Package: StrainRanking (via r-universe)

August 31, 2024

Type Package

Title Ranking of Pathogen Strains

Version 1.2

Date 2017-11-25

Author Soubeyrand, S., Tollenaere, C., Haon-Lasportes, E. and Laine, A.-L.

Maintainer Samuel Soubeyrand <samuel.soubeyrand@inra.fr>

Description Regression-based ranking of pathogen strains with respect to their contributions to natural epidemics, using demographic and genetic data sampled in the curse of the epidemics. This package also includes the GMCPIC test.

License GPL (>= 2.0) | file LICENSE

Depends R (>= 3.0.0), methods, stats, graphics

NeedsCompilation yes

Repository CRAN

Date/Publication 2017-11-29 23:45:52 UTC

Contents

`StrainRanking-package` *Ranking of Pathogen Strains*

Description

Regression-based ranking of pathogen strains with respect to their contributions to natural epidemics, using demographic and genetic data sampled in the curse of the epidemics. This package also includes the GMCPIC test.

Details

Package:	StrainRanking
Type:	Package
Version:	1.2
Date:	2017-11-25
License:	GPL (>=2.0)
Depends:	methods

To rank pathogen strains using the method of Soubeyrand et al. (2014), create a DG object (Demographic and Genetic data set) with one of the three construction functions (DGobj.rawdata, DGobj.simul.regression and DGobj.simul.mechanistic) and apply the ranking.strains function. Other construction functions returning a DG object might be written to extend the approach proposed by Soubeyrand et al. (2014).

Author(s)

Soubeyrand, S., Tollenaere, C., Haon-Lasportes, E. and Laine, A.-L.

Maintainer: `<samuel.soubeyrand@inra.fr>`

References

Soubeyrand S., Tollenaere C., Haon-Lasportes E. & Laine A.-L. (2014). Regression-based ranking of pathogen strains with respect to their contributions to natural epidemics. PLOS ONE 9(1): e86591.

Soubeyrand S, Garreta V, Monteil C, Suffert F, Goyeau H, Berder J, Moinard J, Fournier E, Tharreau D, Morris C, Sache I (2017). Testing differences between pathogen compositions with small samples and sparse data. Phytopathology 107: 1199-1208. http://doi.org/10.1094/PHYTO-02-17-0070-FI

DGobj-class *Class* "DGobj"

Description

Class of objects containing demographic and genetic data and used as input of the function ranking.strains for ranking pathogen strains.

Objects from the Class

Objects can be created by calls of the form `new("DGobj", ...)` and by calls of the constructors DGobj.rawdata, DGobj.simul.mechanistic and DGobj.simul.regression.

Slots

`demographic`**:** Object of class `"matrix"`. The first two columns give the coordinates of sites where demographic data are available. The third column gives the values of the demographic growth at these sites.

`genetic`**:** Object of class `"matrix"`. The first two columns give the coordinates of sites where genetic data are available. Each following column (3, 4, ...) gives the frequencies of a given strain at these sites.

Methods

[`signature(x = "DGobj")`: ...

[<- `signature(x = "DGobj")`: ...

names `signature(x = "DGobj")`: ...

show `signature(object = "DGobj")`: ...

summary `signature(object = "DGobj")`: ...

Author(s)

Soubeyrand, S., Tollenaere, C., Haon-Lasportes, E. and Laine, A.-L.

References

Soubeyrand S., Tollenaere C., Haon-Lasportes E. & Laine A.-L. (2014). Regression-based ranking of pathogen strains with respect to their contributions to natural epidemics. PLOS ONE 9(1): e86591.

See Also

DGobj.rawdata, DGobj.simul.mechanistic, DGobj.simul.regression, ranking.strains

Examples

```
showClass("DGobj")

## load powderymildew data
data(powderymildew)

## construct a DG object from raw data
DGdata=DGobj.rawdata(demographic.coord=powderymildew$demographic.coord,
 genetic.coord=powderymildew$genetic.coord,
 demographic.measures=powderymildew$demographic.measures,
 genetic.frequencies=powderymildew$genetic.frequencies)

## show
DGdata
## summary
summary(DGdata)
## show the demographic slot
DGdata["demographic"]
## show the genetic slot
DGdata["genetic"]
## modify the demographic slot
#DGdata["demographic"]=DGdata["demographic"][1:50,]
## names of slots
names(DGdata)
```

`DGobj.rawdata` *Construction of a DG object from raw data*

Description

Construction of a DG object from raw demographic and genetic data.

