Package 'ncRNAtools'

May 6, 2024

Type Package

Title An R toolkit for non-coding RNA

Version 1.14.0

Date 2021-06-27

Description ncRNAtools provides a set of basic tools for handling and analyzing non-coding RNAs. These include tools to access the RNAcentral database and to predict and visualize the secondary structure of non-coding RNAs. The package also provides tools to read, write and interconvert the file formats most commonly used for representing such secondary structures.

License GPL-3

Imports httr, xml2, utils, methods, grDevices, ggplot2, IRanges, GenomicRanges, S4Vectors

Suggests knitr, BiocStyle, rmarkdown, RUnit, BiocGenerics

VignetteBuilder knitr

biocViews FunctionalGenomics, DataImport, ThirdPartyClient, Visualization, StructuralPrediction

NeedsCompilation no

Encoding UTF-8

LazyData true

BugReports https://github.com/LaraSellesVidal/ncRNAtools/issues

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

git_branch RELEASE_3_19

git_last_commit 693bd78

git_last_commit_date 2024-04-30

Repository Bioconductor 3.19

Date/Publication 2024-05-05

Author Lara Selles Vidal [cre, aut] (<https://orcid.org/0000-0003-2537-6824>), Rafael Ayala [aut] (<https://orcid.org/0000-0002-9332-4623>), Guy-Bart Stan [aut] (<https://orcid.org/0000-0002-5560-902X>), Rodrigo Ledesma-Amaro [aut] (<https://orcid.org/0000-0003-2631-5898>)

Maintainer Lara Selles Vidal <lara.selles@oist.jp>

Contents

`findPairedBases` *Determines paired bases from secondary structure string*

Description

Determines bases that form pairs in a given RNA sequence from the secondary structure string of the RNA sequence.

Usage

```
findPairedBases(secondaryStructureString, sequence)
```

Arguments

`secondaryStructureString`	A string representing the secondary structure of the RNA sequence in the Dot-Bracket format.
`sequence`	string with the RNA sequence corresponding to the provided secondary structure string. Should contain only standard RNA symbols (i.e., "A", "U", "G" and "C").

Value

A dataframe where each row denotes a base pair. The dataframe comprises the following 4 columns:

`Position1`	position in the sequence of the nucleotide involved in the base pair closest to the 5' end of the RNA.
`Position2`	position in the sequence of the nucleotide involved in the base pair closest to the 3' end of the RNA.

Nucleotide1	nucleotide type of the base in Position1.
Nucleotide2	nucleotide type of the base in Position2.

Examples

```
# Read a Dot Bracket file with the secondary structure of an RNA sequence:

exampleDotBracketFile <- system.file("extdata", "exampleDotBracket.dot",
                                     package="ncRNAtools")

exampleDotBracket <- readDotBracket(exampleDotBracketFile)

# Generate a dataframe of paired bases from the returned secondary structure
# string:

pairedBasesTable <- findPairedBases(secondaryStructureString=exampleDotBracket$secondaryStructure,
sequence=exampleDotBracket$sequence)
```

---

flattenDotBracket *Converts an extended Dot-Bracket secondary structure to basic Dot-Bracket notation*

Description

Generates a string with the secondary structure of an RNA sequence in the basic Dot-Bracket notation from a string in the extended Dot-Bracket notation.

Usage

```
flattenDotBracket(extendedDotBracketString)
```

Arguments

extendedDotBracketString	A string with a secondary structure representation of an RNA molecule in the extended Dot-Bracket notation. The extended Dot-Bracket notation uses dots (".") for unpaired bases, and multiple pairs of symbols ("("-")", "["-"]", ""-"", "<"-">", "A"-"a", "B"-"b", "C"-"c" and "D"-"d") to indicate paired bases. This allows the format to unambiguously represent highly nested structures, including pseudoknots.

Value

A string representing the secondary structure of the provided RNA in the basic Dot-Bracket format. In such format, dots (".") indicate unpaired bases, and paired bases are represented with pairs of round brackets ("("-")").

