# Package: RTCC (via r-universe)

**Type** Package

**Title** Detecting Trait Clustering in Environmental Gradients

**Version** 0.1.1

**Author** Mateu Menendez-Serra, Vicente J. Ontiveros, Emilio O. Casamayor, David Alonso

**Maintainer** Mateu Menendez-Serra <mateu.menendez@ceab.csic.es>

**Description** The Randomized Trait Community Clustering method (Triado-Margarit et al., 2019, <doi:10.1038/s41396-019-0454-4>) is a statistical approach which allows to determine whether if an observed trait clustering pattern is related to an increasing environmental constrain. The method 1) determines whether exists or not a trait clustering on the sampled communities and 2) assess if the observed clustering signal is related or not to an increasing environmental constrain along an environmental gradient. Also, when the effect of the environmental gradient is not linear, allows to determine consistent thresholds on the community assembly based on trait-values.

**License** GPL-3

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 7.1.0

**Imports** matrixStats, vegan, Rcpp

**Suggests** testthat, knitr, rmarkdown

**LinkingTo** testthat, Rcpp

**NeedsCompilation** yes

**Depends** R (>= 3.5.0)

**Date/Publication** 2020-06-12 17:50:03 UTC

**Repository** https://mateumenendez.r-universe.dev

**RemoteUrl** https://github.com/cran/RTCC

**RemoteRef** HEAD

**RemoteSha** 028d47d02e698ac9e074f0fd70aa60252e8ae298

# Contents

**Index**                     **13**

---

group_information         *Genomic data linked to saline lagoons.*

---

### Description

A dataset containing genomic data of 544 genomes that matched 16s rRNA data from saline lagoons of the Monegros desert area.

### Usage

```
group_information
```

### Format

A data frame with 544 rows and 14 variables:

**genome** Genome IMG code

**Genome_Size** Genome size

**GC_perc** GC percentage

**Coding_base_perc** Conding base percentage

**CDS_perc** CDS percentage

**RNA_perc** RNA percentage

**rRNA_count** rRNA count

**Transporter_perc** Transporter proteins percentage

**Signal_peptide_perc** Signal peptide percentage

**Transmembrane_perc** Transmembrane proteins percentage

**Gene_Count** Gene count

**min_env** Minimum environmental value where the organism has been observed

**max_env** Minimum environmental value where the organism has been observed

**rel_abundance** Relative abundance of the organism on the metacommunity

...

**Source**

Triadó-Margarit, X., Capitán, J.A., Menéndez-Serra, M. et al. A Randomized Trait Community Clustering approach to unveil consistent environmental thresholds in community assembly. ISME J 13, 2681–2689 (2019).

---

metadata                    *Salinity values of saline lagoons.*

---

**Description**

A dataset containing salinity values of 136 lagoons on the Monegros desert area.

**Usage**

```
metadata
```

**Format**

A data frame with 136 rows and 2 variables:

**sample_ID**  Sample internal code

**salinity**  Sample salinity value

**Source**

Triadó-Margarit, X., Capitán, J.A., Menéndez-Serra, M. et al. A Randomized Trait Community Clustering approach to unveil consistent environmental thresholds in community assembly. ISME J 13, 2681–2689.2019.

---

RTCC                    *RTCC: Detecting trait clustering in environmental gradients with the Randomized Trait Community Clustering method*

---

**Description**

A set of functions which allows to determine if the observed traits present clustering/overdispersion patterns on the observed samples, and if so, to stablish if the observed pattern is linked to the effect of an environmental gradient.

## Details

The study of phenotypic similarities and differences within species along environmental gradients might be used as a powerful tool complementing taxon-based approaches when assesing the contribution of stochastic and deterministic processes in community assembly. For this, this package allows an easy implementation of a method for detecting clustering/overdispersion patterns along an environmental gradient (Triado-Margarit et al., 2019). A first function assesses if the observed traits exhibit a clustering/overdispersion pattern on the tested samples. If positive, two subsequent functions determine whether the observed pattern is linked to the effect of an environmental varible and its statistical significance.

## Data entry

The data consists on presence-absence observations along a measured environmental gradient and trait quantitative information of the observed organisms.

