Package 'FcircSEC'

January 31, 2020

Title Full Length Circular RNA Sequence Extraction and Classification

Version 1.0.0

Author Md. Tofazzal Hossain, Yin Peng, Shengzhong Feng and Yanjie Wei

Maintainer Md. Tofazzal Hossain <tofazzal.stat@gmail.com>

Description Extract full length circular RNA sequences and classify circular RNA
using the output of circular RNA prediction tools, reference genome and the annotation file corresponding to the reference genome.
This package uses the output of circular RNA prediction tools such as 'CIRI', 'CIRCExplorer' and the output of other state-of-the-art circular RNA prediction tools.
Details about the circular RNA prediction procedure can be found in
'Yuan Gao, Jinfeng Wang and Fangqing Zhao' (2015) <doi:10.1186/s13059-014-0571-3>
and 'Zhang XO, Wang HB, Zhang Y, Lu X, Chen LL and Yang L' (2014) <doi:10.1016/j.cell.2014.09.001>.

URL https://github.com/tofazzal4720/FcircSEC

BugReports https://github.com/tofazzal4720/FcircSEC/issues

Depends R (>= 3.6.0), Biostrings, seqRFLP, stringi

License GPL-2

Encoding UTF-8

LazyData true

RoxygenNote 7.0.0

NeedsCompilation no

Repository CRAN

Date/Publication 2020-01-31 15:50:02 UTC

R topics documented:

---

chr1 *chr1 of Reference genome*

Description

part of chr1 (first 2000000 nucleotides) from the reference genome.

Usage

```
data("chr1")
```

Format

A data frame with 1 observations on the following 2 variables.

`seq_name` Name of the chromosome

`sequence` Nucleotides sequence

Details

This data contains only the first 2000000 nucleotides of the reference chromosome 1.

Source

The full reference genome was downloaded from UCSC.

References

UCSC genome browser.

Examples

```
data(chr1)
## maybe str(chr1) ; plot(chr1) ...
```

`circClassification` *circRNA classification using trancript information and the bed file from the circRNA prediction tools*

Description

This function classifies circRNAs using the transcript information obtained from annotation file and the bedfile obtained from the circRNA prediction tools

Usage

```
circClassification(transcriptdata, bedfile, outfiletxt, outfilebed)
```

Arguments

`transcriptdata`	The transcript data (obtained from function `transcriptExtract`)
`bedfile`	The bed file (obtained from the circRNA prediction tools) having four columns chromosome, circRNA start, circRNA end position and circRNA strand
`outfiletxt`	The output file with the detailed information of circRNA classification
`outfilebed`	The output file with chromosome, start and end position of each circRNAs

Value

The detailed information of circRNA classification will be written in outfiletxt and only chromosome, start and end position of each circRNAs will be written in outfilebed

Examples

```
#Loading and example transcript data and write to a file
#Here temporary directory is created as input-output
#directory. Please provide you own directory instead.
out_dir<-tempdir()
t_data<-data("transcript_data")
t_data<-transcript_data
write.table(t_data, file.path(out_dir,"transcript_data.txt"), row.names=FALSE)

#Loading an example bedfile obtained form the circRNA prediction tool and write to a file
b_file<-data("output_CIRI")
b_file<-output_CIRI
write.table(b_file, file.path(out_dir,"output_CIRI.bed"), col.names=FALSE, row.names=FALSE)

#Classification of circRNAs. Here, the output will be written in two files
#circRNA_class.txt and circRNA_class.bed in out_dir directory
circClassification (file.path(out_dir,"transcript_data.txt"),
   file.path(out_dir,"output_CIRI.bed"), file.path(out_dir, "circRNA_class.txt"),
   file.path(out_dir, "circRNA_class.bed"))
```

`circRNA_classb` *circRNA classification data*

Description

This data is generated by r package FcircSEC.

Usage

```
data("circRNA_classb")
```

Format

A data frame with 14 observations on the following 3 variables.

`V1` Chromosome names

`V2` Start position

`V3` End position

Details

This is the output generated by the circClassification function of the r package FcircSEC. It has three columns: chromosome name, start position and end position of circRNAs.

