Package ‘genomeIntervals’

January 21, 2017

Version 1.30.0
Date 2015-07-15
Type Package
Title Operations on genomic intervals
Author Julien Gagneur <gagneur@genzentrum.lmu.de>, Joern Toedling, Richard Bourgon, Nicolas Delhomme
Maintainer Julien Gagneur <gagneur@genzentrum.lmu.de>
Depends R (>= 2.15.0), methods, intervals (>= 0.14.0), BiocGenerics
       (>= 0.15.2)
Imports GenomeInfoDb (>= 1.5.8), GenomicRanges (>= 1.21.16),
       IRanges(>= 2.3.14), S4Vectors (>= 0.7.10)
Suggests
Description This package defines classes for representing genomic intervals
and provides functions and methods for working with these.
Note: The package provides the basic infrastructure for and is
enhanced by the package ‘girafe’.
License Artistic-2.0
Collate AllGenerics.R Genome_intervals-class.R
       Genome_intervals-coercion-methods.R Genome_intervals-package.R
       core_annotated.R distance_to_nearest-methods.R
       interval_complement-methods.R interval_overlap-methods.R
       interval_intersection-methods.R interval_union-methods.R
       writeGff3.R
biocViews DataImport, Infrastructure, Genetics
LazyLoad yes
NeedsCompilation no

R topics documented:

  genomeIntervals-package .................................................. 2
  c ................................................................. 3
  core_annotated .................................................. 4
  distance_to_nearest ........................................ 5
Operations on genomic intervals

Description

Tools for operation on genomic intervals.

Details

Package: genomeIntervals
Version: 1.25.3
Date: 2015-07-15
Type: Package
Depends: R (>= 2.15.0), intervals (>= 0.14.0), BiocGenerics, methods
Imports: GenomeInfoDb (>= 1.5.8)
Suggests: 
License: Artistic 2.0
BiocViews: DataImport, Infrastructure, Genetics
LazyLoad: yes

Index:

GenomeIntervals Class "Genome\_intervals"
GenomeIntervals\_stranded Class "Genome\_intervals\_stranded"
distance\_to\_nearest Distance in bases to the closest interval(s)
gen\_ints Genome Intervals examples
getGffAttribute Pull one or more key/value pairs from gffAttributes strings
interval\_overlap Assess overlap from one set of genomic intervals to another
interval\_complement Compute the complement of a set of genomic intervals
interval\_intersection Compute the intersection of one or more sets of genomic intervals
interval\_union Compute the union of genomic intervals in one or more genomic interval matrices
**parseGffAttributes**  Parse out the gffAttributes column of a Genome_intervals object

**readGff3**  Make a Genome_intervals_stranded object from a GFF file

**Author(s)**

Julien Gagneur <gagneur@embl.de>, Richard Bourgon, Joern Toedling, Nicolas Delhomme.

Maintainer: Julien Gagneur <gagneur@embl.de>

**See Also**

intervals

---

**Description**

This function combines several genome intervals (stranded or not) objects into a single one.

**Usage**

```r
## S4 method for signature 'Genome_intervals'
c(x, ..., recursive = FALSE)
```

**Arguments**

- `x`  a Genome_intervals or Genome_intervals_stranded object - not mandatory.
- `...` two (one if `x` is defined) or more Genome_intervals or Genome_intervals_stranded objects.
- `recursive` inherited from the base c function definition and not used.

**Details**

If the arguments have mixed classes (both Genome_intervals or Genome_intervals_stranded), then they are coerced to Genome_intervals before combination. Otherwise, the common class is used. If a list is provided with NULL entries, these are discarded. If a vector of object is provided with non GenomeIntervals classes, then a list, ordered as the input vector, is returned.

**Value**

- A single Genome_intervals or Genome_intervals_stranded object. Input objects are combined in their order of appearance in the the argument list.
- If any input argument is not a Genome_intervals, list(...) is returned instead.

```r
##'
```
Examples

```r
##' load toy examples
data("gen_ints")

##' combine i and j returns a Genome_intervals_stranded object
c( i, j )

##' combine a not-stranded and a stranded returns a not-stranded object
c( as(i, "Genome_intervals"), j )
```

---

**core_annotated**

*Genome intervals with minimal annotation*

**Description**

returns a copy of the input (stranded) genome intervals object with annotations restricted to the minimally required ones.

**Usage**

core_annotated(x)

**Arguments**

- `x`: A `Genome_intervals` or `Genome_intervals_stranded` object.

**Value**

A copy of `x` with the annotation slot restricted to `seq_name`, `inter_base` and `strand` (the latter only if `x` is a `Genome_intervals_stranded` object).

