

# Package ‘GMCPIC’

July 7, 2017

**Type** Package

**Title** Generalized Monte Carlo Plug-in Test with Calibration (GMCPIC Test)

**Version** 1.2

**Date** 2017-07-07

**Author** Samuel Soubeyrand <samuel.soubeyrand@inra.fr>, Vincent Garreta

**Maintainer** Samuel Soubeyrand <samuel.soubeyrand@inra.fr>, Jean-Francois Rey <jean-francois.rey@inra.fr>

**Description** GMCPIC is an R-package implementing a computer intensive procedure to test the equality of two unknown vectors of probabilities  $p_1$  and  $p_2$  based on two multinomial draws performed with these probabilities. The GMCPIC package was specifically developed to test differences between pathogen compositions with small samples and sparse data.

**License** GPL ( $\geq 2$ ) | file LICENSE

**Depends** R ( $\geq 3.0.0$ )

## R topics documented:

GMCPIC-package . . . . .	2
gmpic.test . . . . .	3
PathogenCompositionMoryzaeChina . . . . .	4
PathogenCompositionMoryzaeMadagascar . . . . .	5
PathogenCompositionPsyringaeClades . . . . .	6
PathogenCompositionPsyringaeHaplotypes . . . . .	7
PathogenCompositionPsyringaePhylogroups . . . . .	8
PathogenCompositionPtriticinaGalibier . . . . .	9
PathogenCompositionPtriticinaKalango . . . . .	10

<b>Index</b>	<b>12</b>
--------------	-----------

---

GMCPIC-package	<i>Package implementing the Generalized Monte Carlo plug-in test with calibration (GMCPIC test)</i>
----------------	---

---

## Description

Computer intensive procedure to test equality of the probability vectors of two multinomial draws. This procedure was specifically designed to handle small samples and sparse data.

## Details

Package: GMCPIC  
Type: Package  
Version: 1.2  
Date: 2017-07-07  
License: GPL (>=2)

The GMCPIC package implements the Generalized Monte Carlo plug-in test with calibration proposed in Soubeyrand et al. (2017). This procedure can be used to test the equality of the vectors of probabilities of two multinomial draws. This test was developed to test the similarity of two pathogen compositions based on small samples and sparse data. Data sets analyzed in Soubeyrand et al. (2015) are provided in this package.

## Author(s)

Samuel Soubeyrand <samuel.soubeyrand@inra.fr>

Vincent Garreta

Jean-Francois Rey <jean-francois.rey@inra.fr>

## 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*. <http://doi.org/10.1094/PHYTO-02-17-0070-FI>

## See Also

[gmpic.test](#), [stats:chisq.test](#)

## 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=gmpic.test(x, B=10^1, M=10^1, 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)
```

---

gmpic.test	<i>Function implementing the Generalized Monte Carlo plug-in test with calibration (GMCPIC test)</i>
------------	--

---

## 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

```
gmpic.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

**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*. <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*. <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*. <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*. <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*. <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*. <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")
```

---

PathogenCompositionPtritricinaGalibier

*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., 2015).

**Usage**

```
data(PathogenCompositionPtritricinaGalibier)
```

**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*. <http://doi.org/10.1094/PHYTO-02-17-0070-FI>

**See Also**

[PathogenCompositionPtritricinaKalango](#)

**Examples**

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

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

## Total number of different variants
ncol(PathogenCompositionPtritricinaGalibier)

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

---

PathogenCompositionPtritricinaKalango

*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., 2014).

**Usage**

```
data(PathogenCompositionPtritricinaKalango)
```

**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*. <http://doi.org/10.1094/PHYTO-02-17-0070-FI>

**See Also**

[PathogenCompositionPtritricinaGalibier](#)

**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")
```

# Index

## \*Topic **datasets**

PathogenCompositionMoryzaeChina, [4](#)  
 PathogenCompositionMoryzaeMadagascar, [5](#)  
 PathogenCompositionPsyringaeClades, [6](#)  
 PathogenCompositionPsyringaeHaplotypes, [7](#)  
 PathogenCompositionPsyringaePhylogroups, [8](#)  
 PathogenCompositionPtriticinaGalibier, [9](#)  
 PathogenCompositionPtriticinaKalango, [10](#)

## \*Topic **misc**

gmpic.test, [3](#)

## \*Topic **package**

GMCPIC-package, [2](#)

China

(PathogenCompositionMoryzaeChina), [4](#)

Clades

(PathogenCompositionPsyringaeClades), [6](#)

Galibier

(PathogenCompositionPtriticinaGalibier), [9](#)

GMCPIC (GMCPIC-package), [2](#)

GMCPIC-package, [2](#)

gmpic.test, [2](#), [3](#)

Haplotypes

(PathogenCompositionPsyringaeHaplotypes), [7](#)

Kalango

(PathogenCompositionPtriticinaKalango), [10](#)

Madagascar

(PathogenCompositionMoryzaeMadagascar), [5](#)

MoryzaeChina

(PathogenCompositionMoryzaeChina), [4](#)

MoryzaeMadagascar

(PathogenCompositionMoryzaeMadagascar), [5](#)

PathogenCompositionMoryzaeChina, [4](#), [6](#)

PathogenCompositionMoryzaeMadagascar, [5](#), [5](#)

PathogenCompositionPsyringaeClades, [6](#), [8](#), [9](#)

PathogenCompositionPsyringaeHaplotypes, [7](#), [7](#), [9](#)

PathogenCompositionPsyringaePhylogroups, [7](#), [8](#), [8](#)

PathogenCompositionPtriticinaGalibier, [9](#), [10](#)

PathogenCompositionPtriticinaKalango, [9](#), [10](#)

Phylogroups

(PathogenCompositionPsyringaePhylogroups), [8](#)

PsyringaeClades

(PathogenCompositionPsyringaeClades), [6](#)

PsyringaeHaplotypes

(PathogenCompositionPsyringaeHaplotypes), [7](#)

PsyringaePhylogroups

(PathogenCompositionPsyringaePhylogroups), [8](#)

PtriticinaGalibier

(PathogenCompositionPtriticinaGalibier), [9](#)

(PathogenCompositionPtriticinaKalango),

10

stats:chisq.test, 2