal to the number of individuals.
nit	Number of MCMC iterations
thinning	Number of MCMC iterations between two writing steps (if thinning=1, all states are saved whereas if e.g. thinning=10 only each 10 iteration is saved)
freq.model	Character: "Correlated" or "Uncorrelated" (model for frequencies). See also details in detail section of Geneland help page.
varnpop	Logical: if TRUE the number of class is treated as unknown and will vary along the MCMC inference, if FALSE it will be fixed to the initial value npopinit. varnpop = TRUE *should not* be used in conjunction with freq.model = "Correlated" as in this case it seems that large numbers of populations are not penalized enough and there is a serious risk of inferring spurious sub-populations.

`spatial` Logical: if TRUE the colored Poisson-Voronoi tessellation is used as a prior for the spatial organisation of populations. If FALSE, all clustering receive equal prior probability. In this case spatial information (i.e coordinates) are not used and the locations of the nuclei are initialized and kept fixed at the locations of individuals.

`jcf` Logical: if true update of c and f are performed jointly

`filter.null.alleles`
Logical: if TRUE, tries to filter out null alleles. An extra fictive null allele is created at each locus coding for all putative null allele. Its frequency is estimated and can be viewed with function `PlotFreq`. This option is available only with `freq.model="Uncorrelated"`.

`prop.update.cell`
Integer between 0 and 1. Proportion of cell updated. For debugging only.

`write.rate.Poisson.process`
Logical: if TRUE (default) write rate of Poisson process simulated by MCMC

`write.number.nuclei`
Logical: if TRUE (default) write number of nuclei simulated by MCMC

`write.number.pop`
Logical: if TRUE (default) write number of populations simulated by MCMC

`write.coord.nuclei`
Logical: if TRUE (default) write coordinates of nuclei simulated by MCMC

`write.color.nuclei`
Logical: if TRUE (default) write color of nuclei simulated by MCMC

`write.freq` Logical: if TRUE (default is FALSE) write allele frequencies simulated by MCMC

`write.ancestral.freq`
Logical: if TRUE (default is FALSE) write ancestral allele frequencies simulated by MCMC

`write.drifts` Logical: if TRUE (default is FALSE) write drifts simulated by MCMC

`write.logposterior`
Logical: if TRUE (default is FALSE) write logposterior simulated by MCMC

`write.loglikelihood`
Logical: if TRUE (default is FALSE) write loglikelihood simulated by MCMC

`write.true.coord`
Logical: if TRUE (default is FALSE) write true spatial coordinates simulated by MCMC

`write.size.pop` Logical: if TRUE (default is FALSE) write size of populations simulated by MCMC

`write.mean.quanti`
Logical: if TRUE (default is FALSE) write means of quantitatives variables in the various groups simulated by MCMC

`write.sd.quanti`
Logical: if TRUE (default is FALSE) write standard deviations of quantitatives variables in the various groups simulated by MCMC

`write.betaqtc` Logical: if TRUE (default is FALSE) write hyper-parameter beta of distribution of quantitatives variables simulated by MCMC

`miss.loc` A matrix with `nindiv` lines and `nloc` columns of 0 or 1. For each individual, at each locus it says if the locus is genuinely missing (no attempt to measure it). This info is used under the option `filterNA=TRUE` do decide how a double missing value should be treated (genuine missing data or double null allele).

Value

Successive states of all blocks of parameters are written in files contained in `path.mcmc` and named after the type of parameters they contain.

Storage format

All parameters processed by function MCMC are written in the directory specified by '`path.mcmc`' as follows:

- File '`population.numbers.txt`' contains values of the number of populations (`nit` lines, one line per iteration of the MCMC algorithm).
- File '`population.numbers.txt`' contains values of the number of populations (`nit` lines, one line per iteration of the MCMC algorithm).
- File '`nuclei.numbers.txt`' contains the number of points in the Poisson point process generating the Voronoi tessellation.
- File '`color.nuclei.txt`' contains vectors of integers of length `nb.nuclei.max` coding the class membership of each Voronoi tile. Vectors of class membership for successive states of the chain are concatenated in one column. Some entries of the vector containing clas membership for a current state may have missing values as the actual number of polygon may be smaller that the maximum number allowed `nb.nuclei.max`. This file has `nb.nuclei.max*chain/thinning` lines.
- File '`coord.nuclei.txt`' contains coordinates of points in the Poisson point process generating the Voronoi tessellation. It has `nb.nuclei.max*chain/thinning` lines and two columns (hor. and vert. coordinates).
- File '`drifts.txt`' contains the drift factors for each population, (one column per population).
- File '`ancestral.frequencies.txt`' contains allele frequencies in ancestral population. Each line contains all frequencies of the current state. The file has `nit` lines. In each line, values of allele frequencies are stored by increasing allele index and and locus index (allele index varying first).
- File '`frequencies.txt`'contains allele frequencies of present time populations. Column xx contains frequencies of population numer xx. In each column values of allele frequencies are stored by increasing allele index and and locus index (allele index varying first), and values of successive iterations are pasted. The file has `nallmax*nloc*nit/thinning` lines where `nallmax` is the maximum number of alleles over all loci.
- File '`Poisson.process.rate.txt`' contains rates of Poisson process.
- File '`hidden.coord.txt`' contains the coordinates of each individual as updated along the chain if those given as input are not considered as exact coordinates (which is specified by `delta.coord` to a non zero value).

- File 'log.likelihood.txt' contains log-likelihood of data for the current state of parameters of the Markov chain.
- File 'log.posterior.density.txt' contains log of posterior probability (up to marginal density of data) of the current state of parameters in the Markov chain.

Author(s)

Gilles Guillot

References

- G. Guillot, Estoup, A., Mortier, F. Cosson, J.F. A spatial statistical model for landscape genetics. Genetics, 170, 1261-1280, 2005.
- G. Guillot, Mortier, F., Estoup, A. Geneland : A program for landscape genetics. Molecular Ecology Notes, 5, 712-715, 2005.
- Gilles Guillot, Filipe Santos and Arnaud Estoup, Analysing georeferenced population genetics data with Geneland: a new algorithm to deal with null alleles and a friendly graphical user interface Bioinformatics 2008 24(11):1406-1407.
- G. Guillot. Inference of structure in subdivided populations at low levels of genetic differentiation. The correlated allele frequencies model revisited. Bioinformatics, 24:2222-2228, 2008
- G. Guillot and F. Santos A computer program to simulate multilocus genotype data with spatially auto-correlated allele frequencies. Molecular Ecology Resources, 2009
- G. Guillot, R. Leblois, A. Coulon, A. Frantz Statistical methods in spatial genetics, Molecular Ecology, 2009.
- B. Guedj and G. Guillot. Estimating the location and shape of hybrid zones. Molecular Ecology Resources, 11(6) 1119-1123, 2011
- G. Guillot, S. Renaud, R. Ledevin, J. Michaux and J. Claude. A Unifying Model for the Analysis of Phenotypic, Genetic and Geographic Data. Systematic Biology, to appear, 2012.

See Also

simFmodel

---

nullify *Simulates null alleles*

Description

Simulates null alleles

Usage

```
nullify(genotypes, nall.null = 1, nloc.null)
```

Arguments

genotypes	a matrix of genotypes as produced by simFmodel and simIBD
nall.null	number of null alleles on each locus
nloc.null	number of loci carrying null alleles

Details

For diploid data only

Value

A list with component:

genotypes	The new genotypes after alteration.

Author(s)

Gilles Guillot

PlotDrift *Plot the drift factors*

Description

Gives a plot of the trace of the drift factors along the MCMC run

Usage

```
PlotDrift(path.mcmc,printit=FALSE,file)
```

Arguments

path.mcmc	Path to output files directory
printit	Logical : if TRUE, figures are also printed
file	Character : Path to file where figures should be printed

Author(s)

Gilles Guillot

PlotFreq — *Plot the trace of allele frequencies in present time population and optionnaly print it.*

Description

Plot of allele frequencies in present time populations number ipop, for allele iall of locus iloc