**Examples**

```r
# load toy examples
data("gen_ints")

# add some non-core annotations to i
annotation(i)$comment = "some non-core annotation"

# i with all annotations
i

# core annotations only
core_annotated(i)

## Not run:
# with different annotation columns, i and j cannot be combined
c( i, j )

## End(Not run)

# core annotated versions can
c( core_annotated(i), core_annotated(j) )
```
distance_to_nearest

Description

Given two objects, from and to, compute the distance in bases of each from interval to the nearest to interval(s). The distance between a base and the next inter-bases on either side values 0.5. Thus, base - base and inter-base - inter-base intervals distances are integer, whereas base - inter-base intervals distances are half-integers.

Usage

```r
## S4 method for signature 'Genome_intervals,Genome_intervals'
distance_to_nearest(from, to)
## S4 method for signature
## 'Genome_intervals_stranded,Genome_intervals_stranded'
distance_to_nearest(from, to)
```

Arguments

- `from`: A Genome_intervals or Genome_intervals_stranded object.
- `to`: A Genome_intervals or Genome_intervals object.

Details

A wrapper calling intervals::distance_to_nearest by seqnames and by strand (if both from and to are Genome_intervals_stranded objects). Thus, if both are stranded, distances are computed over each strand separately. One object must be coerced to Genome_intervals if this is not wished.

Value

A numeric vector of distances with one element for each row of from.

See Also

 intervals::distance_to_nearest

Examples

```r
## load toy examples
data(gen_ints)

## i in close_intervals notation
close_intervals(i)

## j in close_intervals notation
close_intervals(j)

## distances from i to j
dn = distance_to_nearest(i, j)
dn
```
## distance == 0 if and only if the interval overlaps another one:

```r
io = interval_overlap(i,j)
if( any( ( sapply(io, length) >0 ) != (!is.na(dn) & dn ==0) ) )
  stop("The property 'distance == 0 if and only if the interval overlaps another one' is not followed for at least one instance.
```

## distances without strand-specificity

```r
distance_to_nearest(
  as(i,"Genome_intervals"),
  as(j,"Genome_intervals")
)
```

---

**GenomeIntervals**  
**Constructor function for genomeIntervals objects**

### Description

A user-friendly constructor function for creating both Genome_intervals and Genome_intervals_stranded objects.

### Usage

```r
GenomeIntervals(chromosome, start, end, strand = NULL, inter.base = NULL, leftOpen = NULL, rightOpen = NULL, ...)
```

### Arguments

- **chromosome**: character vector of chromosome names of the intervals; will become the seqnames of the resulting object
- **start**: numeric or integer; start (left-most) coordinate of the intervals
- **end**: numeric or integer; end (right-most) coordinate of the intervals
- **strand**: character; specifies which strand the intervals are located on; if specified an object of class Genome_intervals_stranded is created; if NULL an object of class Genome_intervals is created
- **inter.base**: logical; if TRUE an interval is located between the specified coordinates, instead of spanning them; useful for restriction-enzym cutting sites, for example.
- **leftOpen**: logical; if TRUE an interval is left-open; if NULL all intervals are assumed to be left-closed.
- **rightOpen**: logical; if TRUE an interval is right-open; if NULL all intervals are assumed to be right-closed.
- **...**: any additional annotation for supplied intervals

### Details

The arguments chromosome, start, and end need to be of the same length, with the first element of each vector corresponding to the first interval, the second element to the second interval, and so on.

The same applies to strand, inter.base, leftOpen, rightOpen and any additional vectors in `...`, if they are specified.
Value

An object of class Genome_intervals or Genome_intervals_stranded depending on whether strand has been specified.

Author(s)

J. Toedling

See Also

Genome_intervals-class, Genome_intervals_stranded-class

Examples

## constructing a Genome_intervals object
G <- GenomeIntervals(start=c(1,3,4,5,8,10), end=c(5,5,6,8,9,11),
                     chromosome=rep(c("chr2","chrX","chr1"), each=2),
                     leftOpen=rep(c(FALSE, FALSE, TRUE), 2))
show(G)

## constructing a Genome_intervals_stranded object with
## additional interval annotation
GS <- GenomeIntervals(start=c(1,3,4,5,8,10), end=c(5,5,6,8,9,11),
                       chromosome=rep(c("chr2","chrX","chr1"), each=2),
                       strand=c("-","-","+","+","+","+"),
                       GC.content=round(runif(6), digits=2))
show(GS)
Author(s)
Nicolas Delhomme

See Also
- genomeIntervals object
- readGff3 function

Examples

```r
## Not run:
annot<-readGff3(system.file("extdata", "Dmel-mRNA-exon-r5.52.gff3", package="RnaSeqTutorial")
gAnnot<-as(annot,"RangedData") type(annot)
## End(Not run)
```

Genome_intervals deprecated functions

The following function have been deprecated:
- seq_name
- seq_name<- 

Description

- The seq_name and seq_name<- accessor functions have been replaced by the more generic seqnames and seqnames<- accessor functions, respectively.

