Package 'primerTree'

January 23, 2020

Version 1.0.5

Maintainer Jim Hester <james.f.hester@gmail.com>

Author Jim Hester

Title Visually Assessing the Specificity and Informativeness of Primer Pairs

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

Depends R (>= 3.0.0), directlabels, gridExtra

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

LazyData true

License GPL-2

RoxygenNote 7.0.2

Encoding UTF-8

NeedsCompilation yes

Repository CRAN

Date/Publication 2020-01-23 22:50:06 UTC

R topics documented:

accession2gi 2
bryophytes_trnL 2
calc_rank_dist_ave 3
clustalo 4
filter_seqs 4
get_sequence 5
get_sequences 6
get_taxonomy 7
gi2taxid 7
identify.primerTree_plot 8

layout_tree_ape 8
mammals_16S 9
parse_primer_hits 9
plot.primerTree 9
plot_tree 10
plot_tree_ranks 11
primerTree 12
primer_search 12
search_primer_pair 13
seq_lengths 15
seq_lengths.primerTree 15
summary.primerTree 16
tree_from_alignment 16

Index **18**

`accession2gi` *Maps a genbank accession to a nuclotide database gi.*

Description

Maps a genbank accession to a nuclotide database gi.

Usage

```
accession2gi(accession)
```

Arguments

`accession`	accession character vector to lookup.

Value

named vector of gis.

`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

```
## 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

```
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 supress output to stderr or stdout
original.ordering	use the original ordering of the sequences
...	additional arguments passed to the command line clustalo

filter_seqs — *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

```
filter_seqs(x, ...)

## S3 method for class 'primerTree'
filter_seqs(x, min_length = 0, max_length = Inf, ...)
```

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

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(
  gi,
  start = NULL,
  stop = NULL,
  api_key = Sys.getenv("NCBI_API_KEY")
)
```

Arguments

`gi`	nucleotide gi 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(
  gi,
  start = NULL,
  stop = NULL,
  api_key = Sys.getenv("NCBI_API_KEY"),
  simplify = TRUE,
  .parallel = FALSE,
  .progress = "none"
)
```

Arguments

gi	the gi 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` *Retrieve 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

```
get_taxonomy(gis)
```

Arguments

`gis`	a character vector of the gis to retrieve

Value

data.frame of the 'gis, taxIds, and taxonomy

`gi2taxid` *Maps a nucleotide database gi to a taxonomy database taxId*

Description

Maps a nucleotide database gi to a taxonomy database taxId

Usage

```
gi2taxid(gi)
```

Arguments

`gi`	gi character vector to lookup.

Value

named vector of taxIds.

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

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

Arguments

x	the plot to identify
...	additional arguments passed to annotate

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

plot.primerTree *plot function for a primerTree object, calls plot_tree_ranks*

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 *plots a tree, optionally with colored and labeled points by taxonomic rank*

Description

plots a tree, optionally with colored and labeled points by taxonomic rank

Usage

```
plot_tree(
  tree,
  type = "unrooted",
  main = NULL,
  guide_size = NULL,
  rank = NULL,
  taxonomy = NULL,
  size = 2,
  legend_cutoff = 25,
  ...
)
```

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,
  size = 2,
  guide_size = NULL,
  legend_cutoff = 25,
  ...
)
```

Arguments

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

See Also

plot_tree to plot only a single rank or the just the tree layout.

Examples

```
library(gridExtra)
library(directlabels)
#plot all the common ranks
plot_tree_ranks(mammals_16S$tree, mammals_16S$taxonomy)
#plot specific ranks, with a larger dot size
plot_tree_ranks(mammals_16S$tree, mammals_16S$taxonomy,
  ranks=c('kingdom', 'class', 'family'), size=3)
```

`primerTree` **primerTree** *Visually Assessing the Specificity and Informativeness of Primer Pairs*

Description

`primerTree` has two main commands: `search_primer_pair` which takes a primer pair and returns an primerTree object of the search results `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*

Description

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.

`search_primer_pair` *Automatic primer searching Search a given primer pair, retrieving the alignment results, their product sequences, the taxonomic information for the sequences, a multiple alignment of the products*

Description

Automatic primer searching Search a given primer pair, retrieving the alignment results, their product sequences, the taxonomic information for the sequences, a multiple alignment of the products

Usage

```
search_primer_pair(
  forward,
  reverse,
  name = NULL,
  num_aligns = 500,
  num_permutations = 25,
  simplify = TRUE,
  clustal_options = list(),
  distance_options = list(model = "N", pairwise.deletion = T),
  api_key = Sys.getenv("NCBI_API_KEY"),
  ...,
  .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

name
: name to give to the primer pair

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.

simplify
: use simple names for primer hit results or complex

clustal_options
: a list of options to pass to clustal omega, see link{clustalo} for a list of options

distance_options
: a list of options to pass to dist.dna, see link{dist.dna} for a list of options

api_key
: NCBI api-key to allow faster sequence retrieval

...
: 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

A list with the following elements,

name
: name of the primer pair

BLAST_result
: html blast results from Primer-BLAST as 'a response

object.

taxonomy
: taxonomy for the primer products from NCBI

sequence
: sequence of the primer products

alignment
: multiple alignment of the primer products

tree
: phylogenetic tree of the reconstructed from the 'multiple alignment

See Also

primer_search, clustalo

Examples

```
## Not run:
#simple search
mammals_16S = search_primer_pair(name='Mammals 16S',
 'CGGTTGGGGTGACCTCGGA', 'GCTGTTATCCCTAGGGTAACT')
#returning 1000 alignments, allow up to 3 mismatches in primer
mammals_16S = search_primer_pair(name='Mammals 16S',
 'CGGTTGGGGTGACCTCGGA', 'GCTGTTATCCCTAGGGTAACT',
 num_aligns=1000, total_primer_specificity_mismatch=3)

## End(Not run)
```

seq_lengths *Get a summary of sequence lengths from a primerTree object*

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

Value

a table of sequence length frequencies

Examples

```
# 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

```
## 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

```
## 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
...	Ignored options
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

---

tree_from_alignment *Construct a neighbor joining tree from a dna alignment*

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
...	furthur arguments to dist.dna

See Also

dist.dna, nj

Index

accession2gi, 2

bryophytes_trnL, 2

calc_rank_dist_ave, 3
clustalo, 4, *14*
create_progress_bar, *14*

dist.dna, *17*
DNAbin, *5*, *6*

filter_seqs, 4

get_sequence, 5
get_sequences, 6
get_taxonomy, 7
gi2taxid, 7

identify.primerTree_plot, 8

layout_tree_ape, 8, *11*

mammals_16S, 9

nj, *17*

parse_primer_hits, 9, *13*
phylo, *8*
plot.phylo, *8*
plot.primerTree, 9, *12*
plot_tree, *8*, *10*, 10, *11*
plot_tree_ranks, *10*, 11
primer_search, *9*, 12, *14*
primerTree, 12

response, *14*

search_primer_pair, *12*, 13
seq_lengths, 15
seq_lengths.primerTree, 15
summary.primerTree, 16

tree_from_alignment, 16