Usage

```
PlotFreq(path.mcmc,ipop,iloc,iall,printit=FALSE,path)
```

Arguments

path.mcmc	Path to output files directory
ipop	Integer number : index of population
iloc	Integer number : index of locus
iall	Integer number : index of allele. If MCMC was launched with option filter.null.alleles=TRUE, an extra fictive allele standing for putative null alleles is created. It estimated frequency can be also plotted. If there is say, 5 alleles at a locus, the estimated frequency of null alleles can be seen invoking PlotFreq with iall=6.
printit	Logical : if TRUE, figures are also printed
path	Character : Path to directory where figures should be printed

Author(s)

Gilles Guillot

PlotFreqA — *Plots allele frequencies in ancestral population along the MCMC run and optionally prints it*

Description

Plot frequency of allele iall of locus iloc in ancestral population

Usage

```
PlotFreqA(path.mcmc,iloc,iall,printit=FALSE,path)
```

Arguments

path.mcmc	Path to output files directory
iloc	Integer number : index of locus
iall	Integer number : index of allele
printit	Logical : if TRUE, figures are also printed
path	Character : Path to directory where figures should be printed

Author(s)

Gilles Guillot

Plotnpop *Plot of number of populations along the MCMC run*

Description

Gives a plot of the number of populations along the MCMC run

Usage

```
Plotnpop(path.mcmc,burnin,printit=FALSE,file,format="pdf")
```

Arguments

path.mcmc	Path to output files directory
printit	Logical : if TRUE, figures are also printed
file	Character : Path to file where figures should be printed
format	format of the output file, should be either "ps" or "pdf"
burnin	An integer: number of saved iterations to discard for the representation of the histogram of the chain

Author(s)

Gilles Guillot

References

A spatial statistical model for landscape genetics, Guillot, Estoup, Mortier, Cosson, Genetics, 2005

Guillot, Mortier, Estoup, Geneland : A program for landscape genetics. Molecular Ecology Notes, 2005.

Plotntile *Plot of number of tiles along the MCMC run*

Description

Plot of number of tiles in the Poisson-Voronoi tessellation along the MCMC run

Usage

```
Plotntile(path.mcmc,burnin,printit=FALSE,file)
```

Arguments

path.mcmc	Path to output files directory
printit	Logical : if TRUE, figures are also printed
file	Character : Path to file where figures should be printed
burnin	An integer: number of saved iterations to discard for the representation of the histogram of the chain

Author(s)

Gilles Guillot

PlotTessDyn *Plot tessellation along an MCMC run.*

Description

Plot sequence of tessellations simulated along an MCMC run.

Usage

```
PlotTessDyn(coordinates,path.mcmc,nxgrid,nygrid)
```

Arguments

coordinates	coordinates
path.mcmc	Path to output files directory
nxgrid	nb of pixels in the x direction
nygrid	nb of pixels in the y direction

Author(s)

Gilles Guillot

PlotTessellation *Maps of posterior probability of membership*

Description

Plots maps of posterior probabilities of population membership for each population

Usage

```
PlotTessellation(coordinates, path.mcmc,printit,path)
```

Arguments

coordinates	Spatial coordinates of individuals. A matrix with 2 columns and one line per individual.
path.mcmc	Character : Path to output files directory
printit	Logical : if TRUE, figures are also printed
path	Character : Path to directory where figures should be printed

Author(s)

Gilles Guillot

See Also

PostProcessChain

PosteriorMode *Map of mode of posterior distribution of population membership*

Description

Plots a map giving the modal population as a color code for each pixel.

Usage

```
PosteriorMode(coordinates,path.mcmc,plotit=TRUE,format="pdf",new.dev=TRUE,
                          printit=FALSE,file,main.title)
```

Arguments

coordinates	Spatial coordinates of individuals. A matrix with 2 columns and one line per individual.
path.mcmc	Path to output files directory
plotit	Logical: if TRUE the map is plotted
printit	Logical : if TRUE, figures are also printed
format	"ps" or "pdf"
file	Character : Path to file where figures should be printed
main.title	Character : Title to appear on top of the graphic
new.dev	Logical. Open a new graphical device if TRUE and printit is FALSE

Author(s)

Gilles Guillot

References

G. Guillot, Estoup, A., Mortier, F. Cosson, J.F. A spatial statistical model for landscape genetics. Genetics, 170, 1261-1280, 2005.

