

Package ‘quiddich’

March 25, 2019

Type Package

Title QUick IDentification of DIagnostic CHaracters

Version 1.0.0

Date 2019-03-21

Author A. Luise Kuehn

Maintainer A. Luise Kuehn <luise.kuehn@uni-greifswald.de>

Description Provides tools for an automated identification of diagnostic molecular characters, i.e. such columns in a given nucleotide or amino acid alignment that allow to distinguish taxa from each other. These characters can then be used to complement the formal descriptions of the taxa, which are often based on morphological and anatomical features. Especially for morphologically cryptic species, this will be helpful. QUIDDICH distinguishes between four different types of diagnostic characters. For more information, see “Kuehn, A.L., Haase, M. 2019. QUIDDICH: QUick IDentification of DIagnostic CHaracters.”

Depends ape(>= 5.2), R(>= 0.12.0)

Suggests spider, adegenet

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 6.1.1

NeedsCompilation no

Repository CRAN

Date/Publication 2019-03-25 09:30:03 UTC

R topics documented:

quiddich-package	2
changesAA	2
diagCharAA	4
diagCharNA	5

Index	8
--------------	----------

quiddich-package *QUick IDentification of DIagnostic CHaracters*

Description

The package QUIDDICH provides tools for an automated identification of diagnostic molecular characters, i.e. such columns in a given nucleotide or amino acid alignment that allow to distinguish taxa from each other. These characters can be used to complement formal descriptions of taxa, which are often based on morphological and anatomical features. Especially for morphologically cryptic species, this is helpful. QUIDDICH distinguishes between four different types of diagnostic characters. For more information, see "Kuehn, A.L., Haase, M. 2019. QUIDDICH: QUick IDentification of DIagnostic CHaracters."

Author(s)

A. Luise Kuehn <luise.kuehn@uni-greifswald.de>

changesAA *Identification of diagnostic molecular characters in a nucleotide alignment that also cause diagnostic characters in the corresponding amino acid alignment*

Description

This function is a tool for an automated identification of such diagnostic molecular characters in a nucleotide alignment that also cause diagnostic characters in the corresponding amino acid alignment. For each taxon given in `taxOfInt`, it identifies the mentioned characters and returns their positions and types in the nucleotide alignment and their positions and types in the corresponding amino acid alignment.

Usage

```
changesAA(DNAbin, taxVector = dimnames(DNAbin)[[1]], codonstart = 1,
  taxOfInt = "all", types = c("type1", "type2", "type3"))
```

Arguments

DNAbin	An object (the nucleotide alignment) of class 'DNAbin'.
taxVector	The taxon vector. Default assumes that each row in the alignment belongs to a different taxon.
codonstart	An integer defining where to start the translation of the nucleotide alignment into the corresp. amino acid alignment. Default is 1.
taxOfInt	A vector containing the taxa for which the above mentioned characters shall be extracted. Default is "all".
types	A vector containing the types of diagnostic characters (in the nucleotide alignment) that shall be considered. The types can be "all" or any combination of "type1", "type2", "type3" and "type4". Default is "type1", "type 2" and "type3".

Value

changesAA returns a list, where each entry belongs to one taxon of interest. Each taxon of interest has a matrix assigned to it, in which the first and second column contain the positions and types of the identified diagnostic characters in the nucleotide alignment and the third and fourth column contain the positions and types of the diagnostic characters in the corresponding amino acid alignment.

type1 means that the character is suitable to distinguish each individual of the taxon of interest from all individuals of the remaining taxa, and that it is fixed for one state in the taxon of interest. type2 means that the character is suitable to distinguish each individual of the taxon of interest from all individuals of the remaining taxa, and that it is not fixed for one state in the taxon of interest. type3 means that the character is suitable to distinguish some (but not all) individuals of the taxon of interest from all individuals of the remaining taxa. type4 means that the character is suitable to distinguish each individual of the taxon of interest from all individuals of at least one (but not all) other taxon while being fixed in both the taxon of interest and the compared taxa.

changesAA returns for each taxon of interest the following elements:

posDiNu	The positions of the identified diagnostic characters in the nucleotide alignment.
typeDiNu	The types of the identified diagnostic characters in the nucleotide alignment.
posDiAA	The positions of the corresponding diagnostic characters in the amino acid alignment.
typeDiAA	The types of the corresponding diagnostic characters in the amino acid alignment.