References

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

Examples

```
# A secondary structure string with characters other than dots and round
# brackets, representing unambiguously pseudo-knots:

extendedDotBracketString <- "((((....((((.(((((([[[[...)))))]]]]..))))).....))))."

# Convert it to the basic Dot-Bracket format:

basicDotBracketString <- flattenDotBracket(extendedDotBracketString)
```

`generatePairsProbabilityMatrix` *Generates a matrix of base pair probabilities*

Description

Generates a matrix of base pair probabilities from a table of base pair probabilities in the format returned by predictSecondaryStructure when using centroidFold or centroidHomFold as the prediction method.

Usage

```
generatePairsProbabilityMatrix(basePairProbsTable)
```

Arguments

`basePairProbsTable`
: A dataframe where each line corresponds to a nucleotide of the query RNA sequence. The first column indicates the position number, the second column indicates the corresponding nucleotide type and additional columns indicating the probability of forming a base pair with other nucleotides. The potentially pairing nucleotides and their corresponding probabilities should be provided as strings, with a colon separating both fields.

Value

A symmetric square matrix with a number of rows and columns equal to the number of nucleotides in the corresponding RNA sequence, determined by the number of rows of basePairProbsTable. The names of rows and columns are the nucleotide type and position. The value of each cell of the matrix is the probability that the nucleotides at positions given by the row and columns of the cell form a base pair.

Examples

```
# Load an example table of base pair probabilities, calculated with centroidFold:

basePairProbabilitiesTable <- read.csv(system.file("extdata",
"exampleBasePairProbabilitiesTable.csv", package="ncRNAtools"))
```

```
# Generate a base pair probability matrix from the returned base pair probability
# table:

basePairProbabilityMatrix <- generatePairsProbabilityMatrix(basePairProbabilitiesTable)
```

---

pairsToSecondaryStructure *Generates a string with the secondary structure of an RNA sequence*

Description

Generates a string with the secondary structure of an RNA sequence from a table of paired bases.

Usage

```
pairsToSecondaryStructure(pairedBases, sequence)
```

Arguments

pairedBases	A dataframe where each row contains the information of two bases that form a pair. The dataframe should contain columns named "Position1" and "Position2" indicating respectively the positions of the 5' and 3' bases involved in the base pair.
sequence	string with the RNA sequence corresponding to the provided table of paired bases. Should contain only standard RNA symbols (i.e., "A", "U", "G" and "C"), and no spaces or newlines.

Value

A string representing the secondary structure of the provided RNA in the Dot-Bracket format.

Examples

```
# Read a Dot Bracket file with the secondary structure of an RNA sequence:

exampleDotBracketFile <- system.file("extdata", "exampleDotBracket.dot",
                                     package="ncRNAtools")

exampleDotBracket <- readDotBracket(exampleDotBracketFile)

# Generate a dataframe of paired bases from the returned secondary structure
# string:

pairedBasesTable <- findPairedBases(secondaryStructureString=exampleDotBracket$secondaryStructure,
sequence=exampleDotBracket$sequence)

# Generate a secondary structure string from the table of paired bases:
```

```
secondaryStructureString <- pairsToSecondaryStructure(pairedBasesTable,
exampleDotBracket$sequence)

# Verify that the resulting secondary structure string is equal to the original
# prediction:

secondaryStructureString == exampleDotBracket$secondaryStructure
```

`plotCompositePairsMatrix` *Plots a composite matrix of base pair probabilities and paired bases*

Description

Generates a heatmap-like plot to visualize the probabilities that different bases of an RNA molecule form a pair together with bases that are paired in a given secondary structure for the same RNA. The top-right half of the plot represents base pair probabilities, while in the bottom-left half paired bases are represented with black squares in the corresponding cells. Usually, a correspondance between paired bases and high-probability base pairs is observed.

Usage

```
plotCompositePairsMatrix(basePairProbsMatrix, pairedBases, probabilityThreshold=0.1,
                         colorPalette=paste(rainbow(7, rev=TRUE), "FF", sep=""))
```

Arguments

`basePairProbsMatrix`	A symmetric square matrix containing the probabilities of pairs between different bases. Should be in the same format as output by the generatePairsProbabilityMatrix function.
`pairedBases`	A dataframe indicating paired bases, in the same format as output by the findPairedBases function.
`probabilityThreshold`	Threshold for representing the probability that two given bases form a pair. Pairs with a probability lower than the threshold will not be considered, and their corresponding cell in the plot will be left blank.
`colorPalette`	Color palette to be used for displaying the probabilities above the specified threshold.