G. Guillot, Mortier, F., Estoup, A. Geneland : A program for landscape genetics. Molecular Ecology Notes, 5, 712-715, 2005.

PostProcessChain *Computation for maps of posterior probability of population membership*

Description

Computes posterior probabilities of population membership for each pixel of the spatial domain.

Usage

```
PostProcessChain(coordinates,
path.mcmc,nxdom, nydom,burnin)
```

Arguments

coordinates	Spatial coordinates of individuals. A matrix with 2 columns and one line per individual.
path.mcmc	Path to output files directory
nxdom	Number of pixel for discretization of the spatial domain in the horizontal direction
nydom	Number of pixel for discretization of the spatial domain in the vertical direction

`burnin` Number of iterations of the chain to throw away. WARNING : this argument should be given the number of stored iterations (and not the number of computed iterations which differ if `thinning` !=1). If you have `nit`=100000 and `thinning`=100, then only 1000 iterations are stored. Then `burnin`=10 will throw away 10 stored iterations, namely 100*10 computed iterations.

Value

`Posterior probability of population membership for each pixel:`
They are written in an ascii file called '`proba.pop.membership.txt`'. Two first columns are coordinates of pixels then one column per population (thus `npopmax` values are computed for each pixel). Images in each column of '`proba.pop.membership.txt`' are stored starting from the bottom left pixel. First line of '`proba.pop.membership.txt`' = bottom left pixel , second line of '`proba.pop.membership.txt`' = upward neighboor of the previous pixel, etc...

`Posterior probability of population membership for each individual:`
They are written in a file named '`proba.pop.membership.indiv.txt`'. One line per individuals and `2+npopmax` columns per individual. Two first columns are spatial coordinates.

`Label of modal population for pixels and individuals:`
They are written in files named '`modal.pop.txt`' and '`modal.pop.indiv.txt`' respectively. See the example section of function `MCMC` to see how they can be added in a plot.

Author(s)

Gilles Guillot

References

G. Guillot, Estoup, A., Mortier, F. Cosson, J.F. A spatial statistical model for landscape genetics. Genetics, 170, 1261-1280, 2005.

G. Guillot, Mortier, F., Estoup, A. Geneland : A program for landscape genetics. Molecular Ecology Notes, 5, 712-715, 2005.

See Also

`PlotTessellation`

`show.estimate.hz` *Show estimate of parameters of the hybrid zone model by MCMC simulation*

Description

Show estimate of parameters of a hybrid zone model.

Usage

```
show.estimate.hz(coordinates,
   path.mcmc.adm,
   burnin,
   angle=0,
   plot.distruct=TRUE,
   plot.mcmc=TRUE)
```

Arguments

coordinates	Spatial coordinates of individuals. A matrix with 2
path.mcmc.adm	Path to output files directory for the admixture model. It seems that the path should be given in the Unix style even under Windows (use \/ instead of \). This path *has to* end with a slash (\/) (e.g. path.mcmc="/home/me/Geneland-admixture/")
burnin	burnin
angle	angle of the axis on which sampling sites are projected (in radian)
plot.distruct	Logical
plot.mcmc	Logical

Value

No object is returned. All outputs are stored in ascii files located in the path.mcmc directory

Author(s)

G. Guillot

References

B. Guedj and G. Guillot,

---

show.simdata *Graphical display of data simulated by simdata*

Description

Graphical display of data simulated by simdata

Usage

```
show.simdata(dataset,
plot.coord = FALSE,
file.plot.coord,
plot.tess = FALSE,
file.plot.tess,
plot.freq.grid = FALSE,
file.plot.freq.grid,
loc.grid = 1,
plot.freq.indiv = FALSE,
file.plot.freq.indiv,
loc.indiv=1,
zlim.freq=c(0,1),
plot.gen = FALSE,
file.plot.gen)
```