Usage

```
DGobj.rawdata(demographic.coord, demographic.measures, genetic.coord,
 genetic.frequencies)
```

Arguments

`demographic.coord`	[2-column matrix] Coordinates of sites where demographic measurements were made.
`demographic.measures`	[2-column matrix] Demographic measurements (e.g. pathogen intensity). The first column contains measurements at the first sampling time. The second column contains measurements at the second sampling time.
`genetic.coord`	[2-column matrix] Coordinates of sites where genetic samples were collected.
`genetic.frequencies`	[Matrix] with frequencies of genetic samples from all sampled strains. Each column corresponds to a given strain.

Value

An object from the DG class.

Note

Demographic measurements, say $Y_i(t_1)$ and $Y_i(t_2)$, made at sampling sites $i \in \{1, \ldots, I\}$ and at the first and second sampling times, respectively, are transformed into the values $Z_i = \log\left(\frac{1+Y_i(t_2)}{1+Y_i(t_1)}\right)$ characterizing the temporal growth of the epidemic in space. The growth variable Z_i is given in the thrid column of the demographic slot of the returned DG object.

Author(s)

Soubeyrand, S., Tollenaere, C., Haon-Lasportes, E. and Laine, A.-L.

References

Soubeyrand S., Tollenaere C., Haon-Lasportes E. & Laine A.-L. (2014). Regression-based ranking of pathogen strains with respect to their contributions to natural epidemics. PLOS ONE 9(1): e86591.

See Also

DGobj-class, DGobj.simul.mechanistic, DGobj.simul.regression

Examples

```
## load the powdery mildew data set
data(powderymildew)

## create a DG object from this data set
DGdata=DGobj.rawdata(demographic.coord=powderymildew$demographic.coord,
 genetic.coord=powderymildew$genetic.coord,
 demographic.measures=powderymildew$demographic.measures,
 genetic.frequencies=powderymildew$genetic.frequencies)

summary(DGdata)
```

DGobj.simul.mechanistic

Simulation of a DG object under a mechanistic model

Description

Simulation of a DG object under a mechanistic model generating a multi-strain epidemic with multiple introductions over a square grid.

Usage

```
DGobj.simul.mechanistic(sqrtn, size1, size2, theta, beta, M, delta,
 plots = FALSE)
```

Arguments

sqrtn	[Positive integer] Side size of the square grid over which the epidemic is simulated. The inter-node distance in the grid is one in the horizontal and vertical directions. The total number of grid nodes is sqrtn^2.
size1	[Positive integer] Maximum number of grid nodes where pathogen isolates are collected (sampling sites).
size2	[Positive integer] Maximum number of pathogen isolates sampled in each sampling site.
theta	[Vector of positive numerics] Fitness coefficients of the strains. The length of this vector determines the number of strains in the epidemic.
beta	[Vector of postive numerics of size 2] Immigration parameters. The first component is the expected number of immigration nodes for every strain. The second component is the expected number of pathogen units in each immigration node.
M	[Positive integer] Number of time steps of the epidemic.
delta	[Positive numeric] Dispersal parameter.
plots	[Logical] If TRUE, plots are produced. The plots show the curse of the epidemic for each strain and the proportion of each strain in space at the final time step.

Details

The effective number of sampling sites is the maximum of size1 and the number of sites occupied at the last time of the simulation.

In each sampling site, the effective number of sampled isolates is the maximum of size2 and the number of pathogen isolates in the site.

The immigration time T_s^{immigr} at which the sub-epidemic due to strain s is initiated is randomly drawn between 1 and M with probabilities $P(T_s^{immigr} = t) = (M-t)^2 / \sum_{k=1}^{M}(M-k)^2$.

