

Package ‘seqmagick’

June 27, 2023

Title Sequence Manipulation Utilities

Version 0.1.6

Description

Supports reading and writing sequences for different formats (currently interleaved and sequential formats for 'FASTA' and 'PHYLIP'), file conversion, and manipulation (e.g. filter sequences that contain specify pattern, export consensus sequence from an alignment).

Depends R (>= 3.6.0)

Imports Biostrings, magrittr, utils

Suggests downloader, knitr, rmarkdown, GenomicAlignments, GenomicRanges, IRanges, muscle, Rsamtools, prettydoc

VignetteBuilder knitr

ByteCompile true

URL <https://github.com/YuLab-SMU/seqmagick>

BugReports <https://github.com/YuLab-SMU/seqmagick/issues>

License Artistic-2.0

Encoding UTF-8

RoxygenNote 7.2.3

NeedsCompilation no

Author Guangchuang Yu [aut, cre] (<<https://orcid.org/0000-0002-6485-8781>>)

Maintainer Guangchuang Yu <guangchuangyu@gmail.com>

Repository CRAN

Date/Publication 2023-06-27 05:10:02 UTC

R topics documented:

bam2DNAStrngSet	2
bam2DNAStrngSet2	3
bs_aln	3
bs_filter	4

bs_hamming	5
bs_rename	5
consensus	6
download_genbank	7
fas2phy	8
fa_combine	8
fa_filter	9
fa_read	10
fa_rename	10
fa_to_interleaved	11
fa_write	12
gb_read	13
get_id	13
phy2fas	14
phy_read	15
phy_write	15
renameTXT	16
replaceInside	17
seqlen	17
Index	19

bam2DNAStrngSet	<i>bam2DNAStrngSet</i>
-----------------	------------------------

Description

convert bam file to aligned fasta file

Usage

```
bam2DNAStrngSet(bamfile, refseq)
```

Arguments

bamfile	bam file
refseq	refseq, DNAStrngSet object

Value

DNAStrngSet object

Author(s)

Guangchuang Yu

bam2DNAStrngSet2	<i>bam2DNAStrngSet2</i>
------------------	-------------------------

Description

convert bam file to aligned fasta file

Usage

```
bam2DNAStrngSet2(bamfile, refseq)
```

Arguments

bamfile	bam file
refseq	refseq, DNAStrngSet object

Value

DNAStrngSet object

Author(s)

Guangchuang Yu

bs_aln	<i>bs_aln</i>
--------	---------------

Description

sequence alignment

Usage

```
bs_aln(x, method = "muscle", ...)
```

Arguments

x	XStringSet object
method	alignment method
...	additional parameter

Value

aligned sequences, XStringSet object

Author(s)

Guangchuang Yu

Examples

```
## Not run:  
fa_file <- system.file("extdata/HA.fas", package="seqmagick")  
x <- fa_read(fa_file)  
bs_aln(x)  
  
## End(Not run)
```

bs_filter

bs_filter

Description

biological sequence filter by searching pattern

Usage

```
bs_filter(x, pattern, by = "description", ignore.case = FALSE)
```

Arguments

x	BStringSet object
pattern	keyword for filter
by	one of 'description' and 'sequence'
ignore.case	logical

Value

BStringSet object

Author(s)

Guangchuang Yu

Examples

```
fa_file <- system.file("extdata/HA.fas", package="seqmagick")  
x <- fa_read(fa_file)  
bs_filter(x, 'ATGAAAGTAAAA', by='sequence')
```

bs_hamming	<i>bs_hamming</i>
------------	-------------------

Description

hamming distances of sequences

Usage

```
bs_hamming(x, count_indel = FALSE, ...)
```

Arguments

x	BStringSet object
count_indel	whether count indel or not
...	additional parameters

Value

hamming distance

Author(s)

Guangchuang Yu

Examples

```
## Not run:
fa_file <- system.file("extdata/HA.fas", package="seqmagick")
x <- fa_read(fa_file)
## align first 5 sequences, use `bs_aln(x)` to align all sequences
aln <- bs_aln(x[1:5])
bs_hamming(aln)

## End(Not run)
```

bs_rename	<i>bs_rename</i>
-----------	------------------

Description

rename sequence

Usage

```
bs_rename(x, mapping, position, sep, mode)
```

Arguments

x	BStringSet object
mapping	two column data.frame
position	rename token at specific position
sep	separator to divide token
mode	one of 'replace', 'prefix' or 'suffix'

Value

BStringSet

Author(s)

Guangchuang Yu

consensus

consensus

Description

consensus of aligned sequences

consensus of aligned sequences

Usage

```
consensus(x, type = "DNA")
```

```
bs_consensus(x, type = "DNA", r = 1)
```