Source

This example data is produced by r package FcircSEC.

References

R package FcircSEC.

Examples

```
data(circRNA_classb)
## maybe str(circRNA_classb) ; plot(circRNA_classb) ...
```

`circRNA_classt` *circRNA classification data*

Description

This data is generated by r package FcircSEC.

Usage

```
data("circRNA_classt")
```

Format

A data frame with 14 observations on the following 15 variables.

`ID` Circular RNA ID

`chr` Chromosome

`circ_start` Start position

`circ_end` End position

`circ_strand` Strand

`splice_L` circRNA length

`circ_type` circRNA type

`e_count` Number of exons

`e_sizes` Exon sizes

`e_offsets` Start positions of exons

`b_transcript` Transcript name

`b_strand` Transcript strand

`b_trans_start` Transcript start

`b_trans_end` Transcript end

`b_gene` Host gene

Details

This is the output generated by the circClassification function of the r package FcircSEC. It contains all necessary information for extracting the full length circRNA sequence.

Source

This example data is produced by r package FcircSEC.

References

R package FcircSEC.

Examples

```
data(circRNA_classt)
## maybe str(circRNA_classt) ; plot(circRNA_classt) ...
```

`circRNA_genomic_sequence` *Genomic sequences of circRNAs.*

Description

This data is generated by r package FcircSEC.

Usage

```
data("circRNA_genomic_sequence")
```

Format

A data frame with 14 observations on the following 2 variables.

`seq_name1` IDs of circular RNAs

`sequence1` Genomic sequences of circular RNAs

Details

This is the output generated by the get.fasta function of the r package FcircSEC. It contains the genomic sequences of the circRNAs.

Source

This example data is produced by r package FcircSEC.

References

R package FcircSEC.

Examples

```
data(circRNA_genomic_sequence)
## maybe str(circRNA_genomic_sequence) ; plot(circRNA_genomic_sequence) ...
```

circSeqExt *Generating full length circRNA sequences*

Description

This function can extract the full length circRNA sequences from the output of the circular RNA predictions tools

Usage

```
circSeqExt(genomic_seq, circ_class_txt, out_filename)
```

Arguments

genomic_seq	A fasta file (obtain using function get.fasta) with the genomic sequences for circRNAs
circ_class_txt	The circRNA classification file (obtained from function circClassification)
out_filename	The name of the output file

Value

The fasta file containing the full length circRNA sequences will be written in the output file 'out_filename'

Examples

```
#Loading an example circRNA genomic sequence and write to a file
#Here temporary directory is created as input-output
#directory. Please provide you own directory instead.
out_dir<-tempdir()
circ_genomic_seq<-data("circRNA_genomic_sequence")
circ_genomic_seq<-circRNA_genomic_sequence
df.fasta=dataframe2fas(circ_genomic_seq, file.path(out_dir, "circ_genomic_seq.fasta"))

#Loading an example circ_class_txt data and write to a file
circ_class_txt<-data("circRNA_classt")
circ_class_txt<-circRNA_classt
write.table(circ_class_txt, file.path(out_dir, "circ_class.txt"),
        row.names=FALSE)

#Extracting full length circRNA sequences. Here, the output will be
#written in file circRNA_sequence.fasta in out_dir directory
circSeqExt(file.path(out_dir, "circ_genomic_seq.fasta"),
 file.path(out_dir, "circ_class.txt"), file.path(out_dir, "circRNA_sequence.fasta"))
```

get.fasta *Generating sequences from the reference genome with specific intervals*

Description

This function can extract the sequences from the reference genome for the given intervals (start, end) of chromosomes

Usage

```
get.fasta(ref_genome, circ_class_bed, out_filename)
```

Arguments

ref_genome	The reference genome
circ_class_bed	The bed file having chromosome, start and end position of each circRNAs (obtained from function circClassification)
out_filename	The name of the output file

Value

The fasta file of the sequences extracted from the reference genome for the given intervals will be written in the output file 'out_filename'