Genome_intervals-class

Class "Genome\_intervals"

Description

A set of genomic intervals without specified strand. Genomic intervals are intervals over the integers with two further annotations: seqnames (a chromosome or more generally a sequence of origin) and inter_base (logical) that states whether the interval is to be understood as an interval over bases (such as coding-sequence) or inter-bases (such as restriction sites or insertion positions).

Slots

.Data: See Intervals_full
annotation: A "data.frame" with the same number of rows as .Data. It has a column named seq_name that is a factor and does not contain missing values. seq_name is used to represent the chromosome or more generally the sequence of origin of the intervals. annotation has a column named inter_base that is logical and does not contain missing values. inter_base is FALSE if the interval is to be understood as an interval over bases (such as coding-sequence) and TRUE if it is over inter-bases (such as restriction site or an insertion position). Like base intervals, inter-base interval are encoded over the integers. An inter-base at position n indicates the space between base n and n+1.

closed: See Intervals_full
type: See Intervals_full

Extends
Class "Intervals_full", directly. Class "Intervals_virtual", by class "Intervals\_full", distance 2. Class "matrix", by class "Intervals\_full", distance 3. Class "array", by class "Intervals\_full", distance 4. Class "structure", by class "Intervals\_full", distance 5. Class "vector", by class "Intervals\_full", distance 6, with explicit coerce.

Methods

\[ \text{signature(x = "Genome_intervals")}: ... \]
\[ \text{[[ signature(x = "Genome_intervals")}: ... \]
\[ \text{[<- signature(x = "Genome_intervals")}: ... \]
\$ \text{signature(x = "Genome_intervals")}: ...
\$\text{<- signature(x = "Genome_intervals")}: ...

\text{annotation signature(object = "Genome_intervals")}: ...
\text{annotation<- signature(object = "Genome_intervals")}: ...
\text{coerce signature(from = "Genome_intervals", to = "Intervals_full")}: ...
\text{coerce signature(from = "Genome_intervals", to = "character")}: ...
\text{coerce signature(from = "Genome_intervals", to = "data.frame")}: ...
\text{distance\_to\_nearest signature(from = "Genome_intervals", to = "Genome_intervals")}: ...

\text{inter\_base signature(x = "Genome_intervals")}: ...
\text{inter\_base<- signature(x = "Genome_intervals")}: ...
\text{interval\_complement signature(x = "Genome_intervals")}: ...
\text{interval\_intersection signature(x = "Genome_intervals")}: ...
\text{interval\_overlap signature(from = "Genome_intervals", to = "Genome_intervals")}: ...
\text{interval\_union signature(x = "Genome_intervals")}: ...
\text{seqnames signature(x = "Genome_intervals")}: ...
\text{seqnames<- signature(x = "Genome_intervals")}: ...
\text{size signature(x = "Genome_intervals")}: ...
\text{type< signature(x = "Genome_intervals")}: ...
\text{which\_nearest For each interval in Set1, finds nearest (least distant) interval in Set2. Intervals on different chromosomes are never considered \text{`near'} to each other. The returned value is a data.frame with the number of rows equal to the number of intervals in Set1. Each row specifies the distance to the nearest interval in Set2 (a 0 means that the interval overlaps one or more intervals in Set2), and the indices of near and overlapping intervals in Set2. See Intervals_full for further details.}
**width**  Returns the interval length as the number of bp covered (base interval) or spanned (inter-base interval). Similar to the IRanges package `width` function.

**Note**
A `Genome_intervals` is a "Intervals_full" of type Z (i.e. a set of intervals over the integers). The annotation slot can carry further columns that can serve as annotations.

**See Also**
`Genome_intervals_stranded` for a derived class that allows stranded genomic intervals.

**Examples**

```r
# The "Genome_intervals" class

i <- new("Genome_intervals",
  matrix(
    c(1,2,
      3,5,
      4,6,
      8,9
    ),
    byrow = TRUE,
    ncol = 2
  ),
  closed = matrix(
    c(
      TRUE, FALSE,
      TRUE, FALSE,
      TRUE, TRUE,
      TRUE, FALSE
    ),
    byrow = TRUE,
    ncol = 2
  ),
  annotation = data.frame(
    seq_name = factor(c("chr01","chr01", "chr02","chr02")),
    inter_base = c(FALSE, FALSE, TRUE, TRUE)
  )
)

colnames(i) <- c( "start", "end" )

# print
print(i)

# size (number of bases per interval)
size(i)

## convert to a data.frame
as(i,"data.frame")

## simpler way to construct a Genome_intervals object:
G <- GenomeIntervals(start=c(1,3,4,5,10,8), end=c(5,5,6,8,11,9), chromosome=rep(c("chr2","chrX","chr1"), each=2),
```

---

The `Genome_intervals` class provides methods for working with genomic intervals, including functions for calculating the length of intervals and for creating objects with specific properties. This class is particularly useful for handling genomic data, where intervals can represent regions of interest on a chromosome.

