Package: rchime (via r-universe)

July 23, 2026

Type Package

Title Detect and Remove Chimeras from Amplicon Sequence Analysis Data

Version 0.1.0

Date 2026-07-07

Maintainer Pat Schloss <pschloss@umich.edu>

Description Detect and remove chimeras from your amplicon sequence analysis using reference-based or de novo approaches. The 'rchime' package implements the 'VSEARCH' algorithms described in Rognes et al. (2016) <doi:10.7717/peerj.2584>. 'VSEARCH' builds on the work of Edgar,R.C. et al. (2011) <doi:10.1093/bioinformatics/btr381>.

URL https://github.com/mothur/rchime, https://mothur.org/rchime/

BugReports https://github.com/mothur/rchime/issues

License GPL (>= 3)

Config/testthat/edition 3

Imports cli, parallelly, RcppThread, utils

Depends Rcpp, R (>= 4.5.0), strollur(>= 0.1.3)

Suggests knitr, rmarkdown, xml2, pak, testthat (>= 3.0.0)

LinkingTo Rcpp, RcppThread, RcppXsimd, cli

Encoding UTF-8

VignetteBuilder knitr

Config/roxygen2/version 8.0.0

NeedsCompilation yes

Author Sarah Westcott [aut] (ORCID: <https://orcid.org/0009-0001-1529-8247>), Pat Schloss [cph, cre] (ORCID: <https://orcid.org/0000-0002-6935-4275>), Torbjorn Rognes [cph], Frederic Mahe [cph], Tomas Flouri [cph], Christopher Quince [cph], Ben Nichols [cph]

Repository https://cran.r-universe.dev

Date/Publication 2026-07-23 14:20:02 UTC

RemoteUrl https://github.com/cran/rchime

RemoteRef HEAD

RemoteSha 5bb9e1c29e4e11f0ac7fd041d2b5f4ca1224d55a

Contents

`rchime` *Detect and remove chimeras from your* R*hrefhttps://mothur.org/strollur/strollur object or data.frame using a de novo approach or alternatively a reference model.*

Description

The `rchime()` function allows you to detect and remove chimeras from your data using a de novo approach or alternatively a reference model.

Our preferred way of doing this is to use the abundant sequences as our reference (de novo).

This function uses code from the vsearch tools.

Usage

```
rchime(
  data,
  reference = NULL,
  dereplicate = TRUE,
  verbose = TRUE,
  remove_chimeras = TRUE,
  rchime_options = NULL,
 table_names = list(sequence_name = "sequence_name", sequence = "sequence", abundance =
    "abundance", sample = "sample")
)
```

Arguments

`data`	a strollur dataset object or a data.frame containing your sequence data.
`reference`	a strollur dataset object or a data.frame containing reference sequence data.
`dereplicate`	logical. The dereplicate option allows you to remove chimeras by sample. When `dereplicate=FALSE`, if a sequence is flagged as chimeric in one sample, it is removed from all samples. Our experience suggests that this is a bit aggressive since we've seen rare sequences get flagged as chimeric when they're the most abundant sequence in another sample. For a more conservative approach, we recommend using the default `dereplicate=TRUE` which will only remove sequences from the samples in which they are flagged as chimeric.
`verbose`	logical, allow console outputs. Default = `TRUE`.
`remove_chimeras`	Boolean, remove chimeras from dataset. Default = `TRUE`. Only used when `data` is a strollur object.
`rchime_options`	List, You can fine tune the vsearch specific options using the [`rchime_options()`] function. Default = `NULL`.
`table_names`	named list used to indicate the names of the columns in the data.frame. Only used when `data` or `reference` are a data.frames.

Value

list() containing a chimera report, and vector of the chimeric sequence's names.

The strollur dataset object will also be updated. The sequences flagged as chimeric are removed and the chimera report is added.

Author(s)

Sarah Westcott, <swestcot@umich.edu>

References

Rognes T, Flouri T, Nichols B, Quince C, Mahé F. (2016) VSEARCH: a versatile open source tool for metagenomics. PeerJ 4:e2584. doi: 10.7717/peerj.2584

Edgar,R.C., Haas,B.J., Clemente,J.C., Quince,C. and Knight,R. (2011), UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27:2194.

See Also

`rchime_options()` to set vsearch specific parameters.

