Package ‘rfaRm’

October 11, 2024

Type Package

Title An R interface to the Rfam database

Version 1.17.0

Date 2024-01-25

Author Lara Selles Vidal, Rafael Ayala, Guy-Bart Stan, Rodrigo Ledesma-Amaro

Maintainer Lara Selles Vidal <lara.selles@oist.jp>,
Rafael Ayala <rafael.ayala@oist.jp>

Description rfaRm provides a client interface to the Rfam database of RNA families. Data that can be retrieved include RNA families, secondary structure images, covariance models, sequences within each family, alignments leading to the identification of a family and secondary structures in the dot-bracket format.

License GPL-3

Imports httr, stringi, rsvg, magick, data.table, Biostrings, utils, rvest, xml2, IRanges, S4Vectors, jsonlite

Suggests R4RNA, treeio, knitr, BiocStyle, rmarkdown, BiocGenerics, RUnit

VignetteBuilder knitr

biocViews FunctionalGenomics, DataImport, ThirdPartyClient, Visualization, MultipleSequenceAlignment

NeedsCompilation no

Encoding UTF-8

LazyData true

git_url https://git.bioconductor.org/packages/rfaRm

git_branch devel

git_last_commit 5de5040

git_last_commit_date 2024-04-30

Repository Bioconductor 3.20

Date/Publication 2024-10-10

Contents

rfamConsensusSecondaryStructure 2
rfamCovarianceModel 3
rfamFamilyAccessionToID 4
rfamFamilyIDToAccession 5
rfamFamilySummary 6
rfamPDBMapping 7
rfamSecondaryStructurePlot 8
rfamSecondaryStructureXMLSVG 9
rfamSeedAlignment 10
rfamSeedTree 11
rfamSeedTreeImage 12
rfamSequenceRegions 13
rfamSequenceSearch 15
rfamTextSearchFamilyAccession 16

Index **18**

`rfamConsensusSecondaryStructure`
Retrieve the secondary structure of an Rfam family

Description

Retrieves the consensus secondary structure and sequence of the specified Rfam family. The notation of the secondary structure can be specified through the "format" argument. If a filename is provided, the consensus secondary structure and sequence will be saved in the specified format.

Usage

```
rfamConsensusSecondaryStructure(rfamFamily, filename=NULL, format="DB")
```

Arguments

`rfamFamily`	string with an Rfam family accession or ID for which the consensus secondary structure and sequence should be retrieved.
`filename`	string indicating a path to which the retrieved consensus secondary structure and sequence should be written. The notation used for the RNA secondary structure is determined by the "format" argument. In the default behaviour, filename=NULL, and no file will be written. Instead, the consensus secondary structure and sequence will just be returned as a character vector.
`format`	string indicating the notation to be used for the RNA secondary structure. It can be either "DB" (extended Dot-Bracket notation; default) or "WUSS" (Washington University Secondary Structure notation).For a description of the different notations of RNA secondary structure, see https://www.tbi.univie.ac.at/RNA/ViennaRNA/doc/html/rna_structure_notations.html.

Value

A character vector where the first element is the consensus sequence of the Rfam family, and the second element is the consensus RNA secondary structure in the specified format.

References

Ioanna Kalvari, Joanna Argasinska, Natalia Quinones-Olvera, Eric P Nawrocki, Elena Rivas, Sean R Eddy, Alex Bateman, Robert D Finn, Anton I Petrov, Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families, Nucleic Acids Research, Volume 46, Issue D1, 4 January 2018, Pages D335–D342, https://doi.org/10.1093/nar/gkx1038

https://docs.rfam.org/en/latest/api.html

https://www.tbi.univie.ac.at/RNA/ViennaRNA/doc/html/rna_structure_notations.html

Examples

```
# Generate a character vector with the consensus secondary structure and
# sequence for the Rfam family with ID "FMN", with extended Dot-Bracket
# notation.

rfamConsensusSecondaryStructure("FMN", format="DB")

# Generate a character vector with the consensus secondary structure and
# sequence for the Rfam family with accession "RF00174", with extended
# Dot-Bracket notation, and save it to a file.

rfamConsensusSecondaryStructure("RF00174", filename="RF00174consensusSS.txt",
format="DB")

# Generate a character vector with the consensus secondary structure and
# sequence for the Rfam family with accession "RF00050", with WUSS notation, and
# save it to a file.

rfamConsensusSecondaryStructure("RF00050", filename="RF00050consensusSS.txt",
format="WUSS")
```

---

rfamCovarianceModel *Get the covariance model associated to an Rfam family*

Description

Gets the covariance model generated with the Infernal software of the Rfam family. Covariance models can be used in Infernal to search through a list of sequences for high-scoring hits to the models. If a filename is provided, the covariance model will be saved to the specified path.

