

Package ‘mnmer’

May 12, 2023

Title '(m,n)-mer' - A Simple Statistical Feature for Sequence Classification

Version 0.99.1

Description The (m,n)-mer is a statistical feature calculated from conditional frequency distributions obtained from a FASTA file. The resulting table, along with class information, is used to create the classification feature matrix. For more information on this method and its benchmarking results, refer to Andrade et al.'s upcoming publication titled ``'(m,n)-mer - A Simple Statistical Feature for Sequence Classification''.

License Artistic-2.0

Encoding UTF-8

RoxygenNote 7.2.3

LinkingTo cpp11

VignetteBuilder knitr

Suggests BiocStyle, covr, knitr, rmarkdown, testthat (>= 3.0.0), caret, MLeval, randomForest

Depends R(>= 4.1.0), utils, Biostrings

biocViews Software, StatisticalMethod, FeatureExtraction, Classification, GenomeWideAssociation, GUI

NeedsCompilation yes

Author Amanda Andrade [cre, aut] (<<https://orcid.org/0000-0003-3530-7435>>),
Marco Grivet [aut] (<<https://orcid.org/0000-0001-5130-095X>>),
Otavio Brustolini [aut] (<<https://orcid.org/0000-0001-8132-9753>>),
Ana Vasconcelos [aut, ths] (<<https://orcid.org/0000-0002-4632-2086>>)

Maintainer Amanda Andrade <aandradebio@gmail.com>

Repository CRAN

Date/Publication 2023-05-12 14:30:05 UTC

R topics documented:

mnmer	2
readNumFASTA	2

Index**4**

mnmer	<i>mnmer</i>
-------	--------------

Description

Generates the feature matrix using conditional probability. As default, all sequences with N+IUPAC content higher 10% than are removed.

Usage

```
mnmer(seqs, m, n)
```

Arguments

seqs	DNAStrngSet object seqs
m	Int value of m m
n	Int value of n n

Value

Outputs a dataframe

Examples

```
dir <-system.file("extdata", package="mnmer")
human <- readDNAStrngSet(file.path(dir, "human_vir.fasta"))
human_02mer <- mnmer(human,2,0)
```

readNumFASTA	<i>readNumFASTA</i>
--------------	---------------------

Description

Load data in to the R system.

Usage

```
readNumFASTA(FASTAfile, size = 0, rand = FALSE, pni = 0.2)
```

Arguments

FASTAfile	Path to a multifasta file FASTAfile
size	Number of sequences to be loaded size
rand	Sequences choose mode random or not. TRUE or FALSE rand
pni	Cutoff percentage for maximum of non-ACTG bases in the sequences pni

Value

DNAStrngSet object

Examples

```
dir <-system.file("extdata", package="mnmer")
human <-readNumFASTA((file.path(dir, "human_vir.fasta")), 10, TRUE, 0.50)
```

Index

- * **featureMatrix**

- mnmer, [2](#)

- * **k-mers**

- mnmer, [2](#)

- * **mmmer**

- mnmer, [2](#)

- * **non-ACTG**

- readNumFASTA, [2](#)

- * **random**

- readNumFASTA, [2](#)

- * **sequences**

- readNumFASTA, [2](#)

cmnmer (mnmer), [2](#)

MNmer (mnmer), [2](#)

mnmer, [2](#)

readFASTA (readNumFASTA), [2](#)

readNum (readNumFASTA), [2](#)

readNumFASTA, [2](#)

readrandFASTA (readNumFASTA), [2](#)