Arguments

x	BStringSet object
type	currently, only DNA supported
r	if any NT > r, it will be selected as representative base

Value

consensus sequence string

consensus sequence string

Author(s)

Guangchuang Yu

Examples

```
## Not run:
fa_file <- system.file("extdata/HA.fas", package="seqmagick")
x <- fa_read(fa_file)
## align first 5 sequences, use `bs_aln(x)` to align all sequences
aln <- bs_aln(x[1:5])
## or bs_consensus(aln)
consensus(aln)

## End(Not run)
```

download_genbank *download_genbank*

Description

download genbank or fasta file by accession number

Usage

```
download_genbank(acc, db = "nuccore", format = "genbank", outfile = NULL, ...)
```

Arguments

acc	accession number(s)
db	supported db, currently 'nuccore'
format	one of 'genbank' or 'fasta'
outfile	output file, by default, acc.gb or acc.fa
...	additional parameters for download.file

Value

output file vector

Author(s)

Guangchuang Yu

Examples

```
## Not run:
tmpgb <- tempfile(fileext = '.gb')
tmpfa <- tempfile(fileext = '.fa')
download_genbank(acc='AB115403', format='genbank', outfile=tmpgb)
download_genbank(acc='AB115403', format='fasta', outfile=tmpfa)

## End(Not run)
```

fas2phy	<i>fas2phy</i>
---------	----------------

Description

convert fasta (aligned sequences) to phylip format

Usage

```
fas2phy(fasfile, outfile = "out.phy", type = "sequential")
```

Arguments

fasfile	aligned sequences in fasta format
outfile	output file
type	one of interleaved and sequential

Value

None

Author(s)

```
Guangchuang Yu fa_file <- system.file("extdata/HA.fas", package="seqmagick") phy_file <- temp-  
file(fileext = ".phy") fas2phy(fa_file, phy_file)
```

fa_combine	<i>fa_combine</i>
------------	-------------------

Description

combine 2 fasta files into 1

Usage

```
fa_combine(file1, file2, outfile = NULL, type = "interleaved")
```

Arguments

file1	fasta file 1
file2	fasta file 2
outfile	output file
type	one of interleaved and sequential

Value

BStringSet

Author(s)

Guangchuang Yu

Examples

```
fa_file <- system.file("extdata/HA.fas", package="seqmagick")
x <- fa_read(fa_file)
fa1 <- tempfile(fileext=".fa")
fa2 <- tempfile(fileext=".fa")
fa_write(x[1:5], fa1)
fa_write(x[6:10], fa2)
fa_combine(fa1, fa2)
```

`fa_filter`*fa_filter*

Description

fasta filter by searching pattern

Usage

```
fa_filter(
  fasfile,
  pattern,
  by = "description",
  ignore.case = FALSE,
  outfile = NULL,
  type = "interleaved"
)
```

Arguments

fasfile	input fasta file
pattern	keyword for filter
by	one of 'description' and 'sequence'
ignore.case	logical
outfile	output file
type	one of 'interleaved' and 'sequential'

Value

BStringSet object

Author(s)

Guangchuang Yu

fa_read	<i>fa_read</i>
---------	----------------

Description

read fasta file

Usage

```
fa_read(file, type = "auto")
```

Arguments

file	fasta file
type	one of 'DNA', 'RNA', 'AA', 'unknown' or 'auto'

Value

BStringSet object

Author(s)

Guangchuang Yu

Examples

```
fa_file <- system.file("extdata/HA.fas", package="seqmagick")
fa_read(fa_file)
```

fa_rename	<i>fa_rename</i>
-----------	------------------

Description

rename fasta sequence name

Usage

```
fa_rename(fasfile, mapping_file, position, sep, mode, outfile)
```

Arguments

fasfile	fasta file
mapping_file	mapping file
position	rename token at specific position
sep	separator to divide token
mode	one of 'replace', 'prefix' or 'suffix'
outfile	output file

Value

BStringSet object

Author(s)

Guangchuang Yu

fa_to_interleaved *fa_to_interleaved*

Description

convert fasta file to interleaved format
convert fasta file to sequential format

Usage

```
fa_to_interleaved(file, outfile)  
  
fa_to_sequential(file, outfile)
```

Arguments

file	fasta file
outfile	output file

Value

None
None

Author(s)