The number of immigration nodes is drawn from the binomial distribution with size sqrtn^2 and with expectation given by the first component of beta. The immigration nodes are uniformly drawn in the grid.

At time T_s^{immigr}, the numbers of pathogen units of strain s at the immigration nodes are independently drawn under the Poisson distribution with mean equal to the second component of beta.

Value

An object from the DG class.

Note

Demographic measurements, say $Y_i(M-1)$ and $Y_i(M)$, made at the grid nodes and at times M-1 and M, are transformed into the values $Z_i = \log\left(\frac{1+Y_i(M-1)}{1+Y_i(M)}\right)$ characterizing the temporal growth of the epidemic in space at the end of the epidemic. The growth variable Z_i is given in the thrid column of the demographic slot of the returned DG object.

Author(s)

Soubeyrand, S., Tollenaere, C., Haon-Lasportes, E. and Laine, A.-L.

References

Soubeyrand S., Tollenaere C., Haon-Lasportes E. & Laine A.-L. (2014). Regression-based ranking of pathogen strains with respect to their contributions to natural epidemics. PLOS ONE 9(1): e86591.

See Also

DGobj-class, DGobj.rawdata, DGobj.simul.regression

Examples

```
## Simulation of a data set
DGmech=DGobj.simul.mechanistic(sqrtn=10, size1=30, size2=10, theta=c(1.5,2,3),
 beta=c(5,5), M=7, delta=0.2)
summary(DGmech)

## Simulation of a data set and plots of the sub-epidemics for the strains and their
## proportions in space at the final time step
DGmech=DGobj.simul.mechanistic(sqrtn=10, size1=30, size2=10, theta=c(1.5,2,3),
 beta=c(5,5), M=7, delta=0.2, plots=TRUE)
summary(DGmech)
```

DGobj.simul.regression

Simulation of a DG object under a regression model

Description

Simulation of a DG object under a regression model generating proportions of pathogen strains in each node of a square grid.

Usage

```
DGobj.simul.regression(sqrtn, size1, size2, theta, alpha.function, sigma,
 plots = FALSE)
```

Arguments

`sqrtn`	[Positive integer] Side size of the square grid over which the proportions are simulated. The inter-node distance in the grid is one in the horizontal and vertical directions. The total number of grid nodes is sqrtn^2.
`size1`	[Positive integer] Number of grid nodes where pathogen isolates are collected (sampling sites).
`size2`	[Positive integer] Number of pathogen isolates sampled in each sampling site.
`theta`	[Vector of numerics] Regression coefficients representing the fitness of the strains. The length of this vector determines the number of strains.
`alpha.function`	[Function] Function whose value is a matrix of positive numerics with number of columns equal to the number of strains and the number of rows is the number of grid nodes. Each row of the matrix provides the parameters of the Dirichlet distribution used to draw the proportions of strains at each node. The argument of the function is a 2-column matrix of coordinates.
`sigma`	[Postive numeric] Standard deviation of the white noise.
`plots`	[Logical] If TRUE, plots are produced. The plots show the proportion of each strain in space.

Value

An object from the DG class.

Note

The function `DGobj.simul.regression` generates a growth variable (third column of the demographic slot of the returned DG object) satisfying:

$$Z_i = \left(\sum_{s=1}^{S} p_i(s) \texttt{theta}[s] \right) + \eta_i,$$

for each demographic sampling site i. In this expression, $(p_i(1), ..., p_i(S))$ are the proportions of the strains at sampling site i, where S is the number of different strains. These proportions are drawn in Dirichlet distributions. `theta`$[s]$ denotes the s-th component of `theta`. η_i denotes a centered random normal variable (white noise) with standard deviation `sigma`.

Author(s)

Soubeyrand, S., Tollenaere, C., Haon-Lasportes, E. and Laine, A.-L.

References

Soubeyrand S., Tollenaere C., Haon-Lasportes E. & Laine A.-L. (2014). Regression-based ranking of pathogen strains with respect to their contributions to natural epidemics. PLOS ONE 9(1): e86591.