Value

A ggplot object with a representation of the base pair probability matrix and bases paired in a given secondary structure.

Examples

```
# Create a list with an RNA sequence, its secondary structure and a table of
# base pair probabilities, calculated with centroidFold:

secondaryStructure <- list(sequence="GGGGAUGUAGCUCAUAUGGUAGAGCGCUCGCUUUGCAUGCGAGAGGCACAGGGUUCGAUUCCCUGCAUCUCCA"
secondaryStructure="(((((((..((((........))))).(((((.......)))))).....(((((.......)))))))))))).",
basePairProbabilities=read.csv(system.file("extdata", "exampleBasePairProbabilitiesTable.csv", package="ncRNAto

# Generate a matrix of base pair probabilities:

probabilitiesMatrix <- generatePairsProbabilityMatrix(secondaryStructure$basePairProbabilities)

# Generate a dataframe with paired bases:

pairedBases <- findPairedBases(secondaryStructure$secondaryStructure, secondaryStructure$sequence)

# Plot base pair probabilities and paired bases:

plotCompositePairsMatrix(probabilitiesMatrix, pairedBases)
```

---

plotPairsProbabilityMatrix

Plots a matrix of base pair probabilities

Description

Generates a heatmap-like plot to visualize the probabilities that different bases of an RNA molecule form a pair.

Usage

```
plotPairsProbabilityMatrix(basePairProbsMatrix, probabilityThreshold=0.1,
                           colorPalette=paste(rainbow(7, rev=TRUE), "FF", sep=""))
```

Arguments

basePairProbsMatrix	A symmetric square matrix containing the probabilities of pairs between different bases. Should be in the same format as output by the generatePairsProbabilityMatrix function.
probabilityThreshold	Threshold for representing the probability that two given bases form a pair. Pairs with a probability lower than the threshold will not be considered, and their corresponding cell in the plot will be left blank.
colorPalette	Color palette to be used for displaying the probabilities above the specified threshold.

Value

A ggplot object with a representation of the base pair probability matrix.

Examples

```
# Load an example table of base pair probabilities, calculated with centroidFold:

basePairProbabilitiesTable <- read.csv(system.file("extdata",
"exampleBasePairProbabilitiesTable.csv", package="ncRNAtools"))

# Generate a matrix of base pair probabilities:

probabilitiesMatrix <- generatePairsProbabilityMatrix(basePairProbabilitiesTable)

# Plot the probability matrix

plotPairsProbabilityMatrix(probabilitiesMatrix)
```

---

predictAlternativeSecondaryStructures

Predicts alternative secondary structures of a given RNA sequence

---

Description

Attempts to identify potential alternative secondary structures of the provided RNA sequence using the RintW method, based on the decomposition of the base-pairing probability matrix over the Hamming distance to a reference secondary structure. It should be noted that RintW runs can take considerable amounts of time.

Usage

```
predictAlternativeSecondaryStructures(sequence, gammaWeight=4, inferenceEngine="BL")
```

Arguments

`sequence`	string with an RNA sequence whose secondary structure should be predicted. Should contain only standard RNA symbols (i.e., "A", "U", "G" and "C").
`gammaWeight`	weight factor for predicted base pairs. It directly affects the number of predicted base pairs. A higher value leads to a higher number of base pairs predicted. It should be a positive number. In the default behavior, a value of 4 is used.
`inferenceEngine`	engine used to identify the optimal canonical secondary structure. Possible values are "BL", "Turner" and "CONTRAfold". In the first two cases, a McCaskill partition function is applied, using respectively the Boltzmann likelihood model or Turner's energy model. In the third case, the CONTRAfold engine, based on conditional log-linear models, is applied. In the default behavior, a McCaskill partition function with a Boltzmann likelihood model is used.