Arguments

dataset	An R object produced by function simdata
plot.coord	Logical
file.plot.coord	Character string. Path to the file where the graphic should be stored
plot.tess	Logical
file.plot.tess	Character string. Path to the file where the graphic should be stored
plot.freq.grid	Logical
file.plot.freq.grid	Character string. Path to the file where the graphic should be stored
loc.grid	Vector of integers giving indices of loci for which the map of frequencies is required
plot.freq.indiv	Logical
file.plot.freq.indiv	Character string. Path to the file where the graphic should be stored
loc.indiv	Vector of integers giving indices of loci for which the plot of frequencies at sites of individuals is required
plot.gen	Logical
file.plot.gen	Character string. Path to the file where the graphic should be stored
zlim.freq	A vector of two integers giving the limit of the values for the image of fields of frequencies

Value

No value returned

Author(s)

Gilles Guillot

References

Guillot G. and Estoup A.

`simdata` *Simulation of georeferenced genotypes under an IBD + barrier model*

Description

Simulates coordinates and genotypes for a `npop` populations. Each population is supposed to be under an Isolation by Distance model and different populations are supposed to be separated by impermeable barriers. The barriers are given by a Poisson-Voronoi tessellation.

Usage

```
simdata(nindiv,
        coord.indiv,
        coord.lim=c(0,1,0,1),
        npop,
        rate,
        number.nuclei,
        coord.nuclei,
        color.nuclei,
        allele.numbers,
        sim.gen=FALSE,
        IBD=TRUE,
        model="stable",
        alpha=1,
        beta=1,
        gamma=1.8,
        sim.quanti=FALSE,
        nquanti.var,
        mean.quanti,
        sd.quanti,
        seed.coord,
        seed.tess,
        seed.freq,
        give.tess.grid=FALSE,
        give.freq.grid=FALSE,
        npix,
        comp.Fst=FALSE,
        comp.Dsigma2=FALSE,
        comp.diff=FALSE,
        width,
```

```
plot.pairs.borders=FALSE)
```

Arguments

nindiv	Number of indivuals
coord.indiv	Coordinates of the individuals
coord.lim	Limits of the geographical domain. The domain is supposed to be rectangular and the limits are given as (abs min, abs max, ord min, ord max)
npop	Number of Populations
rate	Rate of the Poisson process governing the hidden tessellation
number.nuclei	Number of nuclei in the tessellation (if given, then rate is ignored)
coord.nuclei	Coordinates of the nuclei (the number of coordinates of the nuclei given here as a matrix has to comply with number.nuclei)
color.nuclei	Population membeship of the nuclei: a vector of integer of length number of nuclei whose values are between 1 and npop
sim.gen	Logical to say whether genetic data should be simulated
allele.numbers	A vector giving the number of alleles observed at each locus
IBD	Logical. If TRUE, then the allele frequencies are simulated according to an IBD model. If FALSE, panmixia is assumed.
model	Model of spatial covariance function used for the underlying Gaussian fields (see documentation of package RandomFields for details)
alpha	Parameter of the spatial Dirichlet vector field of frequencies (a positive real)
beta	Scale parameter of the spatial covariance function used for the underlying Gaussian fields. A positive real number (see documentation of package RandomFields for details)
gamma	Smoothing parameter of spatial covariance function used for the underlying Gaussian fields. (see documentation of package RandomFields for details)
sim.quanti	Logical to say whether quantitative data should be simulated
nquanti.var	Number of quantitative variables to be simulated
mean.quanti	Mean of the quantitative variables in the various groups. A matrix with npop lines and nvar.quant columns
sd.quanti	Standard deviation of the quantitative variables in the various groups. A matrix with npop lines and nvar.quant columns
seed.coord	Random seed to initialise the simulation of the coordinates (mostly for debugging)
seed.tess	Random seed to initialise the simulation of the tessellation (mostly for debugging)
seed.freq	Random seed to initialise the simulation of the frequencies (mostly for debugging)
give.freq.grid	Logical to tell whether frequencies on a grid are also returned

`give.tess.grid` Logical to tell whether population memberships of pixels on a grid are also returned