Usage

```
rfamCovarianceModel(rfamFamily, filename=NULL)
```

Arguments

rfamFamily	string with an Rfam family accession or ID for which the covariance model should be retrieved.
filename	string indicating a path to which the covariance model should be written. In the default behaviour, filename=NULL, the covariance model will not be saved into any file, and instead will only be returned as a string.

Value

A string with the covariance model of the Rfam family in the Infernal format. If a filename is provided, the covariance model file will also be written to the specified path.

References

Ioanna Kalvari, Joanna Argasinska, Natalia Quinones-Olvera, Eric P Nawrocki, Elena Rivas, Sean R Eddy, Alex Bateman, Robert D Finn, Anton I Petrov, Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families, Nucleic Acids Research, Volume 46, Issue D1, 4 January 2018, Pages D335–D342, https://doi.org/10.1093/nar/gkx1038

Eric P Nawrocki, Sean R Eddy, Infernal 1.1: 100-fold faster RNA homology searches, Bioinformatics, Volume 29, Issue 22, 15 November 2013, Pages 2933–2935, https://doi.org/10.1093/bioinformatics/btt509

https://docs.rfam.org/en/latest/api.html

http://eddylab.org/infernal/Userguide.pdf

Examples

```
# Retrieve the covariance model for the Rfam family with ID "FMN" and save it
# to a file

rfamCovarianceModel("FMN", filename="FMNcovarianceModel.cm")

# Retrieve the covariance model for the Rfam family with accession "RF00174" and
# save it to a file

rfamCovarianceModel("RF00174", filename="RF00174covarianceModel.cm")
```

---

rfamFamilyAccessionToID

Convert an Rfam family accession to an Rfam family ID

Description

Gets the Rfam family ID corresponding to the provided Rfam family accession.

Usage

```
rfamFamilyAccessionToID(rfamFamilyAccession)
```

Arguments

`rfamFamilyAccession` string with the Rfam family accession to be converted to an Rfam family ID.

Value

A string with the Rfam family ID corresponding to the provided Rfam family accesion. The Rfam family ID usually describes the type of RNA associated to the Rfam family.

References

Ioanna Kalvari, Joanna Argasinska, Natalia Quinones-Olvera, Eric P Nawrocki, Elena Rivas, Sean R Eddy, Alex Bateman, Robert D Finn, Anton I Petrov, Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families, Nucleic Acids Research, Volume 46, Issue D1, 4 January 2018, Pages D335–D342, https://doi.org/10.1093/nar/gkx1038

https://docs.rfam.org/en/latest/api.html

Examples

```
# Get the Rfam family ID corresponding to the Rfam accession RF00050

rfamFamilyAccessionToID("RF00050")
```

---

rfamFamilyIDToAccession

Convert an Rfam family ID to an Rfam family accession

Description

Gets the Rfam family accession corresponding to the provided Rfam family ID.

Usage

```
rfamFamilyIDToAccession(rfamFamilyID)
```

Arguments

rfamFamilyID	string with the Rfam family ID to be converted to an Rfam family accession.

Value

A string with the Rfam family accession corresponding to the provided Rfam family ID.

References

Ioanna Kalvari, Joanna Argasinska, Natalia Quinones-Olvera, Eric P Nawrocki, Elena Rivas, Sean R Eddy, Alex Bateman, Robert D Finn, Anton I Petrov, Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families, Nucleic Acids Research, Volume 46, Issue D1, 4 January 2018, Pages D335–D342, https://doi.org/10.1093/nar/gkx1038

https://docs.rfam.org/en/latest/api.html

Examples

```
# Get the Rfam family accession corresponding to the Rfam family ID "FMN""

rfamFamilyIDToAccession("FMN")
```

`rfamFamilySummary` *Get a summary describing an Rfam family*

Description

Gets a summary describing an Rfam family. The summary includes information regarding the number of sequences and species contained in the family, a brief description about the function of the family and the corresponding type of RNA.

Usage

```
rfamFamilySummary(rfamFamily)
```

Arguments

`rfamFamily`	string with an Rfam family accession or ID for which a descriptive summary should be retrieved.