See Also

DGobj-class, DGobj.rawdata, DGobj.simul.mechanistic, generation.alpha.3strains

Examples

```
## Simulation of a data set
DGreg=DGobj.simul.regression(sqrtn=10, size1=30, size2=10, theta=c(1.5,2,3),
 alpha.function=generation.alpha.3strains, sigma=0.1)
summary(DGreg)

## Simulation of a data set and plots of the proportions in space the strains
DGreg=DGobj.simul.regression(sqrtn=10, size1=30, size2=10, theta=c(1.5,2,3),
 alpha.function=generation.alpha.3strains, sigma=0.1,plots=TRUE)
summary(DGreg)
```

`generation.alpha.3strains` *Generation of parameters for the simulations under the regression model*

Description

Generation of parameters of the Dirichlet distribution used to draw the proportions of three strains at each site given in a matrix of coordinates.

Usage

```
generation.alpha.3strains(x)
```

Arguments

x	[2-column matrix] Coordinates where Dirichlet parameters are drawn.

Value

Matrix of positive numerics with three columns corresponding to the number of strains that are considered and with number of rows equal to the number of sites given in x. Each row of the matrix provides the parameters of the Dirichlet distribution used to draw the proportions of three strains at each site given in x.

Note

At each site $(x_{1,i}, x_{2,i})$ of x, the proportions of the three strains are defined by:

$$(p_i(1), p_i(2), p_i(3)) \sim Dirichlet[100\{\cos(x_{2,i}) + 1.5, \sin(x_{1,i}) + 1.5, \sin(x_{2,i}) + 1.5\}].$$

Author(s)

Soubeyrand, S., Tollenaere, C., Haon-Lasportes, E. and Laine, A.-L.

References

Soubeyrand S., Tollenaere C., Haon-Lasportes E. & Laine A.-L. (2014). Regression-based ranking of pathogen strains with respect to their contributions to natural epidemics. PLOS ONE 9(1): e86591.

See Also

DGobj.simul.regression

Examples

```
generation.alpha.3strains(expand.grid(1:10,1:10))
```

`gmcpic.test` *Function implementing the Generalized Monte Carlo plug-in test with calibration (GMCPIC test)*

Description

The GMCPIC test is a procedure to test the equality of the vectors of probabilities of two multinomial draws. The test statistics that is used is the multinomial-density statistic.

Usage

```
gmcpic.test(x, B, M, weights, threshold)
```

Arguments

`x`	[2-column matrix] Column 1 (resp. 2) contains the vector of observed frequencies in population 1 (resp. 2).
`B`	[Integer] Number of Monte Carlo simulations.
`M`	[Integer] Number of repetitions for the calibration.
`weights`	[Numeric] Vector of weights in [0,1] that are tried for the calibration.
`threshold`	[Numeric] Targeted risk level of the test; value in [0,1].

Details

The GMCPIC test was developed to test the similarity of two pathogen compositions based on small samples and sparse data.

Value

list with INPUT arguments (x, B, M, weights and threshold) and the following items:

calibrated.weight
: Weight selected by the calibration procedure.

p.value
: Test p-value.

reject.null.hypothesis
: Logical indicating whether the null hypothesis is rejected or not at the risk level specified by threshold.

Message
: Details about the p-value interpretation.

Author(s)

Samuel Soubeyrand <samuel.soubeyrand@inra.fr>

Vincent Garreta

Maintainer: Jean-Francois Rey

References

Soubeyrand S, Garreta V, Monteil C, Suffert F, Goyeau H, Berder J, Moinard J, Fournier E, Tharreau D, Morris C, Sache I (2017). Testing differences between pathogen compositions with small samples and sparse data. Phytopathology 107: 1199-1208. http://doi.org/10.1094/PHYTO-02-17-0070-FI

Examples

```
## Load Pathogen Compositions of M. oryzae collected in Madagascar
data(PathogenCompositionMoryzaeMadagascar)
x=t(PathogenCompositionMoryzaeMadagascar)

## Apply the GMCPIC test (use B=10^3, M=10^4 to get a robust result)
testMada=gmcpic.test(x, B=10^2, M=10^3, weights=seq(0.5,0.99,by=0.01),threshold=0.05)
testMada

## Apply the Chi-squared test
chisq.test(x, simulate.p.value = TRUE, B = 10000)
```

PathogenCompositionMoryzaeChina

Compositions of Magnaporthe oryzae collected in China

Description

Compositions of Magnaporthe oryzae formed from samples collected in Youle, Yunnan Province, China, in August 2008 and September 2009 (Saleh et al., 2014).

