Package ‘primerTree’

April 5, 2022

Title Visually Assessing the Specificity and Informativeness of Primer Pairs

Version 1.0.6

Description Identifies potential target sequences for a given set of primers and generates phylogenetic trees annotated with the taxonomies of the predicted amplification products.

License GPL-2

Depends R (>= 3.5.0), directlabels, gridExtra

Imports ape, foreach, ggplot2, grid, lubridate, plyr, reshape2, scales, stringr, XML

Encoding UTF-8

LazyData true

RoxygenNote 7.1.2

NeedsCompilation yes

Author Jim Hester [aut], Matt Cannon [aut, cre]

Maintainer Matt Cannon <matthewvc1@gmail.com>

Repository CRAN

Date/Publication 2022-04-05 14:30:02 UTC

R topics documented:

- accession2taxid ................................................................. 2
- bryophytes_trnL ............................................................... 2
- calc_rank_dist_ave ............................................................. 3
- clustalo .............................................................................. 4
- filter_seqs ................................................................. 4
- get_sequence ............................................................... 5
- get_sequences ............................................................. 6
- get_taxonomy .............................................................. 7
- identify.primerTree_plot ................................................ 7
accession2taxid

Maps a nucleotide database accession to a taxonomy database taxId

Usage

accession2taxid(accessions)

Arguments

accessions character vector to lookup.

Value
	named vector of taxIds.

bryophytes_trnL

PrimerTree results for the bryophyte trnL primers

Description

PrimerTree results for the bryophyte trnL primers
calc_rank_dist_ave

Summarize pairwise differences.

Description
Summarize pairwise differences.

Usage
calc_rank_dist_ave(x, ranks = common_ranks)

Arguments
- **x**: a primerTree object
- **ranks**: ranks to show unique counts for, defaults to the common ranks

Details
The purpose of this function is to calculate the average number of nucleotide differences between species within each taxa of given taxonomic level.

For example, at the genus level, the function calculates the average number of nucleotide differences between all species within each genus and reports the mean of those values.

There are several key assumptions and calculations made in this function.

First, the function randomly selects one sequence from each species in the primerTree results. This is to keep any one species (e.g. human, cow, etc.) with many hits from skewing the results.

Second, for each taxonomic level tested, the function divides the sequences by each taxon at that level and calculates the mean number of nucleotide differences within that taxa, then returns the mean of those values.

Third, when calculating the average distance, any taxa for which there is only one species is omitted, as the number of nucleotide differences will always be 0.

Value
returns a data frame of results

Examples

```r
## Not run:
calc_rank_dist_ave(mammals_16S)
calc_rank_dist_ave(bryophytes_trnL)

# Note that the differences between the results from these two primers
# the mean nucleotide differences is much higher for the mammal primers
# than the byrophyte primers. This suggests that the mammal primers have
# better resolution to distinguish individual species.

## End(Not run)
```
clustalo  *Multiple sequence alignment with clustal omega*

**Description**

Calls clustal omega to align a set of sequences of class DNAbin. Run without any arguments to see all the options you can pass to the command line clustal omega.

**Usage**

```r
clustalo(x, exec = "clustalo", quiet = TRUE, original.ordering = TRUE, ...)
```

**Arguments**

- `x`: an object of class 'DNAbin'
- `exec`: a character string with the name or path to the program
- `quiet`: whether to suppress output to stderr or stdout
- `original.ordering`: use the original ordering of the sequences
- `...`: additional arguments passed to the command line clustalo

---

filter_segs  *Filter out sequences retrieved by search_primer_pair() that are either too short or too long. The alignment and tree will be recalculated after removing unwanted reads.*

**Description**

Filter out sequences retrieved by search_primer_pair() that are either too short or too long. The alignment and tree will be recalculated after removing unwanted reads.

**Usage**

```r
filter_segs(x, ...)
```

**Arguments**

- `x`: a primerTree object
- `...`: additional arguments passed to methods.
- `min_length`: the minimum sequence length to keep
- `max_length`: the maximum sequence length to keep
get_sequence

Value

  a primerTree object

Methods (by class)

  • primerTree: Method for primerTree objects

Examples

  ## Not run:
  # filter out sequences longer or shorter than desired:
  mammals_16S_filtered <- filter_seqs(mammals_16S, min_length=131, max_length=156)
  ## End(Not run)

get_sequence

Retrieves a fasta sequence from NCBI nucleotide database.

Description

Retrieves a fasta sequence from NCBI nucleotide database.

Usage

  get_sequence(
    accession,
    start = NULL,
    stop = NULL,
    api_key = Sys.getenv("NCBI_API_KEY")
  )

Arguments

  accession        nucleotide accession to retrieve.
  start            start base to retrieve, numbered beginning at 1. If NULL the beginning of the sequence.
  stop             last base to retrieve, numbered beginning at 1. if NULL the end of the sequence.
  api_key          NCBI api-key to allow faster sequence retrieval.