Value

A list containing the following elements that describe the Rfam family:

`rfamReleaseNumber`	Version of Rfam used in the query
`numberSequencesSeedAlignment`	Number of sequences used in the seed alignment of the Rfam family
`sourceSeedAlignment`	Published reference with the seed alignment of the Rfam family
`numberSpecies`	Number of species containing in their genomes sequences that belong to the Rfam family
`RNAType`	Keywords describing the type of RNA corresponding to the Rfam family
`numberSequencesAll`	Total number of sequences included in the Rfam family
`structureSource`	Description of the source for the secondary structure of the family (published, predicted) and PMID of the corresponding publication or prediction method if applicable
`description`	Descriptive name of the RNA included in the Rfam family
`rfamAccession`	Accession of the Rfam family
`rfamID`	ID of the Rfam family
`comment`	Short paragraph describing the characteristics and biological role of the Rfam family

References

Ioanna Kalvari, Joanna Argasinska, Natalia Quinones-Olvera, Eric P Nawrocki, Elena Rivas, Sean R Eddy, Alex Bateman, Robert D Finn, Anton I Petrov, Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families, Nucleic Acids Research, Volume 46, Issue D1, 4 January 2018, Pages D335–D342, https://doi.org/10.1093/nar/gkx1038

https://docs.rfam.org/en/latest/api.html

Examples

```
# Get a summary for the Rfam family with ID "FMN""

rfamFamilySummary("FMN")

# Get a summary for the Rfam family with accession "RF00174""

rfamFamilySummary("RF00174")

# Get a brief paragraph describing the Rfam family with accession "RF00174""

rfamFamilySummary("RF00174")$comment
```

`rfamPDBMapping` *Retrieve PDB entries with the 3D structure of members of an Rfam family*

Description

Retrieves entries of the PDB database with the experimentally solved 3D structure of members of the specified Rfam family, with correspondances between residues of the PDB structure and positions in the covariance model of the Rfam family. If a filename is provided, the list of PDB entries will be saved to the path specified through filename as a tab-delimited file.

Usage

```
rfamPDBMapping(rfamFamily, filename=NULL)
```

Arguments

`rfamFamily`	string with an Rfam family accession or ID for which the matching PDB entries should be retrieved.
`filename`	string indicating a path to which the retrieved PDB entries should be written as a tab-delimited file. In the default behaviour, filename=NULL, and no file will be written. Instead, the PDB entries will just be returned as a data frame.

Value

A data frame with PDB entries with experimentally solved 3D structures of members of the specified Rfam family, as well as correspondances between residues of the PDB structures and positions of the covariance model of the Rfam family. The data frame contains the following columns:

`Rfam accession`	Rfam accession of the Rfam family
`PDB ID`	ID of the PDB entry with a 3D structure of a member of the Rfam family
`Chain`	ID of the chain of the PDB entry that contains the coordinates for the 3D structure of a member of the Rfam family
`PDB start residue`	First residue of the 3D structure that can be matched to the sequence of a member of the Rfam family
`PDB end residue`	Last residue of the 3D structure that can be matched to the sequence of a member of the Rfam family

`CM start position`
: First position of the covariance model of the Rfam family that can be matched to the 3D structure

`CM end position`
: Last position of the covariance model of the Rfam family that can be matched to the 3D structure

`eValue`
: Expectation value for the alignment of the sequence of the 3D structure to the Rfam family

`Bit score`
: Bit score for the alignment of the sequence of the 3D structure to the Rfam family

References

Ioanna Kalvari, Joanna Argasinska, Natalia Quinones-Olvera, Eric P Nawrocki, Elena Rivas, Sean R Eddy, Alex Bateman, Robert D Finn, Anton I Petrov, Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families, Nucleic Acids Research, Volume 46, Issue D1, 4 January 2018, Pages D335–D342, https://doi.org/10.1093/nar/gkx1038

https://docs.rfam.org/en/latest/api.html

Examples

```
# Generate a data frame with the PDB structures of members of the Rfam family
# with ID "FMN"

rfamPDBMapping("FMN")

# Generate a data frame with the PDB structures of members of the Rfam family
# with accession "RF00174", and save the resulting PDB entries as a
# tab-delimited text file.

rfamPDBMapping("RF00174", filename="RF00174PDBentries.txt")
```

---

rfamSecondaryStructurePlot

Generate a diagram of the secondary structure of an Rfam family

---

Description

Plots a diagram of the specified type of the secondary structure of an Rfam family. If a filename is provided, the diagram will be saved to the specified path.