Examples

```
# Let's use a strollur object with 500 sequences

data <- strollur::load_dataset(
  rchime_example("strollur_multi_sample_small.rds")
)
```

```
chimera_report <- rchime(data)
data
```

`rchime.data.frame` *Detect chimeras in your data.frames.*

Description

The `rchime()` function allows you to detect chimeras from your data using a de novo approach or alternatively a reference model.

Our preferred way of doing this is to use the abundant sequences as our reference (de novo).

This function uses code from the vsearch tools.

Usage

```
## S3 method for class 'data.frame'
rchime(
  data,
  reference = NULL,
  dereplicate = TRUE,
  verbose = TRUE,
  remove_chimeras = NULL,
  rchime_options = NULL,
 table_names = list(sequence_name = "sequence_name", sequence = "sequence", abundance =
    "abundance", sample = "sample")
)
```

Arguments

`data`	a data.frame containing your sequence data.
`reference`	a data.frame containing reference sequence data.
`dereplicate`	logical. The dereplicate option allows you to flag chimeras by sample. When `dereplicate=FALSE`, if a sequence is flagged as chimeric in one sample, it is removed from all samples. Our experience suggests that this is a bit aggressive since we've seen rare sequences get flagged as chimeric when they're the most abundant sequence in another sample. For a more conservative approach, we recommend using the default `dereplicate=TRUE` which will only remove sequences from the samples in which they are flagged as chimeric.
`verbose`	logical, allow console outputs. Default = `TRUE`.
`remove_chimeras`	Only used when `data` is a strollur object.
`rchime_options`	List, You can fine tune the vsearch specific options using the [`rchime_options()`] function. Default = NULL.

`table_names` named list used to indicate the names of the columns in the data.frame. Only used when `data` is a data.frame. By default:

table_names <- list(sequence_name = "sequence_name", sequence = "sequence" abundance = "abundance", sample = "sample")

In table_names, 'sequence_name' is a string containing the name of the column in 'table' that contains the sequence names. Default column name is 'sequence_name'.

In table_names, 'sequence' is a string containing the name of the column in 'table' that contains the sequences. Default column name is 'sequence'.

In table_names, 'abundance' is a string containing the name of the column in 'table' that contains the abundances. Default column name is 'abundance'.

In table_names, 'sample' is a string containing the name of the column in 'table' that contains the samples. Default column name is 'sample'.

Value

list() containing a chimera report, and vector of the chimeric sequence's names. If running with multiple samples and dereplicate = TRUE, then a table containing the modified sequence abundances will also be returned.

Author(s)

Sarah Westcott, <swestcot@umich.edu>

References

Rognes T, Flouri T, Nichols B, Quince C, Mahé F. (2016) VSEARCH: a versatile open source tool for metagenomics. PeerJ 4:e2584. doi: 10.7717/peerj.2584

Edgar,R.C., Haas,B.J., Clemente,J.C., Quince,C. and Knight,R. (2011), UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27:2194.

See Also

`rchime_options()` to set vsearch specific parameters.

Examples

```
# Detect chimeras from the dataset using de novo approach by sample
# (recommended)

data <- readRDS(rchime_example("miseq_data_frame_by_sample_small.rds"))

chimera_report <- rchime(data)
```

rchime.strollur *Detect and remove chimeras from your* R*hrefhttps://mothur.org/strollur/strollur dataset object.*

Description

The rchime() function allows you to detect and remove chimeras from your data using a de novo approach or alternatively a reference model.

Our preferred way of doing this is to use the abundant sequences as our reference (de novo).

This function uses code from the vsearch tools.

Usage

```
## S3 method for class 'strollur'
rchime(
  data,
  reference = NULL,
  dereplicate = TRUE,
  verbose = TRUE,
  remove_chimeras = TRUE,
  rchime_options = NULL,
  table_names = list(sequence_name = "sequence_name", sequence = "sequence")
)
```

Arguments

data	a strollur dataset object containing your sequence data.
reference	a strollur dataset object or a data.frame containing reference sequence data.
dereplicate	logical. The dereplicate option allows you to remove chimeras by sample. When dereplicate=FALSE, if a sequence is flagged as chimeric in one sample, it is removed from all samples. Our experience suggests that this is a bit aggressive since we've seen rare sequences get flagged as chimeric when they're the most abundant sequence in another sample. For a more conservative approach, we recommend using the default dereplicate=TRUE which will only remove sequences from the samples in which they are flagged as chimeric.
verbose	logical, allow console outputs. Default = TRUE.
remove_chimeras	logical, remove chimeras from dataset. Default = TRUE.
rchime_options	List, You can fine tune the vsearch specific options using the [rchime_options()] function. Default = NULL.
table_names	Only used when reference is a data.frame. By default: table_names <- list(sequence_name = "sequence_name", sequence = "sequence") In table_names, 'sequence_name' is a string containing the name of the column in 'table' that contains the sequence names. Default column name is 'sequence_name'.

