Package ‘RTCC’

March 31, 2020

Type Package

Title Detecting Trait Clustering in Environmental Gradients

Version 0.1.0

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)

Repository CRAN

Date/Publication 2020-03-31 10:00:13 UTC

R topics documented:

  group_information ................................................................. 2
  metadata .................................................................................. 3
  RTCC .......................................................................................... 3
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  Coding 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

**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**


---

**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 establish 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 assessing 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 variable 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


### rtcc1

<table>
<thead>
<tr>
<th>Trait selection</th>
</tr>
</thead>
</table>

### 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 communities 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

```r
rtcc1(
  table1,
  table2,
  table3,
  traits_columns,
  min_env_col,
  max_env_col,
  env_var_col,
  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.
**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**

```r
tgcc2(
  table1,
  table2,
  table3,
  species_abundances,
  trait_col_number,
  min_env_col,
  max_env_col,
  env_var_col,
  h_iteration,
)```

**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.

**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**

```r
data(group_information)
data(table_presence_absence)
data(metadata)
rtcc1(group_information, table_presence_absence, metadata, 2:11, 12, 12, 2, 100)```
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 eligible.
- Model 4: organism are selected randomly, but only those whose environmental range includes the value of the simulated community are eligible 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**
rtcc3

\begin{verbatim}
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)
\end{verbatim}

---

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 metacommunity presenting significant trait clustering/overdispersion. After \( h \) iterations, computes a 95 obtained h-index for each point of the environmental gradient.

**Usage**

\begin{verbatim}
rtcc3(
table1,
table2,
table3,
species_abundances,
trait_col_number,
min_env_col,
max_env_col,
env_var_col,
h_iteration,
repetitions,
model
)
\end{verbatim}

**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 eligible. - Model 4: organism are selected randomly, but only those whose environmental range includes the value of the simulated community are eligible 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)
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:

<table>
<thead>
<tr>
<th>genome</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MON_10</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_100</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_101</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_103</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_104</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_106</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_107</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_108</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_109</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_11</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_110</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_111</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_112</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_113</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_114</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_116</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_117</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_118</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_119</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_12</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_120</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_122</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_123</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>---------</td>
<td>--------------------------------------</td>
</tr>
<tr>
<td>MON_124</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_125</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_126</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_127</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_128</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_129</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_13</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_130</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_131</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_132</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_133</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_134</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_135</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_136</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_137</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_138</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_139</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_14</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_140</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_141</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_142</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_143</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_144</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_145</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_146</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_147</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_148</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_15</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_16</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_17</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_18</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_19</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_2</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_3</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_4</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_5</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_6</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_7</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_8</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_9</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_10</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_11</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_12</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_13</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_14</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_15</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_16</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_17</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_18</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_19</td>
<td>Sample presence-absence observations</td>
</tr>
<tr>
<td>MON_20</td>
<td>Sample presence-absence observations</td>
</tr>
</tbody>
</table>
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

Index

*Topic datasets
  group_information, 2
  metadata, 3
  table_presence_absence, 9

  group_information, 2
  metadata, 3
  RTCC, 3
  rtcc1, 4
  rtcc2, 5
  rtcc3, 7
  table_presence_absence, 9