Usage

```
rfamSecondaryStructurePlot(rfamFamily, filename=NULL, plotType="norm")
```

Arguments

`rfamFamily`
: string with an Rfam family accession or ID for which the secondary structure should be plotted.

`filename`
: string indicating a path to which the plotted diagram should be written. In the default behaviour, filename=NULL, the diagram will not be saved into any file, and instead will only be plotted on R's graphics display. Must be a file type supported by magick, such as PNG, JPEG, GIF, RGB or RGBA.

`plotType` string indicating the desired type of secondary structure diagram. Can be one of "cons", "fcbp", "cov", "ent", "maxcm", "norm" (default), "rchie", "rscape" or "rscape-cyk". For a description of each type of diagram, see https://docs.rfam.org/en/latest/api.html#secondary-structure-images.

Value

A magick image object with the requested diagram. If a filename was specified, the diagram plot will also be saved to the specified path.

References

Ioanna Kalvari, Joanna Argasinska, Natalia Quinones-Olvera, Eric P Nawrocki, Elena Rivas, Sean R Eddy, Alex Bateman, Robert D Finn, Anton I Petrov, Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families, Nucleic Acids Research, Volume 46, Issue D1, 4 January 2018, Pages D335–D342, https://doi.org/10.1093/nar/gkx1038

https://docs.rfam.org/en/latest/api.html

Examples

```
# Plot a normal secondary structure diagram of the Rfam family with ID "FMN"

rfamSecondaryStructurePlot("FMN")

# Save a basepair conservation secondary structure diagram of the Rfam family
# with accession "RF00174" as a PNG file.

rfamSecondaryStructurePlot("RF00174", filename="RF00174bpcons.png", plotType="fcbp")
```

---

rfamSecondaryStructureXMLSVG

Get an SVG file with a secondary structure representation of an Rfam family

---

Description

Gets a string corresponding to an SVG file (XML format) with a representation of the secondary structure of the Rfam family, which can be read by rsvg or magick. If a filename is provided, the SVG file will also be written to the specified path.

Usage

```
rfamSecondaryStructureXMLSVG(rfamFamily, filename=NULL, plotType="norm")
```

Arguments

`rfamFamily` string with an Rfam family accession or ID for which a SVG file with a representation of the secondary structure should be retrieved.

`filename` string indicating a path to which the SVG graphics should be written. In the default behaviour, filename=NULL, the SVG will not be written to any file, and instead will only be returned as a character string, which can be directly read by R packages such as rsvg and magick.

`plotType` string indicating the desired type of secondary structure diagram. Can be one of "cons", "fcbp", "cov", "ent", "maxcm", "norm", "rchie", "rscape" or "rscape-cyk". For a description of each type of diagram, see https://docs.rfam.org/en/latest/api.html#secondary-structure-images.

Value

A string with the contents of a SVG file (XML format) with the representation of a secondary structure diagram of the Rfam family. The string can be directly read by rsvg or magick. If a filename is provided, the SVG file will also be written to the specified path.

References

Ioanna Kalvari, Joanna Argasinska, Natalia Quinones-Olvera, Eric P Nawrocki, Elena Rivas, Sean R Eddy, Alex Bateman, Robert D Finn, Anton I Petrov, Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families, Nucleic Acids Research, Volume 46, Issue D1, 4 January 2018, Pages D335–D342, https://doi.org/10.1093/nar/gkx1038

https://docs.rfam.org/en/latest/api.html

Examples

```
# Get a SVG representing a normal secondary structure diagram of the Rfam family
# with ID "FMN"

rfamSecondaryStructureXMLSVG("FMN")

# Get a SVG representing a basepair conservation secondary structure diagram
# for the Rfam family with accession "RF00174", and save it to a file

rfamSecondaryStructureXMLSVG("RF00174", filename="RF00174bpcons.svg", plotType="fcbp")
```

`rfamSeedAlignment` *Retrieve the seed alignment of an Rfam family*

Description

Retrieves the seed multiple alignment of the specified Rfam family. The seed alignment is used to determine the covariance model defining each Rfam family, and comprises only a subset of all the members of each family. If a filename is provided, the alignment will be saved in the specified format.