Guangchuang Yu

Examples

```
fa_file <- system.file("extdata/HA.fas", package="seqmagick")
fa1 <- tempfile(fileext = '.fa')
fa2 <- tempfile(fileext = '.fa')
fa_to_interleaved(fa_file, fa1)
fa_to_sequential(fa_file, fa2)
```

fa_write

fa_write

Description

write fasta file

Usage

```
fa_write(x, outfile, type = "interleaved")
```

Arguments

x	XStringSet object
outfile	output file
type	one of interleaved and sequential

Value

None

Author(s)

Guangchuang Yu

References

<http://evolution.genetics.washington.edu/phylip/doc/sequence.html>

Examples

```
phy_file <- system.file("extdata/HA.phy", package="seqmagick")
x <- phy_read(phy_file)
fa_file <- tempfile(fileext = '.fas')
fa_write(x, fa_file)
```

<i>gb_read</i>	<i>gb_read</i>
----------------	----------------

Description

extract accession number and sequence from genbank file

Usage

```
gb_read(file)
```

Arguments

file	input genbank file
------	--------------------

Value

sequence object

Author(s)

Guangchuang Yu

<i>get_id</i>	<i>get_id</i>
---------------	---------------

Description

get id at specific position

Usage

```
get_id(x, sep = " ", position)
```

Arguments

x	sequence description line
sep	separator to split x
position	id position

Value

id

Author(s)

Guangchuang Yu

Examples

```
fa_file <- system.file("extdata/HA.fas", package="seqmagick")
x <- fa_read(fa_file)
get_id(names(x)[1:5], sep = " ", position=1)
```

phy2fas

phy2fas

Description

convert phylip file to fasta file

Usage

```
phy2fas(phyfile, outfile = "out.fas", type = "interleaved")
```

Arguments

phyfile	phylip file
outfile	output file
type	one of interleaved and sequential

Value

None

Author(s)

Guangchuang Yu

Examples

```
phy_file <- system.file("extdata/HA.phy", package="seqmagick")
fa_file <- tempfile(fileext = '.fas')
phy2fas(phy_file, fa_file)
```

phy_read

phy_read

Description

read aligned sequences in phylip format

Usage

```
phy_read(file)
```

Arguments

file phylip file

Value

BStringSet object

Author(s)

Guangchuang Yu

Examples

```
phy_file <- system.file("extdata/HA.phy", package="seqmagick")
phy_read(phy_file)
```

phy_write

phy_write

Description

write phylip file

Usage

```
phy_write(x, outfile, type = "sequential")
```

Arguments

x XStringSet object
outfile output file
type one of interleaved and sequential

Value

None

Author(s)

Guangchuang Yu

Examples

```
## Not run:
fa_file <- system.file("extdata/HA.fas", package="seqmagick")
x <- fa_read(fa_file)
aln <- bs_aln(x[1:5])
phy_file <- tempfile(fileext = '.phy')
phy_write(aln, phy_file)

## End(Not run)
```

*renameTXT**renameTXT*

Description

rename txt file (eg Description line of fasta file) according to first token (eg accession number)

Usage

```
renameTXT(txt_file, name_file, sep = "_", split = TRUE)
```

Arguments

<code>txt_file</code>	txt file
<code>name_file</code>	name file
<code>sep</code>	separator
<code>split</code>	logical, split result or not

Value

None

Author(s)

Guangchuang Yu

replaceInside	<i>replaceInside</i>
---------------	----------------------

Description

replace character for example from '-' to 'N' of fasta sequence that only applied inside sequence any '-' character at start/end of the sequence (aligned seqs may contains '-' at prefix/suffix) will not be replaced

Usage

```
replaceInside(fasfile, from = "-", to = "N", outfile = NULL)
```

Arguments

fasfile	fasta file
from	character to be replaced, '-' by default
to	replace 'from' to 'to', 'N' by default
outfile	output file

Value

DNAStrngSet

Author(s)

Guangchuang Yu

seqLen	<i>seqLen</i>
--------	---------------

Description

sequence length

Usage

```
seqLen(fasfile)
```

Arguments

fasfile	fasta file
---------	------------

Value

numeric vector

Author(s)

Guangchuang Yu

Index

bam2DNAStrngSet, 2
bam2DNAStrngSet2, 3
bs_aln, 3
bs_consensus (consensus), 6
bs_filter, 4
bs_hamming, 5
bs_rename, 5

consensus, 6

download_genbank, 7

fa_combine, 8
fa_filter, 9
fa_read, 10
fa_rename, 10
fa_to_interleaved, 11
fa_to_sequential (fa_to_interleaved), 11
fa_write, 12
fas2phy, 8

gb_read, 13
get_id, 13

phy2fas, 14
phy_read, 15
phy_write, 15

renameTXT, 16
replaceInside, 17

seqlen, 17