### Functionality

- **`width`**: This function returns the interval length as the number of bases covered by the interval. It is similar to the `width` function in the IRanges package, which is commonly used in bioinformatics for genomic data analysis.

- **`Note`**: The `Genome_intervals` class is designed to handle intervals over integers, allowing for complex annotations and calculations. The annotation slot can carry additional information, such as sequence names or specific bases, which can be useful for detailed analysis.

### Examples

The examples demonstrate how to create a `Genome_intervals` object and perform basic operations on it, such as printing the object and calculating the size of intervals. Additionally, a simpler way to construct a `Genome_intervals` object is shown, which can be more convenient for setting up large datasets.

---

This class is part of a larger toolkit designed to facilitate the analysis and manipulation of genomic data. It is an essential component for researchers and practitioners working with genomic information, providing a robust framework for integrating and analyzing data across different scales.
## S4 method for signature 'Genome_intervals'
order(..., na.last=TRUE, decreasing=FALSE)
## S4 method for signature 'Genome_intervals_stranded'
order(..., na.last=TRUE, decreasing=FALSE)

## S4 method for signature 'Genome_intervals'
sort(x, decreasing=FALSE, ...)
## S4 method for signature 'Genome_intervals'
rank(x, na.last=TRUE,
ties.method = c("average", "first", "random", "max", "min") )
## S4 method for signature 'Genome_intervals'
xtfrm(x)

### Arguments

- **x**: Objects of class `Genome_intervals` or `Genome_intervals_stranded`.
- **...**: Objects of class `Genome_intervals`, `Genome_intervals_stranded` or of any other class for order.
- **na.last**: Ignored for ordering `Genome_intervals` and `Genome_intervals_stranded` objects.
- **decreasing**: TRUE or FALSE.
- **ties.method**: A character string specifying how ties are treated. Only "first" is supported.

### Details

An order on Genome_intervals entries is defined by sorting by 1. seqnames 2. start, where closed start & not inter-base < closed start & inter-base < open start & not inter-base < open start & inter-base 3. stop, where open stop & not inter-base < open stop & inter-base < closed stop & not inter-base < closed stop & inter-base 4. strand (for Genome_intervals_stranded object)
The factors seqnames and strand are sorted according to their levels (default R behavior).
The primitive is implemented in xtfrm which is then called by the other methods. Hence, the order, sort and rank methods are consistent.
Genome_intervals_stranded-class

Class "Genome\_intervals\_stranded"

Description

A set of genomic intervals with a specified strand.

order(..., na.last=TRUE, decreasing=TRUE): return a permutation which rearranges its first argument into ascending or descending order, breaking ties by further arguments. See order in the base package for more details. na.last is ignored for Genome_intervals objects.

rank(x, na.last=TRUE, ties.method=c("average", "first", "random", "max", "min")): Return the sample ranks of the (stranded) genome intervals in x. See rank in the base package for more details.

sort(x): Sort x. See sort in the base package for more details.

xtfrm(x): Auxiliary function that produces a numeric vector which will sort in the same order as 'x'. See xtfrm in the base package for more details. Workhorse for the other methods

See Also

Genome_intervals Genome_intervals_stranded order, sort, rank, xtfrm

Examples

```r
## an example with ties
gi = GenomeIntervals(c("chr2", "chr2", "chr1", "chr1"), c(1,1,10,10), c(5,3,12,12) )

sort(gi)
rank(gi)
order(gi)

## Define order on seqnames at your convenience
## by specifying the order of the levels
## compare:
gi = GenomeIntervals(
c("chr2", "chr2", "chr10", "chr10"),
c(1,1,10,10),
c(5,3,12,12)
)

sort(gi)

## with:

gi2 = GenomeIntervals(
 factor(c("chr2", "chr2", "chr10", "chr10"), levels=c("chr2", "chr10")),
c(1,1,10,10),
c(5,3,12,12)
)

sort(gi2)
```
Slots

.Data: See Genome_intervals

annotation: A data.frame (see Genome_intervals for basic requirements). The annotation moreover has a strand column that is a factor with exactly two levels (typically "+" and "+").

closed: See Genome_intervals
type: See Genome_intervals

Extends


Methods

coerce signature(from = "Genome_intervals_stranded", to = "character"): ...
distance\_to\_nearest signature(from = "Genome_intervals_stranded", to = "Genome_intervals_stranded") ...
interval\_complement signature(x = "Genome_intervals_stranded"): ...
interval\_intersection signature(x = "Genome_intervals_stranded"): ...
interval\_overlap signature(to = "Genome_intervals_stranded", from = "Genome_intervals_stranded"): ...
interval\_union signature(x = "Genome_intervals_stranded"): ...
strand signature(x = "Genome_intervals_stranded"): ...
strand<- signature(x = "Genome_intervals_stranded"): ...