Author(s)

A. Luise Kuehn <luise.kuehn@uni-greifswald.de>

References

Kuehn, A.L., Haase, M. (2019) QUIDDICH: QUick IDentification of DIagnostic CHaracters.

Examples

```
#using a dataset from spider
#install.packages("spider")
library(spider)
data("anoteropsis")
anoTax <- sapply(strsplit(dimnames(anoteropsis)[[1]], split="_"),
  function(x) paste(x[1], x[2], sep="_"))
changesAA(anoteropsis, anoTax, codonstart=2, taxOfInt="all")
changesAA(anoteropsis, anoTax, codonstart=2, taxOfInt="Artoria_flavimanus",
  types="type4")
```

diagCharAA	<i>Identification of diagnostic molecular characters in amino acid alignments for the delineation of taxa</i>
------------	---

Description

This function is a tool for an automated identification of diagnostic molecular characters that allow to distinguish taxa within an amino acid alignment. For each taxon given in `taxOfInt`, it identifies the diagnostic characters and returns their alignment positions, their types, the states that are characteristic for the taxon of interest and (in case of type 4 characters) the taxon that it was compared with.

Usage

```
diagCharAA(AAbin, taxVector = dimnames(AAbin)[[1]], taxOfInt = "all",
           types = c("type1", "type2", "type3"), gapValid = TRUE)
```

Arguments

AAbin	An object (the amino acid alignment) of class 'AAbin'.
taxVector	The taxon vector. Default assumes that each row in the alignment belongs to a different taxon.
taxOfInt	A vector containing the taxa for which diagnostic molecular characters shall be extracted. Default is "all".
types	A vector containing the types of diagnostic molecular characters that shall be extracted. The types can be "all" or any combination of "type1", "type2", "type3" and "type4". Default is "type1", "type 2" and "type3".
gapValid	Boolean variable denoting if a gap can be a characteristic state for taxon <i>i</i> (and taxon <i>l</i> in case of type 4). Default is TRUE.

Value

`diagCharAA` returns a list, where each entry belongs to one taxon of interest. Each taxon of interest has a set of diagnostic molecular characters (position, type, characteristic states for taxon of interest, compared taxa) assigned to it. `type1` means that the character is suitable to distinguish each individual of the taxon of interest from all individuals of the remaining taxa, and that it is fixed for one state in the taxon of interest. `type2` means that the character is suitable to distinguish each individual of the taxon of interest from all individuals of the remaining taxa, and that it is not fixed for one state in the taxon of interest. `type3` means that the character is suitable to distinguish some (but not all) individuals of the taxon of interest from all individuals of the remaining taxa. `type4` means that the character is suitable to distinguish each individual of the taxon of interest from all individuals of at least one (but not all) other taxon while being fixed in both the taxon of interest and the compared taxa.

`diagCharAA` returns for each taxon of interest the following elements:

position	The positions of its diagnostic molecular characters.
----------	---

type	The types of the diagnostic molecular characters.
states	The states that are characteristic for the taxon <i>i</i> of interest, i.e. states that are distinct from "X" and unique to the taxon of interest (in case of type 1, 2 or 3), or fixed in the taxon of interest (type 4), resp.
compared taxa	Only relevant for type 4 characters. It contains the name <i>x</i> if the character is found to be a type 4 character of the taxon of interest when being compared to taxon <i>x</i> .

Author(s)

A. Luise Kuehn <luise.kuehn@uni-greifswald.de>

References

Kuehn, A.L., Haase, M. 2019. QUIDDICH: QUick IDentification of DIagnostic CHaracters.

