

Package ‘odseq’

December 28, 2015

Type Package

Title Outlier detection in multiple sequence alignments

Version 0.9

Date 2015-12-20

Author José Jiménez

Maintainer José Jiménez <jose@jimenezluna.com>

Description Performs outlier detection using bootstrap of distance metrics of the multiple alignment.

License MIT + file LICENSE

LazyData True

Encoding UTF-8

Depends R (>= 3.2.3), msa (>= 1.2.1), kebabs (>= 1.4.1)

NeedsCompilation no

R topics documented:

odseq-package	1
odseq	3
odseq_unaligned	4
seqs	5

Index	6
--------------	----------

odseq-package	<i>Outlier detection in multiple sequence alignments</i>
---------------	--

Description

Performs outlier detection using bootstrap of distance metrics of the multiple alignment.

Details

The DESCRIPTION file:

Package: odseq
Type: Package
Title: Outlier detection in multiple sequence alignments
Version: 0.9
Date: 2015-12-20
Author: José Jiménez
Maintainer: José Jiménez <jose@jimenezluna.com>
Description: Performs outlier detection using bootstrap of distance metrics of the multiple alignment.
License: MIT + file LICENSE
LazyData: True
Encoding: UTF-8
Depends: R (>= 3.2.3), msa (>= 1.2.1), kebabs (>= 1.4.1)

Index of help topics:

odseq	Outlier detection in a multiple sequence alignment
odseq-package	Outlier detection in multiple sequence alignments
odseq_unaligned	Outlier detection provided a distance/similarity matrix of sequences.
seqs	PFAM plus random data.

Author(s)

José Jiménez

Maintainer: José Jiménez <jose@jimenezluna.com>

References

[1] OD-seq: outlier detection in multiple sequence alignments. *Peter Jehl, Fabian Sievers and Desmond G. Higgins*. BMC Bioinformatics. 2015.

See Also

[odseq](#) [odseq_unaligned](#)

Examples

```
data(seqs)
al <- msa(seqs)
odseq(al, distance_metric = "affine", B = 1000, threshold = 0.025)
```

odseq	<i>Outlier detection in a multiple sequence alignment</i>
-------	---

Description

This function will first compute a distance metric among every sequence in the multiple alignment. Then it will bootstrap an average score of these distance to provide information on the distribution of scores, which is used to distinguish outlier sequences with a certain threshold

Usage

```
odseq(msa_object, distance_metric = "linear", B = 100, threshold = 0.025)
```

Arguments

<code>msa_object</code>	An object of formal class <code>MsaAAMultipleAlignment</code> , as provided by the msa package.
<code>distance_metric</code>	A string indicating the type of distance metric to be computed. Either 'linear' and 'affine' is supported at the moment.
<code>B</code>	Integer indicating the number of bootstrap replicates to be run. The higher the more robust the detection should be.
<code>threshold</code>	Float indicating the probability to be left at the right of the bootstrap scores distribution when computing outliers. This parameter may need some tuning depending on each specific problem

Value

Returns a logical vector, where TRUE indicates an outlier.

Author(s)

José Jiménez <jose@jimenezluna.com>

References

[1] OD-seq: outlier detection in multiple sequence alignments. *Peter Jehl, Fabian Sievers and Desmond G. Higgins*. BMC Bioinformatics. 2015.

See Also

[odseq_unaligned](#)

Examples

```
data(seqs)
al <- msa(seqs)
odseq(al, distance_metric = "affine", B = 1000, threshold = 0.025)
```

odseq_unaligned	<i>Outlier detection provided a distance/similarity matrix of sequences.</i>
-----------------	--

Description

Provided a similarity matrix (like the ones provided using string kernels in **kebabs**). It will then compute a score for each sequence and perform bootstrap to provide information on the distribution of the scores, which is used to distinguish outlier sequences.

Usage

```
odseq_unaligned(distance_matrix, B = 100, threshold = 0.025, type = "similarity")
```

Arguments

distance_matrix	A numeric matrix representing either similarity or distance among unaligned sequences. Package kebabs may be useful for this task.
B	Integer indicating the number of bootstrap replicates to be run. The higher the more robust the detection should be.
threshold	Float indicating the probability to be left at the right of the bootstrap scores distribution when computing outliers. This parameter may need some tuning depending on each specific problem
type	A string indicating the type of distance metric used. Either 'similarity' or 'distance'.

Value

Returns a logical vector, where TRUE indicates an outlier.

Author(s)

José Jiménez <jose@jimenezluna.com>

References

[1] OD-seq: outlier detection in multiple sequence alignments. *Peter Jehl, Fabian Sievers and Desmond G. Higgins*. BMC Bioinformatics. 2015.

See Also

[odseq](#)

Examples

```
data(seqs)
sp <- spectrumKernel(k = 3)
mat <- getKernelMatrix(sp, seqs)
odseq_unaligned(mat, B = 1000, threshold = 0.025, type = "similarity")
```

seqs

PFAM plus random data.

Description

Sequences from a certain PFAM family plus 100 random sequences.

Usage

```
data("seqs")
```

Format

An object of class AAStringSet.

Examples

```
data(seqs)
```

Index

odseq, [2](#), [3](#), [4](#)

odseq-package, [1](#)

odseq_unaligned, [2](#), [3](#), [4](#)

seqs, [5](#)