

Package ‘GeoGenetix’

July 21, 2014

Title Quantification of the effect of geographic versus environmental isolation on genetic differentiation

Author Filippo Botta, Gilles Guillot

Description

Quantification of the effect of geographic versus environmental isolation on genetic differentiation

Maintainer ORPHANED

URL <http://www2.imm.dtu.dk/~gigu/Tangle>

License GPL

Version 0.0.2

Date Fri Jun 27 17:23:53 2014

Depends RandomFields

NeedsCompilation yes

Repository CRAN

Date/Publication 2014-07-21 07:35:43

X-CRAN-Original-Maintainer Filippo Botta <filippo.botta@gmail.com>

X-CRAN-Comment Orphaned and repaired on 2014-07-21 as Fortran errors were ignored.

R topics documented:

GeoGenetix-package	2
D_E	2
D_G	2
gen	3
Reformat23	3
Reformat32	4
SimTangleData	4
TangleInference	5
TangleInferenceGauss	7

Index**9**

GeoGenetix-package *Inference and model selection for analysis of geographical genetic variation*

Description

MCMC inference and model selection by cross-validation for the analysis of geographical genetic variation

Details

Package: GeoGenetix
 Type: Package
 Version: 1.0
 Date: 2014-06-20
 License: What license is it under?

~~ An overview of how to use the package, including the most important ~~ functions ~~

Author(s)

Filippo Botta and Gilles Guillot Who wrote it
 Maintainer: Filippo Botta <filippo.botta@gmail.com>

References

Cf. citation(GeoGenetix)

D_E *Matrix of environmental distances*

Description

Matrix of environmental distances

D_G *Matrix of geographical distances*

Description

Matrix of geographical distances

gen	<i>An array of allele counts</i>
-----	----------------------------------

Description

An array of allele counts in the format suitable for the inference functions

Reformat23	<i>Reformat data</i>
------------	----------------------

Description

Reformat data from the 2-way table format (containing allelic states) to the 3-way table format (containing counts of the various allelic states)

Usage

Reformat23(x)

Arguments

x A 2-way table containing allelic states

Value

A 3-way table with dimensions (n,l,a) n: number of geographical locations, l: number of loci, a: max number of alleles.

Author(s)

Casper Eriksen

 Reformat32

Reformat data

Description

Reformat data from the 3-way table format (containing counts of the various allelic states) to the 2-way table format (containing allelic states)

Usage

Reformat32(g)

Arguments

g An array in the 3-way table format

Value

A matrix in the 2-way table format (containing allelic states)

Author(s)

Casper Eriksen

SimTangleData

Simulation of data

Description

Simulate data under the model assumed by the MCMC inference program

Usage

SimTangleData(mod, theta, nsite, nloc, nind, nalm, nalm, var.par, scale.par)

Arguments

mod	A character string equal to "G+E", "G", "E"
theta	A vector of length 5 containing values for (alpha,beta_G,beta_E,gamma,delta)
nsite	Number of geographical locations
nloc	Number of loci
nind	Haploid population size at each combination location x locus
nalm	Maximum number of alleles over the various loci
nalm	Minimum number of alleles over the various loci
var.par	Variance of the random field model for the environmental variable
scale.par	Scale parameter in the exponential covariance of the random field model for the environmental variable

Value

A list containing genotypes, geographical distance, environmental distances, model and parameters.

Author(s)

Filippo Botta

Examples

```
data(toydata,package='GeoGenetix')
```

TangleInference	<i>Inference and model selection for analysis of geographical genetic variation</i>
-----------------	---

Description

MCMC inference and model selection by cross-validation for the analysis of geographical genetic variation

Usage

```
TangleInference(gen, D_G, D_E, nit, thinning, theta.max, theta.init, run, ud, set)
```

Arguments

gen	An array with dimensions (n,l,a) n: number of geographical locations, l: number of loci, a: max number of alleles
D_G	A matrix of geographical distances
D_E	A matrix of environmental distances
nit	Number of iterations
thinning	Thinning of MCMC iterations
theta.max	Upper bounds for the parameters in theta
theta.init	Initial state of theta
run	A vector of 0/1 of length 3 stating which sub-model is investigated among G+E,G,E (in this order)
ud	A vector of length 5 stating which entries in theta=(alpha,beta_E,beta_G,gamma,delta) will be updated in the MCMC iterations.
set	The number of pairs (sites x locus) used as validation set

Value

A list with a component named mod.lik containing likelihoods on the validation set for the various models compared.