## References

Triado-Margarit, X., Capitan, J.A., Menendez-Serra, M. et al. (2019) A Randomized Trait Community Clustering approach to unveil consistent environmental thresholds in community assembly. *ISME J* **13**, 2681–2689 . https://doi.org/10.1038/s41396-019-0454-4

---

rtcc1                               *Trait selection*

---

## Description

This function determines whether the selected traits exhibit or not a clustering/overdispersion signal on the tested samples. For each trait, compares the observed Mean Pairwise Distance (MPD) of each sample against a distribution of synthetic commmunities MPDs obtained by a randomization test. Each synthetic community is build maintaining the original sample richness and randomly selecting organisms form the global pool.

## Usage

```
rtcc1(table1, table2, table3, traits_columns, repetitions)
```

## Arguments

| | |
|---|---|
| table1 | A data frame containing organisms names on the first column and its trait values on the consecutive ones. It also has to contain two columns with the maximum and the minimum values of the tested environmental variable where the organisms have been observed. |
| table2 | A presence-absence observations table with the organisms names on the first column and the sample names as consecutive colnames. |
| table3 | A dataframe containing sample names on the first column and environmental parameters on the consecutive ones. |
| traits_columns | Table 1 column numbers where different trait values appear. |
| repetitions | Number of simulated synthetic communities distributions. |

## Value

The function returns a dataframe with trait names as colnames and the p-value distribution of the different traits.

## Examples

```
data(group_information)
data(table_presence_absence)
data(metadata)
rtcc1(group_information, table_presence_absence, metadata, 2:11, 100)
```

---

rtcc2 *Clustering signal along an environmental gradient*

---

## Description

For a given trait, this function determines whether the observed trait clustering/overdispersion on the metacommunity is linked to an environmental gradient. For this, it sequentially remove samples in decreasing order of the environmental variable and computes at each step the remaining metacommunity h-index. This index is based on the percentage of samples on a metacommunity presenting significant trait clustering/overdispersion.

## Usage

```
rtcc2(
  table1,
  table2,
  table3,
  species_abundances,
  trait_col_number,
  min_env_col,
  max_env_col,
  env_var_col,
  h_iteration,
  repetitions,
  model
)
```

## Arguments

table1          A data frame containing organisms names on the first column and its trait values on the consecutive ones. It also has to contain two columns with the maximum and the minimum values of the tested environmental variable where the organisms have been observed.

| | |
|---|---|
| table2 | A presence-absence observations table with the organisms names on the first column and the sample names as consecutive colnames. |
| table3 | A dataframe containing sample names on the first column and environmental parameters on the consecutive ones. |
| species_abundances | |
| | A vector containing the relative abundance of the organisms on the whole data set on the same order as appear on Table 1. |
| trait_col_number | |
| | Table 1 column number of the tested trait. |
| min_env_col | Table 1 column number indicating the minimum value of the environmental variable were each organism has been observed. |
| max_env_col | Table 1 column number indicating the maximum value of the environmental variable were each organism has been observed. |
| env_var_col | Table 2 column number indicating the tested environmental variable. |
| h_iteration | Number of h-index calculations for computing a confidence interval. |
| repetitions | Number of simulated synthetic communities distributions. |
| model | Model selection. All models build synthetic communities based on the organisms richness of the observed communities. |
| | - Model 1: organism are selected randomly from the global pool. - Model 2: organism are selected randomly with a probability based on its relative abundance on the global pool. - Model 3: organism are selected randomly, but only those whose environmental range includes the value of the simulated community are elegible. - Model 4: organism are selected randomly, but only those whose environmental range includes the value of the simulated community are elegible and the selection probability is based on its relative abundance on the global pool. |

### Value

The function returns a dataframe with the maximum of the environmental variable on the remaining metacommunity after the sequential removal, h-index calculation for each environmental value, and its confidence standard deviation.

### Examples

```
data(group_information)
data(table_presence_absence)
data(metadata)
rtcc2(group_information, table_presence_absence, metadata, group_information$sums,
9, 12, 13, 2, 100, 100, model = 1)
```

---

rtcc3 *Clustering signal significance.*

---

### Description

For a given trait and environmental variable, this function creates a null model of the clustering/overdispersion pattern in order to test if the observed pattern statistically differs from the expected by random. For this, it sequentially remove random samples from the metacommunity and computes at each step the remaining metacommunity h-index. This index is based on the percentage of samples on a metacoomunity presenting significant trait clustering/overdispersion. After h iterations, computes a 95 obtained h-index for each point of the environmental gradient.