Value

A list of two-element lists, where each element of the upper level list represents a potential secondary structure. The first top-level element always represents the canonical secondary structure. If no alternative secondary structures are found, simply a list of two elements is returned, comprising the query sequence and the canonical secondary structure.

When alternative secondary structure elements are found, each top-level element comprises the following two elements:

`sequence` Query RNA sequence

`secondaryStructure`
Predicted secondary structure

References

Andronescu M, Condon A, Hoos HH, Mathews DH, Murphy KP. Computational approaches for RNA energy parameter estimation. RNA. 2010;16(12):2304-2318. doi:10.1261/rna.1950510

Do CB, Woods DA, Batzoglou S. CONTRAfold: RNA secondary structure prediction without physics-based models. Bioinformatics. 2006;22(14):e90-e98. doi:10.1093/bioinformatics/btl246

Hagio T, Sakuraba S, Iwakiri J, Mori R, Asai K. Capturing alternative secondary structures of RNA by decomposition of base-pairing probabilities. BMC Bioinformatics. 2018;19(Suppl 1):38. Published 2018 Feb 19. doi:10.1186/s12859-018-2018-4

Hamada M, Ono Y, Kiryu H, et al. Rtools: a web server for various secondary structural analyses on single RNA sequences. Nucleic Acids Res. 2016;44(W1):W302-W307. doi:10.1093/nar/gkw337

Mathews DH, Sabina J, Zuker M, Turner DH. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J Mol Biol. 1999;288(5):911-940. doi:10.1006/jmbi.1999.2700

http://rtools.cbrc.jp/

Examples

```
# Predict alternative secondary structures of an RNA sequence:

alternativeStructures <- predictAlternativeSecondaryStructures("AAAGGGGUUUCCC")

# Count the number of potential alternative structures identified:

length(alternativeStructures)
```

---

predictSecondaryStructure

Predicts the secondary structure of a given RNA sequence

Description

Predicts the secondary structure of the provided RNA sequence, using the chosen prediction method. Secondary structure predictions are carried out with the rtools RNA Bioinformatics Web server.

Usage

```
predictSecondaryStructure(sequence, method, gammaWeight=NULL, inferenceEngine=NULL,
alignmentEngine=NULL, eValueRfamSearch=NULL, numHomSeqsRfamSearch=NULL)
```

Arguments

sequence	string with an RNA sequence whose secondary structure should be predicted. Should contain only standard RNA symbols (i.e., "A", "U", "G" and "C").
method	method that should be used for the prediction of RNA secondary structure. Possible values are "centroidFold", "centroidHomFold" and "IPknot". Only IPknot is able to predict pseudoknots. For a detailed description of each method, see respectively *Hamada et al., 2008*; *Hamada et al., 2009*, and *Sato et al., 2011*.
gammaWeight	weight factor for predicted base pairs. It directly affects the number of predicted base pairs. A higher value leads to a higher number of base pairs predicted. It should be a positive number. In the default behavior, when no specific value is provided, the default value for each secondary structure prediction method in the rtools webserver is used (4 for centroidFold and IPknot, and 8 forcentroidHomFold).
inferenceEngine	engine used to identify optimal secondary structures. Possible values are "BL", "Turner" and "CONTRAfold". In the first two cases, a McCaskill partition function is applied, using respectively the Boltzmann likelihood model or Turner's energy model. In the third case, the CONTRAfold engine, based on conditional log-linear models, is applied. Additionally, if IPknot is chosen as the method for secondary structure prediction, "NUPACK" is also a possible value if the sequence has 100 nucleotides or less. In this case, the NUPACK scoring model is used. In the default behavior, when no specific value is provided, the default inference engine in the rtools webserver is, which is a McCaskill partition function with a Boltzmann likelihood model.
alignmentEngine	engine used to perform pairwise alignments of the query sequence and Rfam homologous sequences during the application of centroidHomFold. Possible values are "CONTRAlign" and "ProbCons". For details on each alignment engine, see *Do et al., 2006* and *Do et al., 2005* respectively. In the default behavior, when no value is specified, CONTRAlign is used.
eValueRfamSearch	e-value used to select homologous sequences from the Rfam database during the application of centroidHomFold. Should be a number equal to or greater than 0. In the default behavior, when no value is specified, a value of 0.01 is used.
numHomSeqsRfamSearch	maximum number of homologous sequences to be considered during the application of centroidHomFold. Should be a positive integer.In the default behavior, when no value is specified, a value of 30 is used.