Usage

```
data(PathogenCompositionMoryzaeChina)
```

Format

A data frame with two rows, each row providing the pathogen composition (PC) at a given date (1st row: PC collected in August 2008; 2nd row: PC collected in September 2008).

References

Saleh D, Milazzo J, Adreit H, Fournier E, Tharreau D (2014). South-East Asia is the center of origin, diversity and dispersion of the rice blast fungus, Magnaporthe oryzae. New Phytologist 201: 1440-1456.

Soubeyrand S, Garreta V, Monteil C, Suffert F, Goyeau H, Berder J, Moinard J, Fournier E, Tharreau D, Morris C, Sache I (2017). Testing differences between pathogen compositions with small samples and sparse data. Phytopathology 107: 1199-1208. http://doi.org/10.1094/PHYTO-02-17-0070-FI

See Also

PathogenCompositionMoryzaeMadagascar

Examples

```
## Load Pathogen Compositions of M. oryzae collected in China
data(PathogenCompositionMoryzaeChina)

## Size of the first sample
sum(PathogenCompositionMoryzaeChina[1,])

## Size of the second sample
sum(PathogenCompositionMoryzaeChina[2,])

## Total number of different variants
ncol(PathogenCompositionMoryzaeChina)

## Display pathogen compositions
x=PathogenCompositionMoryzaeChina
barplot(t(x), col=rainbow(ncol(x)), main="M. oryzae - China")
```

PathogenCompositionMoryzaeMadagascar

Compositions of Magnaporthe oryzae collected in Madagascar

Description

Compositions of Magnaporthe oryzae formed from samples collected in Andranomanelatra, Madagascar, in February and April 2005 (Saleh et al., 2014).

Usage

```
data(PathogenCompositionMoryzaeMadagascar)
```

Format

A data frame with two rows, each row providing the pathogen composition (PC) at a given date (1st row: PC collected in February 2005; 2nd row: PC collected in April 2005).

References

Saleh D, Milazzo J, Adreit H, Fournier E, Tharreau D (2014). South-East Asia is the center of origin, diversity and dispersion of the rice blast fungus, Magnaporthe oryzae. New Phytologist 201: 1440-1456.

Soubeyrand S, Garreta V, Monteil C, Suffert F, Goyeau H, Berder J, Moinard J, Fournier E, Tharreau D, Morris C, Sache I (2017). Testing differences between pathogen compositions with small samples and sparse data. Phytopathology 107: 1199-1208. http://doi.org/10.1094/PHYTO-02-17-0070-FI

See Also

PathogenCompositionMoryzaeChina

Examples

```
## Load Pathogen Compositions of M. oryzae collected in Madagascar
data(PathogenCompositionMoryzaeMadagascar)

## Size of the first sample
sum(PathogenCompositionMoryzaeMadagascar[1,])

## Size of the second sample
sum(PathogenCompositionMoryzaeMadagascar[2,])

## Total number of different variants
ncol(PathogenCompositionMoryzaeMadagascar)

## Display pathogen compositions
x=PathogenCompositionMoryzaeMadagascar
barplot(t(x), col=rainbow(ncol(x)), main="M. oryzae - Madagascar")
```

PathogenCompositionPsyringaeClades

Compositions of Pseudomonas syringae at the clade resolution

Description

Compositions of Pseudomonas syringae formed from samples collected in South-East France, in Lower Durance River valley and in Upper Durance River valley (Monteil et al., 2014).

Usage

```
data(PathogenCompositionPsyringaeClades)
```

Format

A data frame with two rows, each row providing the pathogen composition (PC) at a given date (1st row: PC collected in Lower Durance River valley; 2nd row: PC collected in Upper Durance River valley).