Usage

```
rfamSeedAlignment(rfamFamily, filename=NULL, format="stockholm")
```

Arguments

`rfamFamily` string with an Rfam family accession or ID for which the seed alignment should be retrieved.

`filename` string indicating a path to which the retrieved seed alignment should be written. The format of the written file is determined by the "format" argument. In the default behaviour, filename=NULL, and no file will be written. Instead, the seed alignment will just be returned as a Biostrings MultipleAlignment object.

`format` string indicating the format into which the seed alignment should be retrieved and written to filename Can be one of "stockholm" (default), "pfam", "fasta" or "fastau". For a description of each type of format, see https://docs.rfam.org/en/latest/api.html#stockholm-format-alignment.

Value

A Biostrings MultipleAlignment object with the aligned RNA sequences included in the seed alignment. The names of the sequences contain their GenBank accession numbers and the start and end positions of the region corresponding to the sequence region member of the Rfam family.

References

Ioanna Kalvari, Joanna Argasinska, Natalia Quinones-Olvera, Eric P Nawrocki, Elena Rivas, Sean R Eddy, Alex Bateman, Robert D Finn, Anton I Petrov, Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families, Nucleic Acids Research, Volume 46, Issue D1, 4 January 2018, Pages D335–D342, https://doi.org/10.1093/nar/gkx1038

https://docs.rfam.org/en/latest/api.html

Examples

```
# Generate a Biostrings MultipleAlignment object with the aligned RNA sequences
# used to generate the seed alignment for the Rfam family with ID "FMN"

rfamSeedAlignment("FMN")

# Generate a Biostrings MultipleAlignment object with the aligned RNA sequences
# used to generate the seed alignment for the Rfam family with accession
# "RF00174", and save it to a file in Stockholm format

rfamSeedAlignment("RF00174", filename="RF00174seedAlignment.stk", format="stockholm")

# Generate a Biostrings MultipleAlignment object with the aligned RNA sequences
# used to generate the seed alignment for the Rfam family with accession
# "RF00174", and save it to a file in FASTA format

rfamSeedAlignment("RF00174", filename="RF00174seedAlignment.fasta", format="fasta")
```

---

rfamSeedTree — *Get the phylogenetic tree of the seed multiple alignment of an Rfam family*

Description

Gets the phylogenetic tree of the seed multiple alignment associated to the Rfam family. The tree is retrieved in the NHX format (New Hampshire extended format. The tree will be saved to the path specified through filename. The generated file can be read directly with the treeio R package.

Usage

```
rfamSeedTree(rfamFamily, filename)
```

Arguments

`rfamFamily`	string with an Rfam family accession or ID for which the phylogenetic tree of the seed alignment should be retrieved.
`filename`	string indicating a path to which the phylogenetic tree in the NHX format should be written.

Value

A string with the phylogenetic tree of the seed alignment of the Rfam family in the NHX format. The tree will also be written to the specified path, producing a file that can be read by the treeio package.

References

Ioanna Kalvari, Joanna Argasinska, Natalia Quinones-Olvera, Eric P Nawrocki, Elena Rivas, Sean R Eddy, Alex Bateman, Robert D Finn, Anton I Petrov, Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families, Nucleic Acids Research, Volume 46, Issue D1, 4 January 2018, Pages D335–D342, https://doi.org/10.1093/nar/gkx1038

https://docs.rfam.org/en/latest/api.html

Examples

```
# Retrieve the phylogenetic tree of the seed alignment for the Rfam family with
# ID "FMN" and save it to a file

rfamSeedTree("FMN", filename="FMNseedTree.nhx")

# Retrieve the phylogenetic tree of the seed alignment for the Rfam family with
# accession "RF00174" and save it to a file

rfamSeedTree("RF00174", filename="RF00174seedTree.nhx")
```

`rfamSeedTreeImage` *Get an image of the phylogenetic tree of the seed multiple alignment of an Rfam family*

Description

Plots an image of the phylogenetic tree of the seed multiple alignment associated to the Rfam family. If a filename is provided, the image will be saved to the path specified through filename. The tree can be labeled with species names or sequence accessions.

Usage

```
rfamSeedTreeImage(rfamFamily, filename=NULL, label="species")
```

Arguments

`rfamFamily`	string with an Rfam family accession or ID for which the phylogenetic tree of the seed alignment should be plotted.
`filename`	string indicating a path to which the plot of the phylogenetic tree should be written.
`label`	string indicating the type of labels that should be added to the plot of the phylogenetic tree. Should be either "species" for labeling with species names (default) or "acc" for labeling with sequence accessions.