Author(s)

Filippo Botta

Examples

```

## Not run:
data(toydata,package='GeoGenetix')

#### Computing options
nit <- 10^2
run <- c(1,1,1)
thinning <- max(nit/10^3,1);
ud <- c(0,1,1,0,0)
theta.init <- c(1,2,1,1,0.01)
set <- dim(gen)[1]*dim(gen)[2]/10
theta.max <- c(10,10*max(D_G),10*max(D_E),1,0.01)

plot <- TRUE
trace <- FALSE

#### Call Tangle ####
output <- TangleInference(gen,D_G,D_E,
                          nit,thinning,
                          theta.max,
                          theta.init,
                          run,ud,set)

mod.lik <- output$mod.lik
tvt <- output$tvt

## plotting outputs
upd=matrix(nrow=sum(run), ncol=length(theta.init), data=1)
upd[2,3]=upd[3,2]=0

plot(as.vector(D_G),as.vector(cor(t(gen[, ,1]))),
     bg=colorRampPalette(c("blue", "red"))(dim(D_E)[1]^2)[order(order(as.vector(D_E)))],
     pch=21,
     xlab='Geographic distance',
     ylab='Empirical covariance genotypes')

kol=c(4,2,3)
xseq=seq(thinning,nit,thinning)
ylab=c(expression(paste(alpha)),
       expression(paste(beta[D])),
       expression(paste(beta[E])),
       expression(paste(gamma)),
       expression(paste(delta)))

par(mfrow=c(sum(run),length(theta.init)))
for (j in 1:sum(run))

```

```

{
  for(k in 1:length(theta.init))
  {
    if (sum(upd[,k]==1)>0)
    {
      if(upd[j, k]==1)
      {
        if(exists("theta"))
          ylim=c(min(tvt[k,,j],theta[k]),max(tvt[k,,j],theta[k])) else
          ylim=c(min(tvt[k,,j]),max(tvt[k,,j]))
        plot(0, type="n",xlab="",ylab="", xlim=c(0,nit), ylim=ylim)
        lines(xseq,tvt[k,,j],col=col[j],xlab="",ylab="")
        if(exists("theta")) abline(h=theta[k],lty=2)
        title(xlab="iterations")
        mtext(ylab[k], side=2, line=2.3,las=1)} else plot.new()
      }
    }
  }
}

## End(Not run)

```

TangleInferenceGauss *Inference in Tangle package under the Gaussian approximation*

Description

Computes the mean squared cross validation error using leave-one-out cross validation for each of the three models "e+g", "e" and "g". Prediction is performed using linear projection based on the normal assumption.

Usage

```
TangleInferenceGauss(data, e, g, ...)
```

Arguments

data	An array with dimensions (n,l,a) n: number of geographical locations, l: number of loci, a: max number of alleles
e	A vector of environmental variables with n entries (n: number of geographical locations)
g	A matrix of geographical coordinates with n rows and 2 columns (n: number of geographical locations)
...	

Value

A vector containing Mean Square Errors in predicting validation set genotypes for the three models

Author(s)

Casper Eriksen

Index

- *Topic **Covariance function**
 - SimTangleData, [4](#)
 - *Topic **Cross-validation**
 - TangleInferenceGauss, [7](#)
 - *Topic **Data format**
 - Reformat23, [3](#)
 - Reformat32, [4](#)
 - *Topic **Gaussian approximation**
 - TangleInferenceGauss, [7](#)
 - *Topic **MCMC inference**
 - TangleInference, [5](#)
 - *Topic **Model selection**
 - TangleInference, [5](#)
 - *Topic **Random field**
 - SimTangleData, [4](#)
 - *Topic **package**
 - GeoGenetix-package, [2](#)
- D_E, [2](#)
D_G, [2](#)
- gen, [3](#)
GeoGenetix (GeoGenetix-package), [2](#)
GeoGenetix-package, [2](#)
- Reformat23, [3](#)
Reformat32, [4](#)
- SimTangleData, [4](#)
- TangleInference, [5](#)
TangleInferenceGauss, [7](#)