References

Monteil C L, Lafolie F, Laurent J, Clement J C, Simler R, Travi Y, Morris C E (2014). Soil water flow is a source of the plant pathogen Pseudomonas syringae in subalpine headwaters. Environ. Microbiol. 16: 203862052.

Soubeyrand S, Garreta V, Monteil C, Suffert F, Goyeau H, Berder J, Moinard J, Fournier E, Tharreau D, Morris C, Sache I (2017). Testing differences between pathogen compositions with small samples and sparse data. Phytopathology 107: 1199-1208. http://doi.org/10.1094/PHYTO-02-17-0070-FI

See Also

PathogenCompositionPsyringaeHaplotypes, PathogenCompositionPsyringaePhylogroups

Examples

```
## Load Pathogen Compositions of P. syringae at the clade resolution
data(PathogenCompositionPsyringaeClades)

## Size of the first sample
sum(PathogenCompositionPsyringaeClades[1,])

## Size of the second sample
sum(PathogenCompositionPsyringaeClades[2,])

## Total number of different variants
ncol(PathogenCompositionPsyringaeClades)

## Display pathogen compositions
x=PathogenCompositionPsyringaeClades
barplot(t(x), col=rainbow(ncol(x)), main="P. syringae - Clades")
```

PathogenCompositionPsyringaeHaplotypes

Compositions of Pseudomonas syringae at the haplotype resolution

---

Description

Compositions of Pseudomonas syringae formed from samples collected in South-East France, in Lower Durance River valley and in Upper Durance River valley (Monteil et al., 2014).

Usage

```
data(PathogenCompositionPsyringaeHaplotypes)
```

Format

A data frame with two rows, each row providing the pathogen composition (PC) at a given date (1st row: PC collected in Lower Durance River valley; 2nd row: PC collected in Upper Durance River valley).

References

Monteil C L, Lafolie F, Laurent J, Clement J C, Simler R, Travi Y, Morris C E (2014). Soil water flow is a source of the plant pathogen Pseudomonas syringae in subalpine headwaters. Environ. Microbiol. 16: 203862052.

Soubeyrand S, Garreta V, Monteil C, Suffert F, Goyeau H, Berder J, Moinard J, Fournier E, Tharreau D, Morris C, Sache I (2017). Testing differences between pathogen compositions with small samples and sparse data. Phytopathology 107: 1199-1208. http://doi.org/10.1094/PHYTO-02-17-0070-FI

See Also

PathogenCompositionPsyringaeClades, PathogenCompositionPsyringaePhylogroups

Examples

```
## Load Pathogen Compositions of P. syringae at the haplotype resolution
data(PathogenCompositionPsyringaeHaplotypes)

## Size of the first sample
sum(PathogenCompositionPsyringaeHaplotypes[1,])

## Size of the second sample
sum(PathogenCompositionPsyringaeHaplotypes[2,])

## Total number of different variants
ncol(PathogenCompositionPsyringaeHaplotypes)

## Display pathogen compositions
x=PathogenCompositionPsyringaeHaplotypes
barplot(t(x), col=rainbow(ncol(x)), main="P. syringae - Haplotypes")
```

PathogenCompositionPsyringaePhylogroups

Compositions of Pseudomonas syringae at the phylogroup resolution

---

Description

Compositions of Pseudomonas syringae formed from samples collected in South-East France, in Lower Durance River valley and in Upper Durance River valley (Monteil et al., 2014).

Usage

```
data(PathogenCompositionPsyringaePhylogroups)
```

Format

A data frame with two rows, each row providing the pathogen composition (PC) at a given date (1st row: PC collected in Lower Durance River valley; 2nd row: PC collected in Upper Durance River valley).

References

Monteil C L, Lafolie F, Laurent J, Clement J C, Simler R, Travi Y, Morris C E (2014). Soil water flow is a source of the plant pathogen Pseudomonas syringae in subalpine headwaters. Environ. Microbiol. 16: 203862052.