Value

A magick image object with the requested phylogenetic tree plot in GIF format. If a filename was specified, the phylogenetic tree plot will also be saved to the specified path.

References

Ioanna Kalvari, Joanna Argasinska, Natalia Quinones-Olvera, Eric P Nawrocki, Elena Rivas, Sean R Eddy, Alex Bateman, Robert D Finn, Anton I Petrov, Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families, Nucleic Acids Research, Volume 46, Issue D1, 4 January 2018, Pages D335–D342, https://doi.org/10.1093/nar/gkx1038

https://docs.rfam.org/en/latest/api.html

Examples

```
# Display a plot of the phylogenetic tree of the seed alignment for the Rfam
# family with ID "FMN" labeled with species names

rfamSeedTreeImage("FMN", label="species")

# Display a plot of the phylogenetic tree of the seed alignment for the Rfam
# family with accession "RF00147" labeled with sequence accessions and save it
# to a PNG file.

rfamSeedTreeImage("RF00147", filename="RF00147seedTreeImage.png", label="acc")
```

`rfamSequenceRegions` *Retrieve all sequence regions belonging to an Rfam family*

Description

Retrieves all sequence regions encoding an RNA assigned to be a member of the specified Rfam family. If a filename is provided, the list of sequence regions will be saved to the path specified through filename as a tab-delimited file.

Usage

```
rfamSequenceRegions(rfamFamily, filename=NULL)
```

Arguments

`rfamFamily` string with an Rfam family accession or ID for which the member sequence regions should be retrieved.

`filename` string indicating a path to which the retrieved sequence regions should be written as a tab-delimited file. In the default behaviour, filename=NULL, and no file will be written. Instead, the sequence regions will be returned as a data frame.

Value

A data frame with the sequence regions belonging to the specified Rfam family. The data frame contains the following columns:

`Sequence GenBank accession`
: GenBank accession of the sequence with the sequence region belonging to the Rfam family

`Bit score` Bit score for the alignment of the sequence region to the Rfam family

`Region start position`
: Position of the sequence specified by the GenBank accession where the sequence region belonging to the Rfam family starts

`Region end position`
: Position of the sequence specified by the GenBank accession where the sequence region belonging to the Rfam family ends

`Sequence description`
: Description of the sequence where the sequence region belonging to the Rfam family is found

`Species` Name of the species where the sequence region is present

`NCBI tax ID` NCBI taxonomy ID of the species where the sequence region is present

References

Ioanna Kalvari, Joanna Argasinska, Natalia Quinones-Olvera, Eric P Nawrocki, Elena Rivas, Sean R Eddy, Alex Bateman, Robert D Finn, Anton I Petrov, Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families, Nucleic Acids Research, Volume 46, Issue D1, 4 January 2018, Pages D335–D342, https://doi.org/10.1093/nar/gkx1038

https://docs.rfam.org/en/latest/api.html

Examples

```
# Generate a data frame with the sequence regions belonging to the Rfam family
# with ID "FMN"

head(rfamSequenceRegions("FMN"))

# Generate a data frame with the sequence regions belonging to the Rfam family
# with accession "RF00360", and save the sequence regions as a tab-delimited
# text file.

head(rfamSequenceRegions("RF00360", filename="RF00360sequenceRegions.txt"))
## Not run:
# Some Rfam families have too many associated sequence regions, in which case
# the list cannot be retrieved from the server.
```

```
rfamSequenceRegions("RF00174")

## End(Not run)
```

`rfamSequenceSearch` *Performs a sequence search of the Rfam database*

Description

Performs a search of the Rfam database by a provided RNA sequence, and retrieves high-scoring hits of Rfam families with different regions of the provided sequence.

Usage

```
rfamSequenceSearch(sequence, fragmentsOverlap=1000, clanCompetitionFilter=TRUE, clanOverlapThresh
```

Arguments

`sequence` string with an RNA sequence to be searched against the Rfam database. Should contain only standard RNA symbols (i.e., "A", "U", "G" and "C"), and no spaces or newlines.

`fragmentsOverlap` when a sequence larger than 10000 nucleotides is provided, it is internally split into smaller fragments before using them to search the Rfam database. This argument controls the number of overlapping bases between consecutive fragments.

