Package ‘GenomicRanges’

August 31, 2017

Title Representation and manipulation of genomic intervals and variables defined along a genome

Description The ability to efficiently represent and manipulate genomic annotations and alignments is playing a central role when it comes to analyzing high-throughput sequencing data (a.k.a. NGS data). The GenomicRanges package defines general purpose containers for storing and manipulating genomic intervals and variables defined along a genome. More specialized containers for representing and manipulating short alignments against a reference genome, or a matrix-like summarization of an experiment, are defined in the GenomicAlignments and SummarizedExperiment packages respectively. Both packages build on top of the GenomicRanges infrastructure.

Version 1.28.4

Encoding UTF-8

Author P. Aboyoun, H. Pagès, and M. Lawrence

Maintainer Bioconductor Package Maintainer <maintainer@bioconductor.org>

biocViews Genetics, Infrastructure, Sequencing, Annotation, Coverage, GenomeAnnotation

Depends R (>= 2.10), methods, stats4, BiocGenerics (>= 0.21.2), S4Vectors (>= 0.9.47), IRanges (>= 2.9.11), GenomeInfoDb (>= 1.11.5)

Imports utils, stats, XVector

LinkingTo S4Vectors, IRanges

Suggests Biobase, AnnotationDbi (>= 1.21.1), annotate, Biostrings (>= 2.25.3), Rsamtools (>= 1.13.53), SummarizedExperiment (>= 0.1.5), Matrix, GenomicAlignments, rtracklayer, BSgenome, GenomicFeatures, Gviz, VariantAnnotation, AnnotationHub, DESeq2, DEXSeq, edgeR, KEGGgraph, BiocStyle, digest, RUnit, BSgenome.Hsapiens.UCSC.hg19, BSgenome.Scerevisiae.UCSC.sacCer2, KEGG.db, hgu95av2.db, org.Hs.eg.db, org.Mm.eg.db, org.Sc.sgd.db, pasilla, pasillaBamSubset, TxDb.Athaliana.BioMart.plantsmart22, TxDb.Dmelanogaster.UCSC.dm3.ensGene, TxDb.Hsapiens.UCSC.hg19.knownGene, BSgenome.Mmusculus.UCSC.mm10, TxDb.Mmusculus.UCSC.mm10.knownGene, RNAseqData.HNRNPC.bam.chr14, hgu95av2probe
absoluteRanges

Transform genomic ranges into "absolute" ranges
**Description**

`absoluteRanges` transforms the genomic ranges in `x` into *absolute* ranges i.e. into ranges counted from the beginning of the virtual sequence obtained by concatenating all the sequences in the underlying genome (in the order reported by `seqlevels(x)`).

`relativeRanges` performs the reverse transformation.

NOTE: These functions only work on *small* genomes. See Details section below.

**Usage**

```
absoluteRanges(x)
relativeRanges(x, seqlengths)
```

### Related utility:

```
isSmallGenome(seqlengths)
```

**Arguments**

- **x**
  - For `absoluteRanges`: a `GenomicRanges` object with ranges defined on a *small* genome (see Details section below).
  - For `relativeRanges`: a `Ranges` object.

- **seqlengths**
  - An object holding sequence lengths. This can be a named integer (or numeric) vector with no duplicated names as returned by `seqlengths()`, or any object from which sequence lengths can be extracted with `seqlengths()`.
  - For `relativeRanges`, `seqlengths` must describe a *small* genome (see Details section below).

**Details**

Because `absoluteRanges` returns the *absolute* ranges in an `IRanges` object, and because an `IRanges` object cannot hold ranges with an end > `.Machine$integer.max` (i.e. >= 2^31 on most platforms), `absoluteRanges` cannot be used if the size of the underlying genome (i.e. the total length of the sequences in it) is > `.Machine$integer.max`. Utility function `isSmallGenome` is provided as a mean for the user to check upfront whether the genome is *small* (i.e. its size is <= `.Machine$integer.max`) or not, and thus compatible with `absoluteRanges` or not.

`relativeRanges` applies the same restriction by looking at the `seqlengths` argument.

**Value**

- An `IRanges` object for `absoluteRanges`.
- A `GRanges` object for `relativeRanges`.

`absoluteRanges` and `relativeRanges` both return an object that is *parallel* to the input object (i.e. same length and names).

`isSmallGenome` returns TRUE if the total length of the underlying sequences is <= `.Machine$integer.max` (e.g. Fly genome), FALSE if not (e.g. Human genome), or NA if it cannot be computed (because some sequence lengths are NA).

**Author(s)**

H. Pagès
See Also

- GRanges objects.
- IRanges objects in the IRanges package.
- Seqinfo objects and the seqlengths getter in the GenomeInfoDb package.
- genomicvars for manipulating genomic variables.
- The tileGenome function for putting tiles on a genome.