`npix` A vector of two integers telling how many horizontal and vertical pixel should contain the grid for the graphical representations

`comp.Fst` Logical to tell whether Fst, Fis and Fit should be computed

`comp.Dsigma2` Logical to tell whether IBD index Dsgma2 should be computed

`comp.diff` Logical to tell whether the local differentiation across the barriers should be computed

`width` Real number specifying the width around the barrier in the computation of its local differentiation

`plot.pairs.borders`
Logical to tell whether the pairs of individuals coming into the computation of the differentiation of the barriers should be plotted

Value

A list whose components can be seen using `summary`

Author(s)

Arnaud Estoup, Gilles Guillot, Filipe Santos

References

G. Guillot, F. Santos, A. Estoup. Inference in population genetics with Geneland: a sensitivity analysis to spatial sampling scheme, null alleles and isolation by distance. Submitted.

See Also

Function `show.simdata` to make graphical display of simulated data.

Examples

```
## Not run:
dataset <- simdata(nindiv=100,
             sim.gen=TRUE,
             number.nuclei=10,
             allele.numbers=rep(5,3),
             model="stable",
             IBD=TRUE,
             alpha=1,
             beta=1,
             gamma=1,
             npop=3,
             sim.gen=TRUE,
             give.tess.grid=TRUE,
             give.freq.grid=TRUE,
             npix=c(10,10),
             comp.Fst=TRUE,
```

```
              comp.Dsigma2=TRUE,
              comp.diff=TRUE,
              width=0.1,
              plot.pairs.borders=TRUE)

## End(Not run)
```

---

simFmodel *Simulation of multi-locus genetic data from the spatial F-model*

Description

Simulates multi-locus genotypes and spatial coordinates for individuals belonging to some spatially organised populations.

Usage

```
simFmodel(nindiv, coordinates, coord.lim, number.nuclei,
coord.nuclei, color.nuclei,nall, npop, freq.model="Uncorrelated",
drift,dominance="Codominant", plots = FALSE, ploth = FALSE)
```

Arguments

nindiv	Integer: Number of individuals
coordinates	Matrix (2 rows, nindiv columns) of spatial coordinates of individuals
coord.lim	Vector of limits of spatial domain to be considered (x min, x max, y min, y max)
number.nuclei	Integer: number of nuclei in the Voronoi tessellation
coord.nuclei	Coordinates of nuclei of Voronoi tessellation
color.nuclei	Population labels of the nuclei (vector of integers of size number.nuclei)
nall	Vector of integers giving number of alleles at each locus
npop	Number of populations
freq.model	model for frequencies:"Correlated" or "Uncorrelated"
drift	Vector (of size npop) of drift factors between 0 and 1 (only for the Correlated model)
dominance	A character string "Codominant" or "Dominant". If "Dominant" is chosen, the first allele is treated as a recessive allele and all heterozigous are converted into homozigous for the dominant allele. The presence of the "dominant" allele is coded as 1, the absence of the "dominant" allele is coded as 0.
plots	Logical: if TRUE, spatial coordinates are ploted
ploth	Logical: if TRUE, barplots for allele frequencies are ploted

Details

number.nuclei uniform i.i.d points are randomly spread on the rectangular domain. These points generates the so called Voronoi tessellation of the domain in number.nuclei polygonal sub-domains. Each polygon is given a color uniformly on {1, npop}. The union of polygons of the color k gives the domain of population k. Then nindiv uniform i.i.d points are randomly spread on the domain and stand for the locations of individuals. Allele frequencies in the ancestral population are sampled from independent Dirichlet D(1,...,1). Allele frequencies in the present time population are drawn from Dirichlet distrubution whose parameters depend on drift factors drift and allele frequencies in the ancestral population. Individual genotypes in each population are drawn from the allele frequencies of the corresponding population assuming Hardy-Weinberg equilibrium and linkage equilibrium.

Value

A list of variables involved in the simulation. The elements of this list are: coordinates, genotypes, allele.numbers, number.nuclei, coord.nuclei, color.nuclei, frequencies, ancestral.frequencies, drifts, index.nearest.nucleus

Author(s)

Gilles Guillot

Index