Examples

```
#Loading an example reference genome and write to a file
#Here temporary directory is created as input-output
#directory. Please provide you own directory instead.
out_dir<-tempdir()
ref_genom<-data("chr1")
ref_genom<-chr1
df.fasta=dataframe2fas(ref_genom, file.path(out_dir, "ref_genome.fasta"))

#Loading an example circRNA classification bed file and write to a file
circ_class_bed<-data("circRNA_classb")
circ_class_bed<-circRNA_classb
write.table(circ_class_bed, file.path(out_dir, "circ_class.bed"),
    col.names=FALSE, row.names=FALSE)

#Getting genomic sequences of circRNAs. The output will be
#generated in file circRNA_genomic_seq.fasta in out_dir directory
get.fasta(file.path(out_dir, "ref_genome.fasta"),
   file.path(out_dir, "circ_class.bed"),
   file.path(out_dir, "circRNA_genomic_seq.fasta"))
```

output_CIRI *Output obtained from circRNA prediction tool.*

Description

Four columns of the circRNA prediction tool.

Usage

```
data("output_CIRI")
```

Format

A data frame with 14 observations on the following 4 variables.

`V1` Chromosome name

`V2` Start position

`V3` End position

`V4` Strand

Details

Four columns of the circRNA prediction tool. The columns have information about chromosome, circRNA start and end positions, and strand.

Source

An example data from the circRNA prediction tool CIRI.

References

This data is obtained from the circRNA prediction tool CIRI.

Examples

```
data(output_CIRI)
## maybe str(output_CIRI) ; plot(output_CIRI) ...
```

refGenchr1 *Gene annotation of chromosome 1 of human reference genome.*

Description

This data is a part of the gene annotation of chr1 of human reference genome.

Usage

```
data("refGenchr1")
```

Format

A data frame with 3066 observations on the following 9 variables.

`V1` Name of chromosome

`V2` Source

`V3` Feature

`V4` Start position

`V5` End position

`V6` Score

`V7` Strand

`V8` Frame

`V9` Attribute

Details

This is an example annotation file of chr1 (human reference genome) whose start position is less than or equal 2000000.

Source

Downloaded from UCSC genome browser.

References

UCSC genome browser.

Examples

```
data(refGenchr1)
## maybe str(refGenchr1) ; plot(refGenchr1) ...
```

`transcriptExtract` *Extracting transcript information from the annotation file*

Description

This function extracts transcript information from the annotation file corresponding to reference genome

Usage

```
transcriptExtract(annotationFile, databaseName, outputfile)
```

Arguments

`annotationFile`	The annotation file (in gtf, gff or gff3 fromat) corresponding to the reference genome
`databaseName`	The database name from where the annotation file was downloaded (the possible options are "ncbi", "ucsc" and "other")
`outputfile`	The name of the output file

Value

The transcript information from the annotation file will be written in the output file 'outputfile'

Examples

```
#Loading an example annotation file and write to a file
#Here temporary directory is created as input-output
#directory. Please provide your own directory instead.
out_dir<-tempdir()
annotation_file<-data(refGenchr1)
annotation_file<-refGenchr1
write.table(annotation_file, file.path(out_dir,"annotation_file.gtf"),
         row.names=FALSE, sep="\t",quote=FALSE, col.names=FALSE)

#Extraction of transcript information. Here, the output will be generated in file
#transcriptdata.txt in out_dir directory
transcriptExtract(file.path(out_dir,"annotation_file.gtf"), "ucsc",
    file.path(out_dir, "transcriptdata.txt"))
```

`transcript_data` *Transcript data obtained from gene annotation file.*

Description

This data is produced by the r package FcircSEC.

Usage

```
data("transcript_data")
```

Format

A data frame with 177 observations on the following 9 variables.

`transcript_id` Transcript name

`chr` Chromosome

`strand` Strand

`trans_start` Transcript start

`trans_end` Transcript end

`exon_count` Number of exons

`exon_starts` Start position of exons

`exon_ends` End position of exons

`gene` Gene name

Details

This is an example data produced by the function transcriptExtract of the r package FcircSEC.

Source

This data is produced by r package FcircSEC.

References

R package FcircSEC.

Examples

```
data(transcript_data)
## maybe str(transcript_data) ; plot(transcript_data) ...
```

Index