

Package ‘SmithWaterman’

January 27, 2023

Type Package

Title Pariwise Local Sequence Alignment of Two Nucleic Acid Sequences

Description Contains a function that performs local sequence alignment according to the Smith Waterman greedy algorithm.
It finds the optimal local alignment of two nucleotide sequences in terms of score.
This package was created for the Scientific Programming course delivered by Politecnico of Milan,
included in the Bioinformatics for Computational Genomics MSc.

Version 0.1.0

Date 2023-01-25

License GPL-2

Encoding UTF-8

Depends R(>= 4.1.0), Biostrings, methods

Suggests knitr, rmarkdown, testthat, BiocStyle

VignetteBuilder knitr

RoxygenNote 7.2.3

biocViews Software, BiologicalQuestion, SequenceMatching,
BiomedicalInformatics, Genetics, Alignment, Sequencing,
DataImport

NeedsCompilation no

Author Sara Ometto [aut, cre]

Maintainer Sara Ometto <sara.ometto@mail.polimi.it>

R topics documented:

best_alignment	2
fill_matrices	2
pairwise_score	3
smith_waterman	3

best_alignment	<i>Find the best alignment</i>
----------------	--------------------------------

Description

Starting from the element with the highest score, this function performs the traceback based on the source of each score recursively, until 0 is encountered. The function returns the optimal local alignment between the two sequences and its score.

Usage

```
best_alignment(scoringMat, tracebackMat)
```

Arguments

scoringMat	filled Scoring Matrix obtained from fill_matrices applying the Smith-Waterman algorithm.
tracebackMat	filled Traceback Matrix obtained from fill_matrices applying the Smith-Waterman algorithm.

Value

A list of the two aligned sequences, the score of the optimal alignment and a matrix with starting ending position.

fill_matrices	<i>Fill the Scoring Matrix and Traceback Matrix</i>
---------------	---

Description

For each pair of nucleotides, the function assigns a score using the Smith Waterman greedy algorithm. It inserts the value in the scoring matrix and, depending on the score, it stores the direction of the alignment in the traceback matrix.

Usage

```
fill_matrices(scoringMat, tracebackMat, seqA, seqB, match, mismatch, gap)
```

Arguments

scoringMat	Initialized empty Scoring Matrix of integers, with the first row and first column set to 0.
tracebackMat	Initialized empty Traceback Matrix of characters, with the first row and column set to 0.
seqA	A vector of the first nucleotide sequence with "-" as first element.
seqB	A vector of the second nucleotide sequence with "-" as first element.
match	A positive number associated with a nucleotide match.
mismatch	A negative number associated with a nucleotide mismatch.
gap	A negative number associated with the gap penalty.

Value

A list with the filled Scoring Matrix and the filled Traceback Matrix.

pairwise_score	<i>Calculate Pairwise Score</i>
----------------	---------------------------------

Description

This function assigns to a pair of nucleotides a score, whether the two nucleotides match or not. The value depends on the parameters match and mismatch, which are part of the scoring system decided a priori.

Usage

```
pairwise_score(nt1, nt2, match, mismatch)
```

Arguments

nt1	Nucleotide of the first sequence, a character.
nt2	Nucleotide of the second sequence, a character.
match	A positive number associated with a nucleotide match.
mismatch	A negative number associated with a nucleotide mismatch.

Value

An integer representing the score associated to the comparison of the two nucleotides.

smith_waterman	<i>Local Sequence Alignment</i>
----------------	---------------------------------

Description

This function performs pairwise local sequence alignment of two nucleotide sequences given as input. It uses a scoring system and the Smith Waterman greedy algorithm to compute the scoring and traceback matrices and returns one of the possible optimal alignments.

Usage

```
smith_waterman(sequenceA, sequenceB, match = 2, mismatch = -1, gap = -1)
```

Arguments

sequenceA	First nucleotide sequence. Can be a string, DNASTring or DNASTringSet object.
sequenceB	Second nucleotide sequence. Can be a string, DNASTring or DNASTringSet object.
match	A positive number associated with a nucleotide match. Default: 2
mismatch	A negative number associated with a nucleotide mismatch. Default: -1
gap	A negative number associated with the gap penalty. Default: -1

Value

A list containing the optimal alignment (DNASTring object), the resulting score and a matrix with the start and end positions. If the two sequences are shorter than 20 nucleotides, the scoring matrix and the traceback matrix are added to the list.

Author(s)

Sara Ometto
Politecnico di Milano
Maintainer: Sara Ometto
E-Mail: <sara.ometto@mail.polimi.it> or <sara.ometto@studenti.unimi.it>

References

https://en.wikipedia.org/wiki/Smith-Waterman_algorithm

See Also

[pairwise_score](#)
[fill_matrices](#)
[best_alignment](#)

Examples

```
library(Biostrings)
smith_waterman('GCATGCG', DNASTring('GATTACA'))
```