See Also

Genome_intervals the parent class without strand.

Examples

# The "Genome_intervals_stranded" class
j <- new(
  "Genome_intervals_stranded",
  matrix(
    c(1,2,
      3,5,
      4,6,
      8,9
    ),
    byrow = TRUE,
    ncol = 2
  ),
  closed = matrix(
    c(FALSE, FALSE,
      TRUE, FALSE,
      FALSE, FALSE,
      TRUE, FALSE,
      TRUE, FALSE))
)
TRUE, TRUE,
TRUE, FALSE
),
byrow = TRUE,
ncol = 2
),
annotation = data.frame(
  seq_name = factor( c("chr01","chr01", "chr02","chr02") ),
  strand = factor( c("+","+","+","-" ) ),
  inter_base = c(FALSE,FALSE,FALSE,TRUE)
)
)

## print
print(j)

## size of each interval as count of included bases
size(j)

## close intervals left and right (canonical representation)
close_intervals(j)

## simpler way to construct a Genome_intervals_stranded object
GS <- GenomeIntervals(start=c(1,3,4,5,8,10), end=c(5,5,6,8,9,11),
  chromosome=rep(c("chr2","chrX","chr1"), each=2),
  strand=c("-","-","+","+","+","+" ) )
show(GS)

---

**Description**

Toy examples for testing functions and running examples of the package genomeIntervals.

**Usage**

```r
data(gen_ints)
```

**Format**

Two Genome_intervals_stranded objects, i and j, without inter-base intervals and a third one, k, with.
getGffAttribute

Pull one or more key/value pairs from gffAttributes strings

Description

GFF files contain a string, with key/value pairs separated by “;”, and the key and value separated by “=”.
This function quickly extracts one or more key/value pairs.

Usage

getGffAttribute(gi, attribute)

Arguments

- gi: A Genome_intervals object.
- attribute: A vector of key names.

Value

A matrix with the same number of rows as gi, and one column per element of attribute.

See Also

- See parseGffAttributes for more complete parsing.
- See the function readGff3 for loading a GFF file.

Examples

```r
# Get file path
libPath <- installed.packages()["genomeIntervals", "LibPath"]
filePath <- file.path(
  libPath,
  "genomeIntervals",
  "example_files"
)

# Load gff
gff <- readGff3( filepath( filePath, "sgd_simple.gff" ), isRightOpen=FALSE )

## head of full gff annotations
head( annotation(gff) )

# extract ID and Parent attributes
idpa = getGffAttribute( gff, c( "ID", "Parent" ) )

head(idpa)
```
### interval_overlap

Assess overlap from one set of genomic intervals to another

**Description**

Given two objects, a 'from' and a 'to', assess which intervals in 'to' overlap which of 'from'.

**Usage**

```r
## S4 method for signature 'Genome_intervals,Genome_intervals'
interval_overlap(
  from, to,
  check_valid = TRUE
)
## S4 method for signature
## 'Genome_intervals_stranded,Genome_intervals_stranded'
interval_overlap(
  from, to,
  check_valid = TRUE
)
```

**Arguments**

- `from` A Genome_intervals or Genome_intervals_stranded object.
- `to` A Genome_intervals or Genome_intervals_stranded object.
- `check_valid` Should `validObject` be called before passing to compiled code?

**Details**

A wrapper calling `intervals:interval_overlap` by seq_name and by strand (if both `to` and `from` are “Genome_intervals_stranded” objects).

**Value**

A list, with one element for each row of `from`. The elements are vectors of indices, indicating which to rows overlap each from. A list element of length 0 indicates a from with no overlapping to intervals.

**Examples**

```r
data(gen_ints)
# i as entered
i

# i in close_intervals notation
close_intervals(i)

# j in close_intervals notation
close_intervals(j)

# list of intervals of j overlapping intervals of i
```
interval_union

interval_overlap(i,j)

interval_union

Genome interval set operations

Description

Compute interval set operations on "Genome_intervals" or "Genome_intervals_stranded" objects.

Usage

## S4 method for signature 'Genome_intervals'
interval_union(x, ...)
## S4 method for signature 'Genome_intervals_stranded'
interval_union(x, ...)

## S4 method for signature 'Genome_intervals'
interval_complement(x)
## S4 method for signature 'Genome_intervals_stranded'
interval_complement(x)

## S4 method for signature 'Genome_intervals'
interval_intersection(x,...)
## S4 method for signature 'Genome_intervals_stranded'
interval_intersection(x,...)

