Package ‘sequences’

February 20, 2015

Title  Generic and Biological Sequences
Version  0.5.9
Description  Educational package used in R courses to illustrate object-oriented programming and package development. Using biological sequences (DNA and RNA) as a working example.
Depends  methods, Rcpp
LinkingTo  Rcpp
Suggests  testthat, knitr
License  GPL-3
biocViews  Infrastructure, DataRepresentation, DataImport
VignetteBuilder  knitr
URL  https://github.com/lgatto/sequences/
Author  Laurent Gatto [aut, cre],
Robert Stojnic [aut]
Maintainer  Laurent Gatto <lg390@cam.ac.uk>
NeedsCompilation  yes
Repository  CRAN
Date/Publication  2014-12-03 13:58:03

R topics documented:

sequences-package .................................................. 2
dnaseq ............................................................... 2
DnaSeq-class ......................................................... 3
gcount ............................................................... 4
gcount2 .............................................................. 4
GenericSeq-class .................................................... 5
readFasta ............................................................ 6
Seq-class ............................................................ 7

Index  9
sequences-package  
A short demo package featuring biological sequences.

Description

Dummy package used in an R course to illustrate OO programming and package development. The course is aimed at bioinformaticians and biologists. Relevant illustrative examples used in the package are generic sequences as a top virtual class and specific biological (DNA and RNA) sequences.

Details

Package: sequences  
Type: Package  
License: GPL-3  
LazyLoad: yes

Author(s)

Laurent Gatto <lg390@cam.ac.uk> Robert Stojnic <rs550@cam.ac.uk> Maintainer: Laurent Gatto <lg390@cam.ac.uk>

dnaseq

Description

dnaseq is an example instance of an object of class DnaSeq. It has been generated by reading the adnaSeq.fasta file (available in syste.file("extdata",package="sequences")) with the readFasta function.

Usage

data(dnaseq)

Format

The format is: chr "dnaseq"
Examples

```r
data(dnaseq)
dnaseq
print(dnaseq)
```

DnaSeq-class  Class "DnaSeq" and "RnaSeq"

Description

The DnaSeq and RnaSeq are instances of the virtual GenericSeq class for "DNA" and "RNA" respectively. Their alphabets are defined accordingly.

Objects from the Class

Instance of DnaSeq and RnaSeq can be created by reading fasta files using the readFasta function.

Slots

See the GenericSeq class for details about the generic slots and methods.

Extends

Class GenericSeq, directly.

Methods

See GenericSeq for inherited methods.

- **comp** signature(object = "DnaSeq"): return the complement of the sequence string of the sequence instance.
- **transcribe** signature(object = "DnaSeq"): transcribes the DnaSeq instance into a RnaSeq instance.

Author(s)

Laurent Gatto <lg390@cam.ac.uk>

See Also

See also the GenericSeq virtual class.

Examples

```r
data("dnaseq")
dnaseq
rnaseq <- transcribe(dnaseq)
rnaseq
```
Description

Returns the number of 'A', 'C', 'G' and 'T' bases in the 'inseq' sequence string.

Usage

gccount(inseq)

Arguments

inseq a DNA sequence string.

Details

This function calls a C primitive

Value

A numeric of length 4

Author(s)

Laurent Gatto <lg390@cam.ac.uk>

Examples

s <- "AAAACCCGCT"
cnt <- gccount(s)
cnt
stopifnot(cnt==table(strsplit(s,"")))
Arguments

inseq a DNA sequence string.

Details

This function calls a C primitive

Value

A numeric of length 4

Author(s)

Robert Stojnic <rs550@cam.ac.uk>

Examples

s <- "AAAAACCCGCT"
cnt <- gccount2(s)
cnt
stopifnot(cnt==table(strsplit(s,""))

---

GenericSeq-class  Class "GenericSeq"

Description

This virtual class defines a generic sequence object composed by an identifier, a sequence and an alphabet that defines the letters of the sequence. A set of accessor methods and setters are described below.

Objects from the Class

A virtual Class: No objects may be created from it.

Slots

id: Object of class "character" that identifies the sequence instance.
alphabet: Object of class "character" that defines all the letters that a valid sequence can be made of.
sequence: Object of class "character" storing the actual sequence string of the sequence instance.
Methods

alphabet signature(object = "GenericSeq"): retrieves the alphabet of valid sequence letters, returned as character.

id signature(object = "GenericSeq"): retrieves the identifier of the sequence instance.

id<- signature(object = "GenericSeq", value = "character"): sets the identifier of the sequence instance to value.

length signature(x = "GenericSeq"): returns the length the sequence instance, i.e. the number of characters of the sequence string.

print signature(x = "GenericSeq"): prints the sequence of the instance.

rev signature(x = "GenericSeq"): return the reverse of the sequence string of the sequence instance.

seq signature(... = "GenericSeq"): returns the sequence character string of the instance.

seq<- signature(object = "GenericSeq", value = "character"): changes the sequence string of the sequence instance to value. The new string must be compatible with the sequence instance alphabet.

show signature(object = "GenericSeq"): displays the object content.

[ signature(x = "GenSeq"): Subsets the sequence of x and returns the corresponding subsetted object of same class as x.

Author(s)

Laurent Gatto <lg390@cam.ac.uk>

See Also

See DnaSeq and RnaSeq classes for examples of subclasses and readfasta on how to create instances.