Soubeyrand S, Garreta V, Monteil C, Suffert F, Goyeau H, Berder J, Moinard J, Fournier E, Tharreau D, Morris C, Sache I (2017). Testing differences between pathogen compositions with small samples and sparse data. Phytopathology 107: 1199-1208. http://doi.org/10.1094/PHYTO-02-17-0070-FI

See Also

PathogenCompositionPsyringaeClades, PathogenCompositionPsyringaeHaplotypes

Examples

```
## Load Pathogen Compositions of P. syringae at the phylogroup resolution
data(PathogenCompositionPsyringaePhylogroups)

## Size of the first sample
sum(PathogenCompositionPsyringaePhylogroups[1,])

## Size of the second sample
sum(PathogenCompositionPsyringaePhylogroups[2,])

## Total number of different variants
ncol(PathogenCompositionPsyringaePhylogroups)

## Display pathogen compositions
x=PathogenCompositionPsyringaePhylogroups
barplot(t(x), col=rainbow(ncol(x)), main="P. syringae - Phylogroups")
```

PathogenCompositionPtriticinaGalibier

Compositions of Puccinia triticina in Galibier crops

Description

Compositions of Puccinia triticina formed from samples collected in Lomagne, South-West France, from 2007 to 2013 (Soubeyrand et al., 2017).

Usage

```
data(PathogenCompositionPtriticinaGalibier)
```

Format

A data frame with 28 rows, each row providing the pathogen composition (PC) at a given date in years 2007-2013. The dates are provided in Soubeyrand et al. (2017).

References

Soubeyrand S, Garreta V, Monteil C, Suffert F, Goyeau H, Berder J, Moinard J, Fournier E, Tharreau D, Morris C, Sache I (2017). Testing differences between pathogen compositions with small samples and sparse data. Phytopathology 107: 1199-1208. http://doi.org/10.1094/PHYTO-02-17-0070-FI

See Also

PathogenCompositionPtriticinaKalango

Examples

```
## Load Pathogen Compositions of P. triticina in Galibier crops
data(PathogenCompositionPtriticinaGalibier)

## Size of the first sample
sum(PathogenCompositionPtriticinaGalibier[1,])

## Total number of different variants
ncol(PathogenCompositionPtriticinaGalibier)

## Display pathogen compositions
x=PathogenCompositionPtriticinaGalibier
barplot(t(x), col=rainbow(ncol(x)), las=2, main="P. triticina - Galibier")
```

PathogenCompositionPtriticinaKalango

Compositions of Puccinia triticina in Kalango crops

Description

Compositions of Puccinia triticina formed from samples collected in Lomagne, South-West France, from 2007 to 2013 (Soubeyrand et al., 2017).

Usage

```
data(PathogenCompositionPtriticinaKalango)
```

Format

A data frame with 28 rows, each row providing the pathogen composition (PC) at a given date in years 2007-2013. The dates are provided in Soubeyrand et al. (2017).

References

Soubeyrand S, Garreta V, Monteil C, Suffert F, Goyeau H, Berder J, Moinard J, Fournier E, Tharreau D, Morris C, Sache I (2017). Testing differences between pathogen compositions with small samples and sparse data. Phytopathology 107: 1199-1208. http://doi.org/10.1094/PHYTO-02-17-0070-FI

See Also

PathogenCompositionPtriticinaGalibier

Examples

```
## Load Pathogen Compositions of P. triticina in Kalango crops
data(PathogenCompositionPtriticinaKalango)

## Size of the first sample
sum(PathogenCompositionPtriticinaKalango[1,])

## Total number of different variants
ncol(PathogenCompositionPtriticinaKalango)

## Display pathogen compositions
x=PathogenCompositionPtriticinaKalango
barplot(t(x), col=rainbow(ncol(x)), las=2, main="P. triticina - Kalango")
```

---

powderymildew *Demographic and genetic real data*

Description

Demographic and genetic data collected during an epidemic of powdery mildew of Plantago lanceolata.

Usage

```
data(powderymildew)
```

Format

The format is: List of 4 components

`$demographic.coord` 'data.frame': 216 obs. of 2 variables (coordinates of the 216 sites with demographic data).

`$genetic.coord` 'data.frame': 22 obs. of 2 variables (coordinates of the 22 sites with genetic data).

