Description

The Needleman-Wunsch simple gap algorithm was one of the first methods introduced for global alignment of biological sequences. The same algorithm can be used to match cell line names or sample names from two related data sets; we provide examples in the documentation, using data that accompanies this package.

While the NameNeedle package can be used for biological sequence alignment, the Biostrings package from Bioconductor contains much more sophisticated tools for that problem.

Details
# cellLineNames-data

<table>
<thead>
<tr>
<th>Package:</th>
<th>NameNeedle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type:</td>
<td>Package</td>
</tr>
<tr>
<td>Version:</td>
<td>1.1.1</td>
</tr>
<tr>
<td>Date:</td>
<td>2012-05-01</td>
</tr>
<tr>
<td>License:</td>
<td>GPL Version 2 or newer</td>
</tr>
<tr>
<td>LazyLoad:</td>
<td>yes</td>
</tr>
</tbody>
</table>

## Author(s)
Kevin R. Coombes <krc@silicovore.com>, P. Roebuck <proebuck@mdanderson.org>

## References
Wang J, Byers LA, Yordy JS, Liu W, Shen L, Baggerly KA, Giri U, Myers JN, Ang KK, Story MD, Girard L, Minna JD, Heymach JV, Coombes KR.
*Blasted cell line names.*

## Description
This dataset contains vectors of cell line names that are used to demonstrate how to use the NameNeedle package.

## Usage
data(cellLineNames)

## Format
This dataset contains four objects: three character vectors (`sf2Names`, `rppaNames`, and `illuNames`) and one factor (`illuType`).

## Details
The three character vectors, `sf2Names`, `rppaNames`, and `illuNames` contain the names of cell lines used in three different but related experiments. The factor, `illuType`, indicates whether the cell lines named in the `illuNames` vector were derived from lung cancer (with the value "Lung") or from head and neck cancer ("HNSCC").

## Examples
data(cellLineNames)
head(rppaNames)
head(sf2Names)
head(illuNames)
summary(illuType)
**needles**

**Needleman-Wunsch global alignment algorithm**

**Description**

This package contains a simple implementation of the Needleman-Wunsch global alignment algorithm.

**Usage**

```r
needles(pattern, subject, params=defaultNeedleParams)
needleScores(pattern, subjects, params=defaultNeedleParams)
defaultNeedleParams
```

**Arguments**

- `pattern`: character string to be matched
- `subject`: character string to be matched against
- `subjects`: character vector where matches are sought
- `params`: list containing four required components. The default values are specified by the object `defaultNeedleParams`, which contains the following values:
  - `$ MATCH`: num 1
  - `$ MISMATCH`: num -1
  - `$ GAP`: num -1
  - `$ GAPCHAR`: chr "*"

**Details**

The Needleman-Wunsch global alignment algorithm was one of the first algorithms used to align DNA, RNA, or protein sequences. The basic algorithm uses dynamic programming to find an optimal alignment between two sequences, with parameters that specify penalties for mismatches and gaps and a reward for exact matches. More elaborate algorithms (not implemented here) make use of matrices with different penalties depending on different kinds of mismatches. The version implemented here is based on the Perl implementation in the first section of Chapter 3 of the book *BLAST*.

**Value**

The `needles` function returns a list with five components:

- `score`: The raw alignment score.
- `align1`: The final (optimal) alignment for the `pattern`.
- `align2`: The final (optimal) alignment for the `subject`.
- `sm`: The score matrix.
- `dm`: The backtrace matrix.

The `needleScores` function returns a numeric vector the same length as the `subjects` argument, with each entry equal to the corresponding raw alignment score.
Author(s)

Kevin R. Coombes <krc@silicovore.com>, P. Roebuck <proebuck@mdanderson.org>

References

Needleman SB, Wunsch CD.
A general method applicable to the search for similarities in the amino acid sequence of two pro-
teins.

Korf I, Yandell M, Bedell J.
BLAST.

Wang J, Byers LA, Yordy JS, Liu W, Shen L, Baggerly KA, Giri U, Myers JN, Ang KK, Story MD, Girard L, Minna JD, Heymach JV, Coombes KR.
Blasted cell line names.

See Also

The Biostrings package from Bioconductor used to contain a function called needwunQS that pro-
vided a simple gap implementation of Needleman-Wunsch, similar to the one presented here. That
function has been deprecated in favor of a more elaborate interface called pairwiseAlignment that
incorporates a variety of other alignment methods in addition. While pairwiseAlignment is much
more useful for applications to biological sequences, it is serious overkill for the application we
have in mind for matching cell line or other sample names.

Examples

data(cellLineNames)
myParam <- defaultNeedleParams
myParam$MATCH <- 2
myParam$MISMATCH <- -2
needles(sf2Names[2], illuNames[1], myParam)
scores <- needleScores(sf2Names[6], illuNames, myParam)
w <- which(scores == max(scores))
w
sf2Names[6]

needles(sf2Names[6], illuNames[w], myParam)
Index

**Topic** character
  needles, 3

**Topic** datasets
  cellLineNames-data, 2

**Topic** package
  NameNeedle-package, 1

  cellLineNames (cellLineNames-data), 2
  cellLineNames-data, 2

  defaultNeedleParams (needles), 3

  illuNames (cellLineNames-data), 2
  illuType (cellLineNames-data), 2

  NameNeedle (NameNeedle-package), 1
  NameNeedle-package, 1
  needles, 3
  needleScores (needles), 3

  pairwiseAlignment, 4

  rppaNames (cellLineNames-data), 2

  sf2Names (cellLineNames-data), 2