

Package ‘MosVir’

March 1, 2024

Title Classify virus sequences in a FASTA format into Arbovirus, mosquito-specific, or other virus.

Version 0.0.0.9000

Description This package applies two groups of machine learning models using soft-voting. The first one is to classify each sequence for mosquito-target or another virus. The second gets only the mosquito-target sequences into the classification models for arbovirus or mosquito-specific viruses. Important: the input sequences must be previously classified as virus sequences.

License ``use_mit_license()``, ``use_gpl3_license()`` or friends to pick a license

Encoding UTF-8

Roxygen `list(markdown = TRUE)`

RoxygenNote 7.2.3

Depends R(>= 4.1), mnmer, Biostrings, caret

R topics documented:

<code>appmodels</code>	1
<code>predict_sequences</code>	2
<code>save_fasta</code>	3

Index	4
--------------	---

<code>appmodels</code>	<i>appmodels</i>
------------------------	------------------

Description

This function convert to a mnmer matrix and apply machine learning models

Usage

```
appmodels(mtx, m, n, len, cls)
```

Arguments

<code>mtx</code>	data.frame of mn-mers dataset
<code>m</code>	number of the mnmer
<code>n</code>	number of the mnmer
<code>len</code>	character of the length of the fragment (550, 1000, 3000, 5000, 10000 bp)
<code>cls</code>	character represents which class is applied class1 or class2

Value

dataframe

<code>predict_sequences</code>	<i>predict_sequences</i>
--------------------------------	--------------------------

Description

This function classifies sequences for the groups of arbovirus, mosquito-specific or othervirus

Usage

```
predict_sequences(seqs, cutoff = 0.5, all.data = TRUE)
```

Arguments

<code>seqs</code>	DNAStrngSet object to perform the classification
<code>cutoff</code>	is the probability limit for the soft voting of the models in ML
<code>all.data</code>	logical indicates if all sequence must be shown or only mosquito-specific and arbovirus

Value

dataframe

Examples

```
seqfile <- system.file ('extdata/PRJNA778885_known.fasta.gz',package="MosVir")
seqs <- readDNAStrngSet(seqfile)
res <- predict_sequences (seqs, cutoff=0.5, all.data=true)
```

save_fasta	<i>save_fasta</i>
------------	-------------------

Description

This function save a FASTA file of the predictions results

Usage

```
save_fasta(seqs, res, file, category = "all")
```

Arguments

seqs	DNAStringSet object for the FASTA sequences
res	data.frame with the results of the predict_sequences function
file	string for the output file to write a FASTA file
group	string of the groups: all (default), mosquito, mosquito-specific or arboviruses

Index

- * **arbovirus**
 - predict_sequences, 2
- * **classification**
 - appmodels, 1
 - predict_sequences, 2
 - save_fasta, 3
- * **learning**
 - appmodels, 1
 - save_fasta, 3
- * **machine**
 - appmodels, 1
 - save_fasta, 3
- * **models**
 - appmodels, 1
 - save_fasta, 3
- * **mosquito**
 - predict_sequences, 2
- * **softvoting**
 - appmodels, 1
 - save_fasta, 3

appmodels, 1

arbovirus (predict_sequences), 2

classes (appmodels), 1

classes (save_fasta), 3

classification (predict_sequences), 2

fasta (save_fasta), 3

format (save_fasta), 3

learning (appmodels), 1

machine (appmodels), 1

models (appmodels), 1

mosquito-specific (predict_sequences), 2

predict_sequences, 2

results (save_fasta), 3

save (save_fasta), 3

save_fasta, 3

soft (appmodels), 1

soft (save_fasta), 3

voting (appmodels), 1

voting (save_fasta), 3