Arguments

x A "Genome_intervals" or "Genome_intervals_stranded" object.

... Optionally, additional objects of the same class as x.

Details

Wrappers calling the corresponding functions of the package intervals by same seq_name, inter_base and if needed strand. Note that the union of single input object x returns the reduced form of x, i.e. the interval representation of the covered set.

Value

A single object of appropriate class, representing the union, complement or intersection of intervals computed over entries with same seq_name, inter_base and also strand if all passed objects are of the class "Genome_intervals_stranded".

See Also

interval_union, interval_complement, interval_intersection and reduce from the package intervals.
Examples

```r
## load toy examples
data(gen_ints)
## content of i object
i

## complement
interval_complement(i)

## reduced form (non-overlapping interval representation of the covered set)
interval_union(i)

## union
interval_union(i[1:2,,] , i[1:4,,])

# map to genome intervals and union again
i.nostrand = as(i,"Genome_intervals")
interval_union(i.nostrand)

## intersection with a second object
# print i and j in closed interval notation
close_intervals(i)
close_intervals(j)

# interval_intersection
interval_intersection(i,,j)

# interval intersection non-stranded
interval_intersection(i.nostrand, as(j, "Genome_intervals"))
```

---

parseGffAttributes  Parse out the gffAttributes column of a Genome\_intervals object

Description

GFF files contain a string, with key/value pairs separated by “;”, and the key and value separated by “=”. This function parses such strings into a list of vectors with named elements.

Usage

```r
parseGffAttributes(gi)
```

Arguments

- `gi` A Genome\_intervals object.

Value

A list, with one element per row of gi. Each element is a character vector with named components. Names correspond to keys, and components correspond to values.
Note

Key/value pairs which are missing the "=" symbol, or which have nothing between it and the ";" delimiter or end of line, will generate a NA value, with a warning. Any key/value "pairs" with more than one "=" cause an error.

See Also

In many cases, getGffAttribute, in this package, is easier and faster. See the function readGff3 for loading a GFF file.

Examples

# Get file path
libPath <- installed.packages()["genomeIntervals", "LibPath"]
filePath <- file.path(
  libPath, 
  "genomeIntervals", 
  "example_files" 
)

# Load gff and parse attributes
gff <- readGff3( filePath, "sgd_simple.gff"), isRightOpen = FALSE )
gfatt <- parseGffAttributes(gff)

head( gfatt )
readGff3, character-method

Details

• readGff3 Make a Genome_intervals_stranded object from a gff file in gff3 format.
• readBasePairFeaturesGff3 Same as readGff3 assuming isRightOpen='FALSE', i.e. no zero length intervals are created. This is the default behaviour since v1.25.1.
• readZeroLengthFeaturesGff3 Same as readGff3 assuming isRightOpen='TRUE', i.e. zero length intervals are created when a feature’s start is the same as its end. This was the default prior to version 1.25.1.
• writeGff3 Write a Genome_intervals object to a gff file in gff3 format.

The file must follow gff3 format specifications as in http://www.sequenceontology.org/gff3.shtml. Due to the imprecise definition and to allow for zero-length features, the default for reading a Gff3 file has been to assume right open intervals (until v1.25.1). As by then, the community consensus has been to use closed intervals, the default behaviour of readGff3 has been changed accordingly. The readGff3 file is now a wrapper that dispatches to two sub functions - which may be used directly - readBasePairFeaturesGff3 and readZeroLengthFeaturesGff3. The former assumes closed intervals and hence does not create zero-length intervals. The latter does the opposite and uses right-open intervals!

Some more noteworthy details:
The file is read as a table and meta-information (lines starting with ###) are not parsed.
A ‘.’ in, for example, the gff file’s score or frame field will be converted to NA.

When the GFF file follows the right-open interval convention (isRightOpen is TRUE), then GFF entries for which end base equals first base are recognized as zero-length features and loaded as inter_base intervals.

Strand entries in the file are expected to be '.', '?', '+' or '-'. The two first are mapped to NA.

It can be that readGff3 is able to construct a Genome_intervals_stranded object from the input file, although not valid. A warning message is then generated and the constructed object is returned to allow inspection of it.

Potential FASTA entries at the end of the file are ignored.

Value

• readGff3 and friendsA Genome_intervals_stranded object image of the gff file. The GFF3 fields seqid, source, type, score, strand, phase and attributes are stored in the annotation slot and renamed as seq_name, source, type, score, strand, phase and gffAttributes respectively.
• writeGff3II dispatches to write.table and hence returns similar values.

See Also

The functions getGffAttribute and parseGffAttributes for parsing GFF attributes.