In table_names, 'sequence' is a string containing the name of the column in 'table' that contains the sequences. Default column name is 'sequence'.

Value

list() containing a chimera report, and vector of the chimeric sequence's names.

The strollur dataset object will also be updated. The sequences flagged as chimeric are removed and the chimera report is added.

Author(s)

Sarah Westcott, <swestcot@umich.edu>

References

Rognes T, Flouri T, Nichols B, Quince C, Mahé F. (2016) VSEARCH: a versatile open source tool for metagenomics. PeerJ 4:e2584. doi: 10.7717/peerj.2584

Edgar,R.C., Haas,B.J., Clemente,J.C., Quince,C. and Knight,R. (2011), UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27:2194.

See Also

`rchime_options()` to set vsearch specific parameters.

Examples

```
# Let's load a strollur object with 500 sequences

data <- strollur::load_dataset(
  rchime_example("strollur_multi_sample_small.rds")
)

# Detect and remove chimeras from the dataset using de novo approach by
# sample (recommended)

chimera_report <- rchime(data)
data
```

`rchime_example` *Get path to rchime example files*

Description

rchime comes bundled with some example files in its `inst/extdata` directory. This function make them easy to access.

Usage

```
rchime_example(file = NULL)
```

Arguments

file	Name of file. If NULL, the example files will be listed.

Value

string, full path to file requested

Examples

```
rchime_example()
rchime_example("silva.gold.rds")
```

---

rchime_options *The* rchime_options() *function allows you to set vsearch specific optional parameters.*

Description

The rchime_options() function allows you to set parameters for the various chimera detection functions. With the exception of dereplicate, we recommend using the defaults, unless you have a compelling reason to do otherwise.

Usage

```
rchime_options(
  processors = parallelly::availableCores(),
  dereplicate = TRUE,
  abskew = 2,
  minh = 0.28,
  mindiv = 0.8,
  xn = 8,
  dn = 1.4,
  maxp = 3
)
```

Arguments

processors	Integer, number of cores to use. Default = all available
dereplicate	Boolean, The dereplicate option allows you to remove chimeras by sample. For example, if dereplicate parameter is FALSE, then if one group finds the sequence to be chimeric, it will be removed from all groups. If dereplicate is set to TRUE, sequences found to be chimeric are only removed from the sample they are found to be chimeric in. Default = TRUE.

abskew	Float, the minimum abundance skew (de novo only). Default = 2.0. abskew <- min (abund(parent1), abund(parent2)) / abund(query)
minh	Float, Mininum score to report chimera. Default 0.28. Values from 0.1 to 5 might be reasonable. Lower values increase sensitivity but may report more false positives. If you decrease –xn, you may need to increase –minh, and vice versa.
mindiv	Float, Minimum divergence ratio, default 0.8. Div ratio is 100%% %%identity between query sequence and the closest candidate for being a parent. If you don't care about very close chimeras, then you could increase –mindiv to, say, 1.0 or 2.0, and also decrease –min h, say to 0.1, to increase sensitivity. How well this works will depend on your data. Best is to tune parameters on a good benchmark.
xn	Float, Weight of a no vote, also called the beta parameter. Default 8.0. Decreasing this weight to around 3 or 4 may give better performance on denoised data.
dn	Float, Pseudo-count prior on number of no votes. Default 1.4. Probably no good reason to change this unless you can retune to a good benchmark for your data. Reasonable values are probably in the range from 0.2 to 2.
maxp	Integer, Maximum number of candidate parents to consider. Default 3.

Value

List containing selected rchime parameters and their values.

Author(s)

Sarah Westcott, <swestcot@umich.edu>

`silva_gold` *silva_gold*

Description

The `silva_gold()` function creates a silva based reference for use with the `rchime()` function in reference mode. It contains 5,181 sequences.

Usage

```
silva_gold()
```

Value

data.frame containing selected rchime parameters and their values.

Author(s)

Sarah Westcott, <swestcot@umich.edu>

Examples

```
reference <- silva_gold()
```

Index