`$demographic.measures` num [1:216, 1:2] Pathogen demographic measurements at week 32 and week 34 for sites whose coordinates are given in `$demographic.coord`.

`$genetic.frequencies` num [1:22, 1:5] Frequencies of strains 1 to 5 for sites whose coordinates are given in `$genetic.coord`.

See the examples section to visualize the data set.

References

Soubeyrand S., Tollenaere C., Haon-Lasportes E. & Laine A.-L. (2014). Regression-based ranking of pathogen strains with respect to their contributions to natural epidemics. PLOS ONE 9(1): e86591.

See Also

DGobj-class, DGobj.rawdata

Examples

```
## load the powderymildew data set
data(powderymildew)

## names of items of powderymildew
names(powderymildew)

## print powderymildew
print(powderymildew)

## alternatives to print one of the items of powderymildew, e.g. the 4th items:
print(powderymildew$genetic.frequencies)
print(powderymildew[[4]])
```

---

ranking.strains *Method for ranking pathogen strains*

Description

Ranking pathogen strains based on demographic and genetic data collected during an epidemic.

Usage

```
ranking.strains(DGobject, bw, nb.mcsimul, plots = FALSE, kernel.type = "Quadratic")
```

Arguments

`DGobject`	Object of the DG class.
`bw`	[Positive numeric] Smoothing bandwidth of the kernel used to estimate strain proportions.

`nb.mcsimul` [Positive integer] Number of permutations to assess the significance of the ranking.

`plots` [Logical] If TRUE, plots are produced. The plots show the growth variable in space, the sampling sites, the estimated values of the fitness coefficients and the corresponding permutation-based distributions obtained under the null hypothesis of coefficient equality.

`kernel.type` [Character string] Type of kernel. Default: Quadratic kernel $K(u) = (1 - u^2)I(0 \leq u \leq 1)$, where I is the indicator function. Other possible kernel types: Linear $K(u) = (1 - u)I(0 \leq u \leq 1)$, Power3 $K(u) = (1 - u^3)I(0 \leq u \leq 1)$, and Power4 $K(u) = (1 - u^4)I(0 \leq u \leq 1)$.

Value

`permutation.estimates` Estimates of the fitness coefficients obtained for the permutations (one row for each permutation).

`estimates` Estimates of the fitness coefficients obtained for the raw data.

`p.values` p.values of pairwise permutation tests of equality of the coefficients.

Author(s)

Soubeyrand, S., Tollenaere, C., Haon-Lasportes, E. and Laine, A.-L.

References

Soubeyrand S., Tollenaere C., Haon-Lasportes E. & Laine A.-L. (2014). Regression-based ranking of pathogen strains with respect to their contributions to natural epidemics. PLOS ONE 9(1): e86591.

See Also

DGobj-class, DGobj.rawdata, DGobj.simul.mechanistic, DGobj.simul.regression

Examples

```
## Application of the ranking method to a real data set
data(powderymildew)
DGdata=DGobj.rawdata(demographic.coord=powderymildew$demographic.coord,
 genetic.coord=powderymildew$genetic.coord,
 demographic.measures=powderymildew$demographic.measures,
 genetic.frequencies=powderymildew$genetic.frequencies)
ranking.strains(DGobject=DGdata, bw=sqrt(2), nb.mcsimul=10^3, plots=TRUE,
 kernel.type="Power4")

## Application of the ranking method to a data set simulated under the
## mechanistic model
DGmech=DGobj.simul.mechanistic(sqrtn=10, size1=30, size2=10, theta=c(1.5,2,3),
 beta=c(5,5), M=7, delta=0.2)
ranking.strains(DGobject=DGmech, bw=sqrt(2), nb.mcsimul=10^3, plots=TRUE,
 kernel.type="Power4")
```

```
## Application of the ranking method to a data set simulated under the
## regression model
DGreg=DGobj.simul.regression(sqrtn=10, size1=30, size2=10, theta=c(1.5,2,3),
 alpha.function=generation.alpha.3strains, sigma=0.1)
ranking.strains(DGobject=DGreg, bw=sqrt(2), nb.mcsimul=10^3, plots=TRUE,
 kernel.type="Power4")
```

Index