### Usage

```
rtcc3(
  table1,
  table2,
  table3,
  species_abundances,
  trait_col_number,
  min_env_col,
  max_env_col,
  env_var_col,
  h_iteration,
  repetitions,
  model
)
```

### Arguments

| | |
|---|---|
| `table1` | A data frame containing organisms names on the first column and its trait values on the consecutive ones. It also has to contain two columns with the maximum and the minimum values of the tested environmental variable where the organisms have been observed. |
| `table2` | A presence-absence observations table with the organisms names on the first column and the sample names as consecutive colnames. |
| `table3` | A dataframe containing sample names on the first column and environmental parameters on the consecutive ones. |
| `species_abundances` | |
| | A vector containing the relative abundance of the organisms on the whole data set on the same order as appear on Table 1. |
| `trait_col_number` | |
| | Table 1 column number of the tested trait. |
| `min_env_col` | Table 1 column number indicating the minimum value of the environmental variable were each organism has been observed. |

| | |
|---|---|
| max_env_col | Table 1 column number indicating the maximum value of the environmental variable were each organism has been observed. |
| env_var_col | Table 2 column number indicating the tested environmental variable. |
| h_iteration | Number of h-index calculations for computing a confidence interval. |
| repetitions | Number of simulated synthetic communities distributions. |
| model | Model selection. All models build synthetic communities based on the organisms richness of the observed communities. |
| | - Model 1: organism are selected randomly from the global pool. - Model 2: organism are selected randomly with a probability based on its relative abundance on the global pool. - Model 3: organism are selected randomly, but only those whose environmental range includes the value of the simulated community are elegible. - Model 4: organism are selected randomly, but only those whose environmental range includes the value of the simulated community are elegible and the selection probability is based on its relative abundance on the global pool. |

**Value**

The function returns a dataframe with the maximum value of environmental variable corresponding to the same number of samples on the ordered remova, h-index calculation for each environmental value, and the percentiles 0.025, 0.5 and 0.975 of the obtained distribution for each point (mean value and 95

**Examples**

```
data(group_information)
data(table_presence_absence)
data(metadata)
rtcc3(group_information, table_presence_absence, metadata, group_information$sums,
9, 12, 13, 2, 50, 20, model = 1)
```

---

table_presence_absence

*Genome presence-absence data of 136 saline lagoons.*

---

**Description**

A dataset containing presence-absence data of 544 genomes on 136 saline lagoons of the Monegros desert area.

**Usage**

table_presence_absence

**Format**

A data frame with 544 rows and 137 variables:

**genome**  Genome IMG code

**MON_10**  Sample presence-absence observations

**MON_100**  Sample presence-absence observations

**MON_101**  Sample presence-absence observations

**MON_103**  Sample presence-absence observations

**MON_104**  Sample presence-absence observations

**MON_106**  Sample presence-absence observations

**MON_107**  Sample presence-absence observations

**MON_108**  Sample presence-absence observations

**MON_109**  Sample presence-absence observations

**MON_11**  Sample presence-absence observations

**MON_110**  Sample presence-absence observations

**MON_111**  Sample presence-absence observations

**MON_112**  Sample presence-absence observations

**MON_113**  Sample presence-absence observations

**MON_114**  Sample presence-absence observations

**MON_116**  Sample presence-absence observations

**MON_117**  Sample presence-absence observations

**MON_118**  Sample presence-absence observations

**MON_119**  Sample presence-absence observations

**MON_12**  Sample presence-absence observations

**MON_120**  Sample presence-absence observations

**MON_122**  Sample presence-absence observations

**MON_123**  Sample presence-absence observations

**MON_124**  Sample presence-absence observations

**MON_125**  Sample presence-absence observations

**MON_126**  Sample presence-absence observations

**MON_127**  Sample presence-absence observations

**MON_128**  Sample presence-absence observations

**MON_129**  Sample presence-absence observations

**MON_13**  Sample presence-absence observations

**MON_130**  Sample presence-absence observations

**MON_131**  Sample presence-absence observations

**MON_133**  Sample presence-absence observations

**MON_134**  Sample presence-absence observations

**MON_135** Sample presence-absence observations

**MON_136** Sample presence-absence observations

**MON_137** Sample presence-absence observations

**MON_138** Sample presence-absence observations

**MON_139** Sample presence-absence observations

**MON_14** Sample presence-absence observations

**MON_140** Sample presence-absence observations

**MON_141** Sample presence-absence observations

**MON_142** Sample presence-absence observations

**MON_144** Sample presence-absence observations

**MON_145** Sample presence-absence observations

**MON_146** Sample presence-absence observations

**MON_147** Sample presence-absence observations

**MON_148** Sample presence-absence observations

**MON_15** Sample presence-absence observations

**MON_17** Sample presence-absence observations

**MON_18** Sample presence-absence observations

**MON_19** Sample presence-absence observations

**MON_2** Sample presence-absence observations

**MON_20** Sample presence-absence observations

**MON_21** Sample presence-absence observations

**MON_22** Sample presence-absence observations

**MON_23** Sample presence-absence observations

**MON_24** Sample presence-absence observations

**MON_25** Sample presence-absence observations

**MON_26** Sample presence-absence observations

**MON_27** Sample presence-absence observations

**MON_28** Sample presence-absence observations

**MON_29** Sample presence-absence observations

**MON_30** Sample presence-absence observations

**MON_31** Sample presence-absence observations

**MON_32** Sample presence-absence observations

**MON_33** Sample presence-absence observations

**MON_34** Sample presence-absence observations

**MON_35** Sample presence-absence observations

**MON_36** Sample presence-absence observations

**MON_37** Sample presence-absence observations

**MON_38** Sample presence-absence observations

**MON_39** Sample presence-absence observations

**MON_4** Sample presence-absence observations

**MON_40** Sample presence-absence observations

**MON_41** Sample presence-absence observations

**MON_42** Sample presence-absence observations

**MON_43** Sample presence-absence observations

**MON_44** Sample presence-absence observations

**MON_45** Sample presence-absence observations

**MON_46** Sample presence-absence observations

**MON_47** Sample presence-absence observations

**MON_48** Sample presence-absence observations

**MON_49** Sample presence-absence observations

**MON_5** Sample presence-absence observations

**MON_50** Sample presence-absence observations

**MON_51** Sample presence-absence observations

**MON_52** Sample presence-absence observations

**MON_53** Sample presence-absence observations

**MON_54** Sample presence-absence observations

**MON_55** Sample presence-absence observations

**MON_56** Sample presence-absence observations

**MON_57** Sample presence-absence observations

**MON_58** Sample presence-absence observations

**MON_59** Sample presence-absence observations

**MON_60** Sample presence-absence observations

**MON_61** Sample presence-absence observations

**MON_62** Sample presence-absence observations

**MON_63** Sample presence-absence observations

**MON_64** Sample presence-absence observations

**MON_65** Sample presence-absence observations

**MON_66** Sample presence-absence observations

**MON_67** Sample presence-absence observations

**MON_68** Sample presence-absence observations

**MON_69** Sample presence-absence observations

**MON_7** Sample presence-absence observations

**MON_70** Sample presence-absence observations

**MON_71** Sample presence-absence observations

**MON_72** Sample presence-absence observations

**MON_73** Sample presence-absence observations

**MON_74** Sample presence-absence observations

**MON_75** Sample presence-absence observations

**MON_76** Sample presence-absence observations

**MON_77** Sample presence-absence observations

**MON_78** Sample presence-absence observations

**MON_79** Sample presence-absence observations

**MON_8** Sample presence-absence observations

**MON_80** Sample presence-absence observations

**MON_81** Sample presence-absence observations

**MON_82** Sample presence-absence observations

**MON_83** Sample presence-absence observations

**MON_84** Sample presence-absence observations

**MON_85** Sample presence-absence observations

**MON_86** Sample presence-absence observations

**MON_88** Sample presence-absence observations

**MON_9** Sample presence-absence observations

**MON_90** Sample presence-absence observations

**MON_91** Sample presence-absence observations

**MON_92** Sample presence-absence observations

**MON_93** Sample presence-absence observations

**MON_94** Sample presence-absence observations

**MON_95** Sample presence-absence observations

**MON_96** Sample presence-absence observations

**MON_97** Sample presence-absence observations

**MON_98** Sample presence-absence observations

**MON_99** Sample presence-absence observations ...

**Source**

Triadó-Margarit, X., Capitán, J.A., Menéndez-Serra, M. et al. A Randomized Trait Community Clustering approach to unveil consistent environmental thresholds in community assembly. ISME J 13, 2681–2689 (2019).

# Index