Value

  an DNAbin object.

See Also

  DNAbin
get_sequences

Retrieves fasta sequences from NCBI nucleotide database.

Description

Retrieves fasta sequences from NCBI nucleotide database.

Usage

get_sequences(
  accession,
  start = NULL,
  stop = NULL,
  api_key = Sys.getenv("NCBI_API_KEY"),
  simplify = TRUE,
  .parallel = FALSE,
  .progress = "none"
)

Arguments

accession the accession number of the sequence to retrieve
start start bases to retrieve, numbered beginning at 1. If NULL the beginning of the sequence.
stop stop bases to retrieve, numbered beginning at 1. if NULL the stop of the sequence.
api_key NCBI api-key to allow faster sequence retrieval.
simplify simplify the FASTA headers to include only the genbank accession.
.parallel if 'TRUE', perform in parallel, using parallel backend provided by foreach
.progress name of the progress bar to use, see 'create_progress_bar'

Value

an DNAbin object.

See Also

DNAbin
### get_taxonomy

Retrieves the taxonomy information from NCBI for a set of nucleotide gis.

#### Description

Retrieve the taxonomy information from NCBI for a set of nucleotide gis.

#### Usage

```r
get_taxonomy(accessions)
```

#### Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>accessions</td>
<td>a character vector of the accessions to retrieve</td>
</tr>
</tbody>
</table>

#### Value

`data.frame` of the 'accessions, taxIds, and taxonomy`

### identify.primerTree_plot

Identify the point closest to the mouse click only works on single ranks.

#### Description

Identify the point closest to the mouse click only works on single ranks.

#### Usage

```r
## S3 method for class 'primerTree_plot'
identify(x, ...)
```

#### Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>the plot to identify</td>
</tr>
<tr>
<td>...</td>
<td>additional arguments passed to annotate</td>
</tr>
</tbody>
</table>
parse_primer_hits

---

**layout_tree_ape**

*layout a tree using ape, return an object to be plotted by plot_tree*

**Description**

layout a tree using ape, return an object to be plotted by plot_tree

**Usage**

`layout_tree_ape(tree, ...)`

**Arguments**

- `tree` The phylo tree to be plotted
- `...` additional arguments to plot.phylo

**Value**

- `edge` list of x, y and xend, yend coordinates as well as ids for the edges
- `tips` list of x, y, label and id for the tips
- `nodes` list of x, y and id for the nodes

---

**mammals_16S**

*PrimerTree results for the mammalian 16S primers*

**Description**

PrimerTree results for the mammalian 16S primers

---

**parse_primer_hits**

*Parse the primer hits*

**Description**

Parse the primer hits

**Usage**

`parse_primer_hits(response)`

**Arguments**

- `response` a httr response object obtained from primer_search
Description

plot function for a primerTree object, calls plot_tree_ranks

Usage

## S3 method for class 'primerTree'
plot(x, ranks = NULL, main = NULL, ...)

Arguments

- **x**: primerTree object to plot
- **ranks**: The ranks to include, defaults to all common ranks, if NULL print all ranks. If 'none' just print the layout.
- **main**: an optional title to display, if NULL displays the name as the title
- **...**: additional arguments passed to plot_tree_ranks

See Also

plot_tree_ranks, plot_tree

Examples

library(gridExtra)
library(directlabels)
#plot with all common ranks
plot(mammals_16S)

#plot only the class
plot(mammals_16S, 'class')

#plot the layout only
plot(mammals_16S, 'none')
plot_tree_ranks

Usage

plot_tree_ranks(
  tree,
  taxonomy,
  main = NULL,
  type = "unrooted",
  ranks = common_ranks,
  ...  
)

Arguments

  tree to be plotted, use layout_tree to layout tree.
  type The type of tree to plot, default unrooted.
  main An optional title for the plot
  guide_size The size of the length guide. If NULL auto detects a reasonable size.
  rank The rank to include, if null only the tree is plotted
  taxonomy A data.frame with an accession field corresponding to the tree tip labels.
  size The size of the colored points
  legend_cutoff The number of different taxa names after which the names are no longer printed.
  ... additional arguments passed to layout_tree_ape

Value

  plot to be printed.

plot_tree_ranks plots a tree along with a series of taxonomic ranks

Description

  plots a tree along with a series of taxonomic ranks

Usage

plot_tree_ranks(
  tree,
  taxonomy,
  main = NULL,
  type = "unrooted",
  ranks = common_ranks,
  ...  
)
primerTree has two main commands: search_primer_pair which takes a primer pair and returns a primerTree object of the search results, and plot.primerTree a S3 method for plotting the primerTree object obtained using search_primer_pair.
primer_search

Query a pair of primers using ncbi's Primer-BLAST, if primers contain iupac ambiguity codes, enumerate all possible combinations and combine the results.