Value

If either centroidFold or centroidHomFold are used for predicting secondary structure, a list of three elements comprising the query RNA sequence, the predicted secondary structure and a table of base

pair probabilities. The secondary structure is represented as a string in the Dot-Bracket format. The three elements of the list are:

`sequence` Query RNA sequence

`secondaryStructure`
Predicted secondary structure

`basePairProbabilities`
A dataframe where each line corresponds to a nucleotide of the query RNA sequence. The first column indicates the position number, the second column indicates the corresponding nucleotide type and additional columns indicating the probability of forming a base pair with other nucleotides. The potentially pairing nucleotides and their corresponding probabilities are provided as strings, where a colon separates both fields

If IPknot is used for predicting secondary structure, no table of base pair probabilities is returned. Therefore, the output is a list of only two elements (sequence and secondaryStructure). Additionally, the secondary structure is provided in the extended Dot-Bracket format if required to represent pseudoknots unambiguously.

References

Andronescu M, Condon A, Hoos HH, Mathews DH, Murphy KP. Computational approaches for RNA energy parameter estimation. RNA. 2010;16(12):2304-2318. doi:10.1261/rna.1950510

Do C.B., Gross S.S., Batzoglou S. CONTRAlign: Discriminative Training for Protein Sequence Alignment. In: Apostolico A., Guerra C., Istrail S., Pevzner P.A., Waterman M. (eds) Research in Computational Molecular Biology. 2006. Lecture Notes in Computer Science, vol 3909. Springer, Berlin, Heidelberg doi:10.1007/11732990_15

Do CB, Mahabhashyam MS, Brudno M, Batzoglou S. ProbCons: Probabilistic consistency-based multiple sequence alignment. Genome Res. 2005;15(2):330-340. doi:10.1101/gr.2821705

Do CB, Woods DA, Batzoglou S. CONTRAfold: RNA secondary structure prediction without physics-based models. Bioinformatics. 2006;22(14):e90-e98. doi:10.1093/bioinformatics/btl246

Hamada M, Kiryu H, Sato K, Mituyama T, Asai K. Prediction of RNA secondary structure using generalized centroid estimators. Bioinformatics. 2009;25(4):465-473. doi:10.1093/bioinformatics/btn601

Hamada M, Ono Y, Kiryu H, et al. Rtools: a web server for various secondary structural analyses on single RNA sequences. Nucleic Acids Res. 2016;44(W1):W302-W307. doi:10.1093/nar/gkw337

Hamada M, Yamada K, Sato K, Frith MC, Asai K. CentroidHomfold-LAST: accurate prediction of RNA secondary structure using automatically collected homologous sequences. Nucleic Acids Res. 2011;39(Web Server issue):W100-W106. doi:10.1093/nar/gkr290

Mathews DH, Sabina J, Zuker M, Turner DH. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J Mol Biol. 1999;288(5):911-940. doi:10.1006/jmbi.1999.2700

Sato K, Kato Y, Hamada M, Akutsu T, Asai K. IPknot: fast and accurate prediction of RNA secondary structures with pseudoknots using integer programming. Bioinformatics. 2011;27(13):i85-i93. doi:10.1093/bioinformatics/btr215

http://rtools.cbrc.jp/

Examples

```
# Predict the secondary structure of an RNA sequence with IPknot:

structurePrediction <- predictSecondaryStructure("UGCGAGAGGCACAGGGUUCGAUUCCCUGCA
UCUCCA", "IPknot", inferenceEngine = "NUPACK")

# Extract the string representing the secondary structure prediction:

structurePrediction$secondaryStructure
```

`readCT` *Reads a file with the secondary structure of an RNA in the CT format*

Description

Reads a file with the secondary structure of an RNA in the CT format (Connectivity Table).