Examples

```
#using a dataset from spider
#install.packages("spider")
#install.packages("ape")
library(spider)
library(ape)
data("anoteropsis")
anoTax <- sapply(strsplit(dimnames(anoteropsis)[[1]], split="_"),
  function(x) paste(x[1], x[2], sep="_"))
anoteropsis_AA <- trans(anoteropsis,code=1,codonstart=2)
diagCharAA(anoteropsis_AA, anoTax, taxOfInt="all")
diagCharAA(anoteropsis_AA, anoTax, taxOfInt="Artoria_flavimanus", types=c("type1","type2"))
```

diagCharNA	<i>Identification of diagnostic molecular characters in nucleotide alignments for the delineation of taxa</i>
------------	---

Description

This function is a tool for an automated identification of diagnostic molecular characters that allow to distinguish taxa within a nucleotide alignment. For each taxon given in `taxOfInt`, it identifies the diagnostic characters and returns their alignment positions, their types, the states that are characteristic for the taxon of interest and (in case of type 4 characters) the taxon that it was compared with.

Usage

```
diagCharNA(DNAbin, taxVector = dimnames(DNAbin)[[1]], taxOfInt = "all",
  types = c("type1", "type2", "type3"), gapValid = TRUE)
```

Arguments

DNABin	An object (the nucleotide alignment) of class 'DNABin'.
taxVector	The taxon vector. Default assumes that each row in the alignment belongs to a different taxon.
taxOfInt	A vector containing the taxa for which diagnostic molecular characters shall be extracted. Default is "all".
types	A vector containing the types of diagnostic molecular characters that shall be extracted. The types can be "all" or any combination of "type1", "type2", "type3" and "type4". Default is "type1", "type 2" and "type3".
gapValid	Boolean variable denoting if a gap can be a characteristic state for taxon i (and taxon l in case of type 4). Default is TRUE.

Value

diagCharNA returns a list, where each entry belongs to one taxon of interest. Each taxon of interest has a set of diagnostic molecular characters (position, type, characteristic states for taxon of interest, compared taxa) assigned to it. type1 means that the character is suitable to distinguish each individual of the taxon of interest from all individuals of the remaining taxa, and that it is fixed for one state in the taxon of interest. type2 means that the character is suitable to distinguish each individual of the taxon of interest from all individuals of the remaining taxa, and that it is not fixed for one state in the taxon of interest. type3 means that the character is suitable to distinguish some (but not all) individuals of the taxon of interest from all individuals of the remaining taxa. type4 means that the character is suitable to distinguish each individual of the taxon of interest from all individuals of at least one (but not all) other taxon while being fixed in both the taxon of interest and the compared taxa.

diagCharNA returns for each taxon of interest the following elements:

position	The positions of its diagnostic molecular characters.
type	The types of the diagnostic molecular characters.
states	The states that are characteristic for the taxon i of interest, i.e. states that are distinct from "n" and unique to the taxon of interest (in case of type 1, 2 or 3), or fixed in the taxon of interest (type 4).
compared taxa	Only relevant for type 4 characters. It contains the name x if the character is found to be a type 4 character of the taxon of interest when being compared to taxon x.

Author(s)

A. Luise Kuehn <luise.kuehn@uni-greifswald.de>

References

Kuehn, A.L., Haase, M. 2019. QUIDDICH: QUick IDentification of DIagnostic CHaracters.

Examples

```
#using a dataset from spider
#install.packages("spider")
library(spider)
data("anoteropsis")
anoTax <- sapply(strsplit(dimnames(anoteropsis)[[1]], split="_"),
function(x) paste(x[1], x[2], sep="_"))
diagCharNA(anoteropsis, anoTax, taxOfInt="all")
diagCharNA(anoteropsis, anoTax, taxOfInt="all", types=c("type1","type2"))
#
#with loading of a fasta file
#install.packages("adegenet")
library(adegenet)
alignment <- fasta2DNABin(paste0(find.package("quiddich"), "/extData/example.fasta"))
taxonVector <- as.vector(sapply(dimnames(alignment)[[1]], function(x) substr(x,1,4)))
diagCharNA(alignment, taxonVector)
```

Index

`changesAA`, [2](#)

`diagCharAA`, [4](#)

`diagCharNA`, [5](#)

`quiddich-package`, [2](#)