Examples

# Get file path
libPath <- installed.packages()["genomeIntervals", "LibPath"]
filePath <- file.path(
  libPath,
  "genomeIntervals",
  "example_files"
)
# Load SGD gff
# SGD does not comply to the GFF3 right-open interval convention
gff <- readGff3( file.path( filePath, "sgd_simple.gff"), isRightOpen = FALSE)

head(gff,10)

head(annotation(gff),10)

## Not run:
## write the gff3 file
writeGff3(gff,file="sgd_simple.gff")

## End(Not run)
Index

*Topic classes
  Genome_intervals-class, 8
  Genome_intervals_stranded-class, 12

*Topic datasets
gen_ints, 14

*Topic manip
  GenomeIntervals, 6

*Topic package
  genomeIntervals-package, 2
  .readGff3(readGff3, character-method), 19
  [, Genome_intervals, ANY-method
    (Genome_intervals-class), 8
  [, Genome_intervals-method
    (Genome_intervals-class), 8
  [<-,. Genome_intervals, ANY, missing, Genome_intervals-method
    (Genome_intervals-class), 8
  [[,] Genome_intervals, ANY, ANY-method
    (Genome_intervals-class), 8
  [[,] Genome_intervals, ANY-method
    (Genome_intervals-class), 8
  [[,] Genome_intervals-method
    (Genome_intervals-class), 8
  [][], Genome_intervals, ANY, ANY-method
    (Genome_intervals-class), 8
  [][,], Genome_intervals-method
    (Genome_intervals-class), 8
  $, Genome_intervals-method
    (Genome_intervals-class), 8
  $<-, Genome_intervals-method
    (Genome_intervals-class), 8

  annotation(Genome_intervals-class), 8
  annotation, Genome_intervals-method
    (Genome_intervals-class), 8
  annotation<-, Genome_intervals-class), 8
  annotation<-, Genome_intervals-method
    (Genome_intervals-class), 8

  array, 9, 13
  as (genomeIntervals coercion methods), 7
  as, Genome_intervals, GRangesList-method
    (genomeIntervals coercion methods), 7
  as, Genome_intervals, RangedData-method
    (genomeIntervals coercion methods), 7
  as, Genome_intervals-method
    (genomeIntervals coercion methods), 7

  c, 3
  c, Genome_intervals-method (c), 3
  c, Genome_intervals_stranded-method (c), 3
  coerce, Genome_intervals, character-method
    (Genome_intervals-class), 8
  coerce, Genome_intervals, data.frame-method
    (Genome_intervals-class), 8
  coerce, Genome_intervals, GRanges-method
    (genomeIntervals coercion methods), 7
  coerce, Genome_intervals, GRangesList-method
    (genomeIntervals coercion methods), 7
  coerce, Genome_intervals, Intervals_full-method
    (Genome_intervals-class), 8
  coerce, Genome_intervals, RangedData-method
    (genomeIntervals coercion methods), 7
  coerce, Genome_intervals_stranded, character-method
    (Genome_intervals_stranded-class), 12

  core_annotated, 4
  core_annotated, Genome_intervals-method
    (core_annotated), 4
  core_annotated, Genome_intervals_stranded-method
    (core_annotated), 4

  distance_to_nearest, 2, 5
  distance_to_nearest, Genome_intervals, Genome_intervals-method
    (distance_to_nearest), 5
  distance_to_nearest, Genome_intervals_stranded, Genome_intervals-method
    (distance_to_nearest), 5

  gen_ints, 2, 14
INDEX

Genome_intervals, 2–5, 7, 11–13, 15, 18–20
Genome_intervals deprecated functions, 8
Genome_intervals-class, 8
Genome_intervals-ordering, 11
Genome_intervals_stranded, 2–5, 10–12, 20
Genome_intervals_stranded-class, 12
GenomeIntervals, 6
genomeIntervals (genomeIntervals-package), 2
genomeIntervals coercion methods, 7
genomeIntervals object, 7, 8
GenomeIntervals-construct
(\texttt{GenomeIntervals}), 6
genomeIntervals-package, 2
getGffAttribute, 2, 15, 19, 20
GRangesList, 7