Usage

```
readCT(filename, sequence=NULL)
```

Arguments

`filename`	A string indicating the path to the CT file to be read. A description of the CT format can be found at http://rna.urmc.rochester.edu/Text/File_Formats.html#CT. The CT file should contain information for all nucleotides of the RNA sequence, including those that do not form base pairs, unless the full-length RNA sequence is provided through the sequence argument. In such case, the CT file may contain only information for nucleotides involved in base pairs.
`sequence`	A string with the full-length sequence of the RNA whose secondary structure is represented in the CT file. Such argument is optional, but if it is not provided, a complete CT file with information for all nucleotides of the RNA must be provided.

Value

A list with the following 4 elements:

`sequenceName`	name of the RNA sequence
`sequence`	RNA sequence
`sequenceLength`	number of nucleotides of the RNA sequence
`pairsTable`	a dataframe indicating the paired bases, in the same format as that returned by the findPairedBases function

References

http://rna.urmc.rochester.edu/Text/File_Formats.html#CT.

Examples

```
exampleCTFile <- system.file("extdata", "exampleCT.ct", package="ncRNAtools")

# If the CT file does not contain information for unpaired nucleotides, the
# original sequence must also be supplied in order to read it (tmRNA of E. coli
# encoded by the ssrA gene):

tmRNASequence <- "GGGGCUGAUUCUGGAUUCGACGGGAUUUGCGAAACCCAAGGUGCAUGCCGAGGGGCGGUUGG
CCUCGUAAAAAGCCGCAAAAAAUAGUCGCAAACGACGAAAACUACGCUUUAGCAGCUUAAUAACCUGCUUAGAGCCCUCU
CUCCCUAGCCUCCGCUCUUAGGACGGGGAUCAAGAGAGGUCAAACCCAAAAGAGAUCGCGUGGAAGCCCUGCCUGGGGUU
GAAGCGUUAAAACUUAAUCAGGCUAGUUUGUUAGUGGCGUGUCCGUCCGCAGCUGGCAAGCGAAUGUAAAGACUGACUAA
GCAUGUAGUACCGAGGAUGUAGGAAUUUCGGACGCGGGUUCAACUCCCGCCAGCUCCACCA"

tmRNASecondaryStructure <- readCT(exampleCTFile, tmRNASequence)
```

`readDotBracket` *Reads a file with the sequence and secondary structure of an RNA in the Dot-Bracket format*

Description

Reads a file with the secondary structure of an RNA in the CT format (Connectivity Table).

Usage

```
readDotBracket(filename)
```

Arguments

`filename`	A string indicating the path to the Dot-Bracket file to be read. A description of the Dot-Bracket format can be found at https://www.tbi.univie.ac.at/RNA/ViennaRNA/doc/html/rna_struc Both basic and extended Dot-Bracket notations are accepted. The Dot-Bracket file should always contain at least two lines: the first one with the RNA sequence, and the second one with the secondary structure representation. The free energy of the secondary structure can be optionally present at the end of the secondary structure line, in kcal/mol. An additional optional line can precede these two, containing ">" as its first character followed by the name or description of the sequence.

Value

A list with the following 4 elements:

`sequenceName`	name of the RNA sequence. Will be set to NA if no string starting with ">" is present at the beginning of the Dot-Bracket file
`sequence`	RNA sequence

`secondaryStructure`
: string representing the secondary structure of the RNA in the Dot-Bracket format

`freeEnergy`
: free energy of the secondary structure of the RNA, usually in kcal/mol. If not present in the Dot-Bracket file, it will be set to NA

References

https://software.broadinstitute.org/software/igv/RNAsecStructure

http://projects.binf.ku.dk/pgardner/bralibase/RNAformats.html

Examples

```
exampleDotBracketFile <- system.file("extdata", "exampleDotBracket.dot",
                                     package="ncRNAtools")

exampleDotBracket <- readDotBracket(exampleDotBracketFile)

# Since the file does not contain an initial line with the name of the sequence,
# sequenceName is set to NA
```