Examples

showClass("GenericSeq")

readFasta

---

readFasta

*Read fasta files.*

Description

Reads sequences data in fasta and create DnaSeq and RnaSeq instances.

Usage

readFasta(infile)
**Arguments**

infile the name of the fasta file which the data are to be read from.

**Details**

This function reads DNA and RNA fasta files and generates valid "DnaSeq" and "RnaSeq" instances.

**Value**

an instance of DnaSeq or RnaSeq.

**Author(s)**

Laurent Gatto <lg390@cam.ac.uk>

**See Also**

GenericSeq, DnaSeq and RnaSeq.

**Examples**

```r
f <- dir(system.file("extdata",package="sequences"),pattern="fasta",full.names=TRUE)
f
aa <- readFasta(f[1])
```

---

**Seq-class**  

*Class* "Seq"

**Description**

The Seq class implements a nucleic acid sequence as well as various sequence transformations. Its aim is to illustrate the usage of reference classes.

**Extends**

All reference classes extend and inherit methods from "envRefClass".

**Fields**

- id: Object of class character defining the sequence identifier.
- alphabet: Object of class character defining the sequence alphabet.
- sequence: Object of class character with the actual sequence string.
Methods

transcribe(): Transcribes a DNA sequence into RNA.

seq(): Returns the sequence string.

comp(): Complements the sequence.

rev(): Reverses the sequence.

valid(): Checks sequence validity.

setAlphabet(): Sets the sequence alphabet.

show(): Shows the 'Seq' instance.

Author(s)

Laurent Gatto <lg390@cam.ac.uk>

See Also

DnaSeq for a S4 implementation of the same subject.

Examples

```r
s <- Seq$new(id="My sequence",
             sequence="AGCAGCTACGACT",
             alphabet=c("A", "C", "G", "T"))

s
s$rev()
```

Index

*Topic IO,
readfasta, 6

*Topic classes
 DnaSeq-class, 3
 GenericSeq-class, 5
 Seq-class, 7

*Topic datasets
dnaseq, 2

*Topic file
readfasta, 6

*Topic package
sequences-package, 2
 [,DnaSeq-method (GenericSeq-class), 5
 [,GenericSeq-method (GenericSeq-class), 5
 [,RnaSeq-method (GenericSeq-class), 5
 [,method (GenericSeq-class), 5

alphabet (GenericSeq-class), 5
alphabet,DnaSeq-method
 (GenericSeq-class), 5
alphabet,GenericSeq-method
 (GenericSeq-class), 5
alphabet,RnaSeq-method
 (GenericSeq-class), 5
alphabet-method (GenericSeq-class), 5

class:DnaSeq (DnaSeq-class), 3
class:GenericSeq (GenericSeq-class), 5
class:RnaSeq (DnaSeq-class), 5
class:Seq (Seq-class), 7

comp (DnaSeq-class), 3
comp,DnaSeq-method (DnaSeq-class), 3
comp,RnaSeq-method (DnaSeq-class), 3
comp-method (DnaSeq-class), 3

DnaSeq, 2, 6–8
dnaseq, 2
DnaSeq-class, 3

envRefClass, 7
gccount, 4
gccount2, 4
GenericSeq, 3, 7
GenericSeq-class, 5

id (GenericSeq-class), 5
id,DnaSeq-method (GenericSeq-class), 5
id,GenericSeq-method
 (GenericSeq-class), 5
id,RnaSeq-method (GenericSeq-class), 5
id<- (GenericSeq-class), 5
id<-,GenericSeq, ANY-method
 (GenericSeq-class), 5

length,DnaSeq-method
 (GenericSeq-class), 5
length,GenericSeq-method
 (GenericSeq-class), 5
length,RnaSeq-method
 (GenericSeq-class), 5

print,DnaSeq-method (GenericSeq-class), 5
print,GenericSeq-method
 (GenericSeq-class), 5
print,RnaSeq-method (GenericSeq-class), 5

readFasta, 2, 3, 6, 6
rev (GenericSeq-class), 5
rev,DnaSeq-method (GenericSeq-class), 5
rev,GenericSeq-method
 (GenericSeq-class), 5
rev,RnaSeq-method (GenericSeq-class), 5
rev-method (GenericSeq-class), 5
RnaSeq, 6, 7
RnaSeq-class (DnaSeq-class), 3
Seq (Seq-class), 7
seq (GenericSeq-class), 5
seq,DnaSeq-method (GenericSeq-class), 5
seq,GenericSeq-method
  (GenericSeq-class), 5
seq,RnaSeq-method (GenericSeq-class), 5
Seq-class, 7
seq-method (GenericSeq-class), 5
seq<- (GenericSeq-class), 5
seq<-, GenericSeq, ANY-method
  (GenericSeq-class), 5
seq<-, GenericSeq, character-method
  (GenericSeq-class), 5
sequences (sequences-package), 2
sequences-package, 2
show,DnaSeq-method (GenericSeq-class), 5
show,GenericSeq-method
  (GenericSeq-class), 5
show,RnaSeq-method (GenericSeq-class), 5

transcribe (DnaSeq-class), 3
transcribe,DnaSeq-method
  (DnaSeq-class), 3
transcribe-method (DnaSeq-class), 3