Usage

primer_search(
  forward,
  reverse,
  num_aligns = 500,
  num_permutations = 25,
  ...,
  .parallel = FALSE,
  .progress = "none"
)

Arguments

forward forward primer to search by 5'-3' on plus strand
reverse reverse primer to search by 5'-3' on minus strand
num_aligns number of alignment results to keep
num_permutations the number of primer permutations to search, if the degenerate bases cause more than this number of permutations to exist, this number will be sampled from all possible permutations.
... additional arguments passed to Primer-Blast
.parallel if 'TRUE', perform in parallel, using parallel backend provided by foreach
.progress name of the progress bar to use, see 'create_progress_bar'

Value

httr response object of the query, pass to parse_primer_hits to parse the results.
<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>forward</td>
<td>forward primer to search by 5'-3' on plus strand</td>
</tr>
<tr>
<td>reverse</td>
<td>reverse primer to search by 5'-3' on minus strand</td>
</tr>
<tr>
<td>name</td>
<td>name to give to the primer pair</td>
</tr>
<tr>
<td>num_aligns</td>
<td>number of alignment results to keep</td>
</tr>
<tr>
<td>num_permutations</td>
<td>the number of primer permutations to search, if the degenerate bases cause more than this number of permutations to exist, this number will be sampled from all possible permutations.</td>
</tr>
<tr>
<td>simplify</td>
<td>use simple names for primer hit results or complex</td>
</tr>
<tr>
<td>clustal_options</td>
<td>a list of options to pass to clustal omega, see link(clustalo) for a list of options</td>
</tr>
<tr>
<td>distance_options</td>
<td>a list of options to pass to dist.dna, see link(dist.dna) for a list of options</td>
</tr>
<tr>
<td>api_key</td>
<td>NCBI api-key to allow faster sequence retrieval</td>
</tr>
<tr>
<td>...</td>
<td>additional arguments passed to Primer-Blast</td>
</tr>
</tbody>
</table>
seq_lengths

Description

Get a summary of sequence lengths from a primerTree object

Usage

seq_lengths(x, summarize = TRUE)

Arguments

x

a primerTree object.

summarize

a logical indicating if a summary should be displayed
**seq_lengths.primerTree**

**Value**

a table of sequence length frequencies

**Examples**

```r
# Show the counts for each length
seq_lengths(mammals_16S)

# Plot the distribution of lengths
seqLengths <- seq_lengths(mammals_16S)
barplot(seqLengths, 
main = "Frequency of sequence lengths for 16S mammal primers",
 xlab="Amplicon length (in bp)",
ylab="(Frequency)"
```

---

**seq_lengths.primerTree**

*Method for primerTree objects*

**Description**

Method for primerTree objects

**Usage**

```r
## S3 method for class 'primerTree'
seq_lengths(x, summarize = TRUE)
```

**Arguments**

- `x` a primerTree object.
- `summarize` a logical indicating if a summary should be displayed

---

**summary.primerTree**

*Summarize a primerTree result, printing quantiles of sequence length and pairwise differences.*

**Description**

Summarize a primerTree result, printing quantiles of sequence length and pairwise differences.

**Usage**

```r
## S3 method for class 'primerTree'
summary(object, ..., probs = c(0, 0.05, 0.5, 0.95, 1), ranks = common_ranks)
```
Arguments

object the primerTree object to summarise
...
probs quantile probabilities to compute, defaults to 0, 5, 50, 95, and 100 probabilities.
ranks ranks to show unique counts for, defaults to the common ranks

Value

invisibly returns a list containing the printed results

Description

Construct a neighbor joining tree from a dna alignment

Usage

tree_from_alignment(dna, pairwise.deletion = TRUE, ...)

Arguments

dna fasta dna object the tree is to be constructed from
pairwise.deletion a logical indicating if the distance matrix should be constructed using pairwise deletion
...

See Also

dist.dna, nj
# Index

accession2taxid, 2
bryophytes_trnL, 2
calc_rank_dist_ave, 3
clustalo, 4, 14
create_progress_bar, 14
dist.dna, 16
DNAbin, 5, 6
filter_seqs, 4
get_sequence, 5
get_sequences, 6
get_taxonomy, 7
identify.primerTree_plot, 7
layout_tree_ape, 8, 10, 11
mammals_16S, 8
nj, 16
parse_primer_hits, 8, 12
phylo, 8
plot.phylo, 8
plot.primerTree, 9, 11
plot_tree, 8, 9, 11
plot_tree_ranks, 9, 10
primer_search, 8, 12, 14
primerTree, 11
response, 14
search_primer_pair, 11, 13
seq_lengths, 14
seq_lengths.primerTree, 15
summary.primerTree, 15
tree_from_alignment, 16