---

rnaCentralGenomicCoordinatesSearch

Retrieves annotated non-coding RNA in a set of genomic ranges

---

Description

Retrieves RNAcentral entries corresponding to non-coding RNA present within a specified set of genomic coordinates

Usage

```
rnaCentralGenomicCoordinatesSearch(genomicRanges, species)
```

Arguments

`genomicRanges`
: GRanges object specifying the genomic coordinates for which known non-coding RNA should be retrieved. Each sequence name must be of the format chr? or simply ?, where ? can be any number or the "X" or "Y" characters. These denote, respectively, standard somatic chromosomes and sexual chromosomes. Additionally, sequence name can also be "MT", which refers to the mitochondrial DNA. Multiple ranges can be provided in the same GRanges object, but they must all belong to the same organism.

`species`
: string specifying the species to which the ranges supplied with the genomicRanges argument refer. The scientific name should be provided. The list of species for which search of RNAcentral by genomic coordinates is supported can be found at https://rnacentral.org/help/genomic-mapping.

Value

A nested list whose length matches the number of genomic ranges provided through genomicRanges. Each top-level element is a list containing all hits found for a given genomic range. In turn, each of the elements of such list representing a hit is a list comprising the following elements that describe the hit:

`rnaCentralID`	RNAcentral ID of the non-coding RNA that was identified as a hit
`species`	species whose genome was searched
`description`	brief description of the RNA sequence of the hit
`RNAType`	category of the RNA sequence of the hit
`genomicCoordinates`	GRanges object containing a single range, within which the hit was found

References

https://rnacentral.org/help/genomic-mapping

https://rnacentral.org/api

Examples

```
# Generate a GRanges object with 2 genomic ranges specifying coordinates of
# the human genome:

genomicCoordinates <- GenomicRanges::GRanges(seqnames=S4Vectors::Rle(c("chr3", "chr4")),
ranges=IRanges::IRanges(rep(39745816, 2), rep(39847679, 2)))

# Retrieve known annotated non-coding RNA present in the specified genomic
# ranges:

knownNonCodingRNA <- rnaCentralGenomicCoordinatesSearch(genomicCoordinates, "Homo sapiens")
```

---

rnaCentralRetrieveEntry

Retrieves information for a specific entry of the RNAcentral database

---

Description

Retrieves information for a specific entry of the RNAcentral database. The retrieved information includes the corresponding sequence, the sequence length, a brief description of the sequence, the species where the sequence is found, the NCBI taxid of the species and the RNA type.

Usage

```
rnaCentralRetrieveEntry(rnaCentralID)
```

Arguments

rnaCentralID	string indicating the RNAcentral ID of the entry that should be retrieved. It must start with "URS".

Value

A list containing the following elements that correspond to the RNAcentral entry associated with the provided RNAcentral ID:

rnaCentralID	Query RNAcentral ID
sequence	RNA sequence of the retrieved entry
sequenceLength	length of the RNA sequence of the retrieved entry
description	short description of the retrieved entry
species	species where the sequence of the retrieved entry is found
ncbiTaxID	NCBI taxonomy identifier of the species corresponding to the retrieved entry
RNATypes	categories of RNA associated to the RNA sequence of the retrieved entry

References

https://rnacentral.org/about-us

https://rnacentral.org/api

Examples

```
# Retrieve information of the RNAcentral entry associated with ID URS00007C2D83_224308:

rnaCentralEntry <- rnaCentralRetrieveEntry("URS00007C2D83_224308")

# Extract the corresponding RNA sequence

rnaCentralEntry$sequence
```

---

rnaCentralTextSearch *Searches the RNAcentral database with a text query*

Description

Searches the RNAcentral database of non-coding RNA with a text query.

Usage

```
rnaCentralTextSearch(query)
```

Arguments

query	string to be used as the text query to search the RNAcentral database. The most basic usage involves providing a word or group of words to be matched against any of the fields of entries of the RNAcentral database. More refined searches can be done by using the field_name:"field value" syntax. Several search terms can be combined with logical operators (and, or, not). Asterisks ("*") can be added to signal any number of any characters. If not added to the query, only exact matches are produced. A detailed description of the syntax can be found at https://rnacentral.org/help/text-search. The entire query should be provided as a single string, and double quotes enclosing field values must be escaped with a backslash (“\“). It should also be noted that logical operators should be capitalized.