\texttt{i(j\texttt{gen_ints})}, 14
\texttt{inter_base(\texttt{Genome_intervals-class})}, 8
\texttt{inter_base(\texttt{Genome_intervals-method}}
(\texttt{Genome_intervals-class}), 8
\texttt{inter_base<- (\texttt{Genome_intervals-class})}, 8
\texttt{inter_base<-(\texttt{Genome_intervals-method}}
(\texttt{Genome_intervals-class}), 8
\texttt{interval_complement}, 2, 17
\texttt{interval_complement (interval_union)}, 17
\texttt{interval_complement (\texttt{Genome_intervals-method}}
(interval_union), 17
\texttt{interval_complement (\texttt{Genome_intervals_stranded-method}}
(interval_union), 17
\texttt{interval_intersection}, 2, 17
\texttt{interval_intersection (interval_union)}, 17
\texttt{interval_intersection (\texttt{Genome_intervals-method}}
(interval_union), 17
\texttt{interval_intersection (\texttt{Genome_intervals_stranded-method}}
(interval_union), 17
\texttt{interval_overlap}, 2, 16
\texttt{interval_overlap, ANY, missing-method}
(interval_overlap), 16
\texttt{interval_overlap, Genome_intervals-method}
(interval_overlap), 16
\texttt{interval_overlap, Genome_intervals_stranded-method}
(interval_overlap), 16
\texttt{interval_overlap, missing, ANY-method}
(interval_overlap), 16
\texttt{interval_union}, 2, 17, 17
\texttt{interval_union (interval_union)}, 17
\texttt{interval_union (\texttt{Genome_intervals-method}}
(interval_union), 17
\texttt{interval_union (\texttt{Genome_intervals_stranded-method}}
(interval_union), 17
\texttt{intervals}, 3
\texttt{intervals::distance_to_nearest}, 5
\texttt{intervals: interval_overlap}, 16
\texttt{Intervals_full}, 8–10, 13
\texttt{Intervals_virtal}, 9, 13

\texttt{j(gen ints)}, 14
\texttt{k(gen ints)}, 14
\texttt{matrix}, 9, 13
\texttt{order}, 12
\texttt{order (\texttt{Genome_intervals-ordering})}, 11
\texttt{order (\texttt{Genome_intervals-method}}
(\texttt{Genome_intervals-ordering}), 11
\texttt{order (\texttt{Genome_intervals_stranded-method}}
(\texttt{Genome_intervals-ordering}), 11
\texttt{parseGffAttributes}, 3, 15, 18, 20

\texttt{RangedData}, 7
\texttt{rank}, 12
\texttt{rank (\texttt{Genome_intervals-ordering})}, 11
\texttt{rank (\texttt{Genome_intervals-method}}
(\texttt{Genome_intervals-ordering}), 11
\texttt{rank (\texttt{Genome_intervals_stranded-method}}
(\texttt{Genome_intervals-ordering}), 11
\texttt{readBasePairFeaturesGff3}
(readGff3, character-method), 19
\texttt{readBasePairFeaturesGff3, character-method}
(readGff3, character-method), 19
\texttt{readGff3 (readGff3, character-method), 19
readGff3, character-method, 19

\texttt{reduce}, 17
\texttt{seq_name (\texttt{Genome_intervals deprecated functions}}
(\texttt{Genome_intervals-class}), 8
\texttt{seq_name (\texttt{Genome_intervals-method}}
(\texttt{Genome_intervals-stranded-method}), 8
\texttt{seq_name<-(\texttt{Genome_intervals deprecated functions}}, 8
\texttt{seq_name<-, \texttt{Genome_intervals-method}}
(\texttt{Genome_intervalsdeprecated functions}), 8
\texttt{seq_names, Genome_intervals-method}
(\texttt{Genome_intervals-class}), 8
seqnames<-, Genome_intervals-method
  (Genome_intervals-class), 8
show, Genome_intervals-method
  (Genome_intervals-class), 8
size, Genome_intervals-method
  (Genome_intervals-class), 8
sort, 12
sort(Genome_intervals-ordering), 11
sort, Genome_intervals-method
  (Genome_intervals-ordering), 11
sort, Genome_intervals_stranded-method
  (Genome_intervals-ordering), 11
strand
  (Genome_intervals_stranded-class), 12
strand, Genome_intervals_stranded-method
  (Genome_intervals_stranded-class), 12
strand<-
  (Genome_intervals_stranded-class), 12
strand<-, Genome_intervals_stranded,ANY-method
  (Genome_intervals-class), 8
strand<-, Genome_intervals_stranded-method
  (Genome_intervals_stranded-class), 12
structure, 9, 13
type<-, Genome_intervals-method
  (Genome_intervals-class), 8
validObject, 16
vector, 9, 13
which_nearest, Genome_intervals, Genome_intervals-method
  (Genome_intervals-class), 8
width, 10
width, Genome_intervals-method
  (Genome_intervals-class), 8
writeGff3(readGff3, character-method), 19
writeGff3, data.frame-method
  (readGff3, character-method), 19
writeGff3, Genome_intervals-method
  (readGff3, character-method), 19
xtfrm, 12
xtfrm(Genome_intervals-ordering), 11
xtfrm, Genome_intervals-method
  (Genome_intervals-ordering), 11
xtfrm, Genome_intervals_stranded-method
  (Genome_intervals-ordering), 11