`clanCompetitionFilter` logical indicating if results should be reduced through a clan competition filter, which removes overlapping hits if they belong to Rfam families of the same clan and have an overlap above a certain threshold.

`clanOverlapThreshold` number indicating the minimum overlap between two hits (as a fraction of the smallest hit) to remove the hit with the worst e-value if their families belong to the same Rfam clan.

Value

A nested list where each element of the top-level list represents a high-scoring hit with the Rfam families. Each of the top-level list elements is a list in itself, containing the following elements that describe the hit:

`rfamAccession` Rfam accession of the Rfam family with which a hit was found

`bitScore` Bit score for the hit with an Rfam family

`eValue` Expectation value for the with an Rfam family

`alignmentStartPositionQuerySequence` Start position in the query sequence of the sequence region that resulted in the hit

`alignmentEndPositionQuerySequence` End position in the query sequence of the sequence region that resulted in the hit

`alignmentStartPositionHitSequence`
: Start position in the Rfam family consensus sequence region with which the hit was found

`alignmentEndPositionHitSequence`
: End position in the Rfam family consensus sequence region with which the hit was found

`alignmentQuerySequence`
: Sequence region of the query RNA sequence provided to search the Rfam database, aligned with the corresponding region of the consensus sequence of the Rfam family with which the hit was found

`alignmentMatch`
: String describing the matches between the aligned regions of the query sequence and the consensus sequence of the Rfam family

`alignmentHitSequence`
: Sequence region of the consensus sequence of the Rfam family with which the hit was found, aligned to the corresponding region of the query RNA sequence

`alignmentSecondaryStructure`
: Secondary structure of the region of the consensus sequence of the Rfam family with which the hit was found

References

Ioanna Kalvari, Joanna Argasinska, Natalia Quinones-Olvera, Eric P Nawrocki, Elena Rivas, Sean R Eddy, Alex Bateman, Robert D Finn, Anton I Petrov, Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families, Nucleic Acids Research, Volume 46, Issue D1, 4 January 2018, Pages D335–D342, https://doi.org/10.1093/nar/gkx1038

https://docs.rfam.org/en/latest/api.html

https://www.tbi.univie.ac.at/RNA/ViennaRNA/doc/html/rna_structure_notations.html

Examples

```
# Search the Rfam database for hits with a specific sequence, and store the
# results in a nested list

searchHits <- rfamSequenceSearch("GGAUCUUCGGGGCAGGGUGAAAUUCCCGACCGGUGGUAUAGUCCAC
GAAAGUAUUUGCUUUGAUUUGGUGAAAUUCCAAAACCGACAGUAGAGUCUGGAUGAGAGAAGAUUC")

# Check number of high-scoring hits

length(searchHits)

# Extract the Rfam family accession and ID for the first hit

searchHits[[1]]$rfamAccession
searchHits[[1]]$rfamID
```

`rfamTextSearchFamilyAccession`
Search the Rfam database by a given keyword

Description

Searches the Rfam database for entries containing a specified keyword, such as "ribozyme" or "FMN", and returns the accession numbers of the Rfam families containing the keyword in its family ID, summary or description.

Usage

```
rfamTextSearchFamilyAccession(query)
```

Arguments

`query`	string with the keyword to be searched in the Rfam database.

Value

A character vector with the accessions of the Rfam families matching the search.

References

Ioanna Kalvari, Joanna Argasinska, Natalia Quinones-Olvera, Eric P Nawrocki, Elena Rivas, Sean R Eddy, Alex Bateman, Robert D Finn, Anton I Petrov, Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families, Nucleic Acids Research, Volume 46, Issue D1, 4 January 2018, Pages D335–D342, https://doi.org/10.1093/nar/gkx1038

https://docs.rfam.org/en/latest/api.html

Examples

```
# Search Rfam families associated to the keyword "FMN"

rfamTextSearchFamilyAccession("FMN")
```

Index

rfamConsensusSecondaryStructure, 2
rfamCovarianceModel, 3
rfamFamilyAccessionToID, 4
rfamFamilyIDToAccession, 5
rfamFamilySummary, 6
rfamPDBMapping, 7
rfamSecondaryStructurePlot, 8
rfamSecondaryStructureXMLSVG, 9
rfamSeedAlignment, 10
rfamSeedTree, 11
rfamSeedTreeImage, 12
rfamSequenceRegions, 13
rfamSequenceSearch, 15
rfamTextSearchFamilyAccession, 16