Value

A character vector where each element is the RNAcentral ID of an entry that matched the query. A maximum of 15 IDs will be returned.

References

https://rnacentral.org/about-us

https://rnacentral.org/help/text-search

Examples

```
# Find RNAcentral entries that correspond to the HOTAIR long noncoding RNA:

rnaCentralTextSearch("HOTAIR")

# Find RNAcentral entries that correspond to FMN riboswitches in Bacillus subtilis:

rnaCentralTextSearch("FMN AND species:\"Bacillus subtilis\"")
```

---

writeCT *Writes a file with the secondary structure of an RNA in the CT format*

Description

Writes a file with the secondary structure of an RNA in the CT format (Connectivity Table).

Usage

```
writeCT(filename, sequence, secondaryStructure=NULL, sequenceName="Sequence",
        pairedBases=NULL)
```

Arguments

filename	A string indicating the path to the CT file to be written. A description of the CT format can be found at http://rna.urmc.rochester.edu/Text/File_Formats.html#CT.
sequence	A string with the full-length sequence of the RNA for which a CT file should be written.
secondaryStructure	A string representing the secondary structure of the RNA sequence in the Dot-Bracket format. A correct value must be provided for either secondaryStructure or pairedBases. If a table of paired bases is provided through pairedBases, the string provided through secondaryStructure will be ignored. If no table of paired bases is provided, one will be internally calculated from the string representation of the secondary structure.
sequenceName	A string with the name of the sequence to be written in the first line of the CT file.
pairedBases	A table of paired bases in the same format as output by the findPairedBases function.

Value

Called for its effect of writing a CT file. Invisibly returns the status code returned by close when closing the file connection. See documentation of close for details.

References

http://rna.urmc.rochester.edu/Text/File_Formats.html#CT.

Examples

```
# Write a CT file by providing an RNA sequence and its secondary structure in
# the Dot-Bracket format:

tempDir <- tempdir()
CTFile <- paste(tempDir, "testCTFile.ct", sep="")
writeCT(CTFile, "AGCGGGUUCCUGGUUCCCCAAGGUUGA",
secondaryStructure="...(((..((.))..))).((..)).."
sequenceName="Test sequence")
```

---

writeDotBracket *Writes a file with the sequence and secondary structure of an RNA in the Dot-Bracket format*

Description

Writes a file with the sequence and secondary structure of an RNA in the Dot-Bracket format.

Usage

```
writeDotBracket(filename, sequence, secondaryStructure, sequenceName="Sequence")
```

Arguments

filename	A string indicating the path to the Dot-Bracket file to be written. A description of the Dot-Bracket format can be found at https://www.tbi.univie.ac.at/RNA/ViennaRNA/doc/html/rna_struc The resulting file will contain three lines. The first line starts with ">", after which the name of the sequence follows. The second line contains the RNA sequence. The third line contains the secondary structure representation in Dot-Bracket notation.
sequence	A string with the full-length sequence of the RNA for which a Dot-Bracket file should be written.
secondaryStructure	A string representing the secondary structure of the RNA sequence in the Dot-Bracket format.
sequenceName	A string with the name of the RNA sequence for which a Dot-Bracket file should be written.

Value

Called for its effect of writing a Dot-Bracket file. Invisibly returns the status code returned by close when closing the file connection. See documentation of close for details.

References

https://software.broadinstitute.org/software/igv/RNAsecStructure

http://projects.binf.ku.dk/pgardner/bralibase/RNAformats.html

Examples

```
# Write a Dot-Bracket file by providing an RNA sequence and its secondary structure in
# the Dot-Bracket format:

tempDir <- tempdir()
DotBracketFile <- paste(tempDir, "testDotBracketFile.dot", sep="")
writeDotBracket(DotBracketFile, "AGCGGGUUCCUGGUUCCCCAAGGUUGA",
secondaryStructure="...(((..((.))..))).((..))..", sequenceName="Test sequence")
```

Index