Package ‘kmeRs’

November 3, 2018

Title  K-Mers Similarity Score Matrix
Version  1.1.0
Description  Contains tools to calculate similarity score matrix for DNA k-mers. The pairwise similarity score is calculated using PAM or BLOSUM substitution matrix. The results are evaluated by similarity score calculated by Needleman-Wunsch (1970) <doi:10.1016/0022-2836(70)90057-4> global or Smith-Waterman (1981) <doi:10.1016/0022-2836(81)90087-5> local alignment. Higher similarity score indicates more similar sequences for BLOSUM and less similar sequences for PAM matrix; 30, 40, 70, 120, 250 and 62, 45, 50, 62, 80, 100 matrix versions are available for PAM and BLOSUM, respectively.
Depends  R (>= 3.4.0)
License  GPL-3
Encoding  UTF-8
URL  https://rafalurniaz.github.io/kmeRs/
BugReports  http://github.com/RafalUrniaz/kmeRs/issues
LazyData  true
biocViews  Software
Imports  tcR, rDNAse, utils, stats, Biostrings
RoxygenNote  6.1.0
Suggests  knitr, ape (>= 4.0), rmarkdown
VignetteBuilder  knitr
NeedsCompilation  no
Author  Rafal Urniaz [aut, cre] (<https://orcid.org/0000-0003-0192-2165>)
Maintainer  Rafal Urniaz <rafal.urniaz@gmail.com>
Repository  CRAN
Date/Publication  2018-11-03 16:00:07 UTC
R topics documented:

- `kmeRs_score_and_sort` ................................. 2
- `kmeRs_show_alignment` ................................. 3
- `kmeRs_similarity_matrix` ............................. 3
- `kmeRs_statistics` ......................................... 4
- `kmeRs_test_package` ..................................... 5
- `kmeRs_transcript_to_complementary` ................. 6

Index

---

**kmeRs_score_and_sort**  Score And Sort The Similarity Matrix

Description

The `kmeRs_score_and_sort` function sums the partial scores and sort the data.frame to indicate the most 'different' k-mers

Usage

```r
kmeRs_score_and_sort(kmeRs_similarity_matrix)
```

Arguments

- `kmeRs_similarity_matrix`

  the similarity matrix calculated by `kmeRs_similarity_matrix` function

Value

sorted similarity matrix with global.score column added; is returned as a data.frame

Examples

```r
# Calculate the example BLOSUM62 matrix and score the result
example <- kmeRs_similarity_matrix(kmers_given = c("A", "T", "C", "G"), submat = "BLOSUM62")
kmeRs_score_and_sort(example)
```
**kmers_show_alignment**

*Calculate and Show Alignment Between Two Compared K-mers*

**Description**

The `kmers_show_alignment` function aligns and shows calculated alignment between two DNA or RNA sequences.

**Usage**

```r
kmers_show_alignment(kmer_A = "", kmer_B = ", submat = "BLOSUM62")
```

**Arguments**

- `kmer_A` given k-mer A
- `kmer_B` given k-mer B
- `submat` substitution matrix version, default is 'BLOSUM62'; could be one of 'BLOSUM45', 'BLOSUM50', 'BLOSUM62', 'BLOSUM80', 'BLOSUM100', 'PAM30', 'PAM40', 'PAM70', 'PAM120', 'PAM250'

**Value**

alignment is returned as a data.frame

**Examples**

```r
# Example alignment
kmers_show_alignment( kmer_A = "AAATTTCGGG", kmer_B = "TCACCC", submat = "BLOSUM62")
```

---

**kmers_similarity_matrix**

*Pairwise Similarity Matrix*

**Description**

The `kmers_similarity_matrix` function generates a pairwise similarity score matrix for for k length given k-mers vs. all possible k-mers combination. The pairwise similarity score is calculated using PAM or BLOSUM substitution matrix; 30, 40, 70, 120, 250 and 62, 45, 50, 62, 80, 100 matrix versions are available for PAM or BLOSUM, respectively. The results are evaluated by global similarity score; higher similarity score indicates more similar sequences for BLOSUM and opposite for PAM matrix.
Usage

kmeRs_similarity_matrix(kmers_given, compare_to = "", alignment_type = "global", k = 3, submat = "BLOSUM62", save_to_file = "")

Arguments

- **kmers_given**: vector with given k-mers
- **compare_to**: this parameter can have 3 different states, when "" - the kmers_given will be compared to each other, default value; ALL - the kmers_given will be compared to all possible combinations given by k parameter e.g. N= 4^6 = 4096 combinations for 6-mers; 3rd option is to provide a list of k-mers which should be compared with the set given by the kmers_given parameter
- **alignment_type**: type of alignment, default is 'global', could be 'local' or 'global', where 'global' represents Needleman-Wunsch global alignment; 'local' represents Smith-Waterman local alignment.
- **k**: length of k-mers to calculate similarity matrix, higher values may slow down the computer, default value is k=3
- **submat**: substitution matrix, default is 'BLOSUM62', but could be one of 'BLOSUM45', 'BLOSUM50', 'BLOSUM62', 'BLOSUM80', 'BLOSUM100', 'PAM30', 'PAM40', 'PAM70', 'PAM120', 'PAM250'
- **save_to_file**: directory and file name; if value is declared the matrix will be saved to the given file name

Value

similarity matrix is returned as a data.frame

Examples

```r
# Display BLOSUM matrix used for calculation
kmeRs_similarity_matrix(kmers_given = c("A", "T", "C", "G"), submat = "BLOSUM62")
```

---

**kmeRs_statistics**

*Calculate And Add Basic Statistics To The K-mers Similarity Matrix*

Description

The kmeRs_statistics function calculates basic statistics and returns the similarity matrix with calculated results or summarized table with statistics only when summary_statistics_only is set to TRUE
Usage

kmeRs_statistics(kmers_similarity_matrix,
summary_statistics_only = FALSE)

Arguments

kmers_similarity_matrix
matrix calculated by kmeRs_similarity_matrix function
summary_statistics_only
when parameter is set to TRUE only the summarized table with statistics is returned

Value

data.frame with results

Examples

# Simple BLOSUM62 similarity matrix for DNA nucleotides
eexample <- kmeRs_similarity_matrix(kmers_given = c("A", "T", "C", "G"), submat = "BLOSUM62")

# Result as a full matrix
ekmeRs_statistics(example)

# Result a summary statistics table
kmeRs_statistics(example, summary_statistics_only = TRUE)

kmeRs_test_package

Simple Demo For KmeRs Package

Description

The kmeRs_test_package function calculates and shows an example report of the kmeRs package for sample given k-mers: "ATA", "CGC", "TGC", "GGA"

Usage

kmeRs_test_package()

Value

eexample report

Examples

# Test package - example raport

kmeRs_test_package()
kmeRs_transcript_to_complementary

Translate Given K-mers To Complementary Sequences

Description
The kmeRs_transcript_to_complementary function transcripts DNA given k-mers to complementary sequences

Usage
kmeRs_transcript_to_complementary(kmers_given)

Arguments

kmers_given vector contains given k-mers

Value
vector contains complementary sequences

Examples

# Returns complementary sequence to GATTACA

kmeRs_transcript_to_complementary('GATTACA')
Index

kmeRs_score_and_sort, 2
kmeRs_show_alignment, 3
kmeRs_similarity_matrix, 3
kmeRs_statistics, 4
kmeRs_test_package, 5
kmeRs_transcript_to_complementary, 6