Examples

```r
## TOY EXAMPLE
gr <- GRanges(Rle(c("chr2", "chr1", "chr3", "chr1"), 4:1),
              IRanges(1:10, width=5),
              seqinfo=Seqinfo(c("chr1", "chr2", "chr3"), c(100, 50, 20)))

ar <- absoluteRanges(gr)
ar

gr2 <- relativeRanges(ar, seqlengths(gr))
gr2

## Sanity check:
stopifnot(all(gr == gr2))

## ON REAL DATA

## With a "small" genome
library(TxDb.Dmelanogaster.UCSC.dm3.ensGene)
txdb <- TxDb.Dmelanogaster.UCSC.dm3.ensGene
ex <- exons(txdb)
ex

isSmallGenome(ex)

## Note that because isSmallGenome can return NA (see Value section
## above), its result should always be wrapped inside isTRUE() when
## used in an if statement:
if (isTRUE(isSmallGenome(ex))) {
  ar <- absoluteRanges(ex)
  ar

  ex2 <- relativeRanges(ar, seqlengths(ex))
  ex2  # original strand is not restored

  ## Sanity check:
  strand(ex2) <- strand(ex)  # restore the strand
  stopifnot(all(ex == ex2))
}

## With a "big" genome (but we can reduce it)
```
library(TxDb.Hsapiens.UCSC.hg19.knownGene)
txdb <- TxDb.Hsapiens.UCSC.hg19.knownGene
ex <- exons(txdb)
isSmallGenome(ex)
## Not run:
  absoluteRanges(ex)  # error!
## End(Not run)

## However, if we are only interested in some chromosomes, we might
## still be able to use absoluteRanges():
seqlevels(ex, pruning.mode="coarse") <- paste0("chr", 1:10)
isSmallGenome(ex)  # TRUE!
ar <- absoluteRanges(ex)
ex2 <- relativeRanges(ar, seqlengths(ex))

## Sanity check:
strand(ex2) <- strand(ex)
stopifnot(all(ex == ex2))

---

**Constraints**

**Enforcing constraints thru Constraint objects**

**Description**

Attaching a Constraint object to an object of class A (the "constrained" object) is meant to be a convenient/reusable/extensible way to enforce a particular set of constraints on particular instances of A.

**THIS IS AN EXPERIMENTAL FEATURE AND STILL VERY MUCH A WORK-IN-PROGRESS!**

**Details**

For the developer, using constraints is an alternative to the more traditional approach that consists in creating subclasses of A and implementing specific validity methods for each of them. However, using constraints offers the following advantages over the traditional approach:

- The traditional approach often tends to lead to a proliferation of subclasses of A.
- Constraints can easily be re-used across different classes without the need to create any new class.
- Constraints can easily be combined.

All constraints are implemented as concrete subclasses of the Constraint class, which is a virtual class with no slots. Like the Constraint virtual class itself, concrete Constraint subclasses cannot have slots.

Here are the 7 steps typically involved in the process of putting constraints on objects of class A:

1. Add a slot named constraint to the definition of class A. The type of this slot must be Constraint.OR_NULL. Note that any subclass of A will inherit this slot.
2. Implements the constraint() accessors (getter and setter) for objects of class A. This is done by implementing the "constraint" method (getter) and replacement method (setter) for objects of class A (see the examples below). As a convenience to the user, the setter should also accept the name of a constraint (i.e. the name of its class) in addition to an instance of that class. Note that those accessors will work on instances of any subclass of A.
3. Modify the validity method for class A so it also returns the result of `checkConstraint(x, constraint(x))` (append this result to the result returned by the validity method).

4. Testing: Create `x`, an instance of class A (or subclass of A). By default there is no constraint on it (`constraint(x)` is `NULL`). `validObject(x)` should return `TRUE`.

5. Create a new constraint (MyConstraint) by extending the Constraint class, typically with `setClass("MyConstraint", contains="Constraint")`. This constraint is not enforcing anything yet so you could put it on `x` (with `constraint(x) <- "MyConstraint"`), but not much would happen. In order to actually enforce something, a “checkConstraint” method for signature `c(x="A", constraint="MyConstraint")` needs to be implemented.

6. Implement a “checkConstraint” method for signature `c(x="A", constraint="MyConstraint")`. Like validity methods, “checkConstraint” methods must return `NULL` or a character vector describing the problems found. Like validity methods, they should never fail (i.e. they should never raise an error). Note that, alternatively, an existing constraint (e.g. SomeConstraint) can be adapted to work on objects of class A by just defining a new “checkConstraint” method for signature `c(x="A", constraint="SomeConstraint")`. Also, stricter constraints can be built on top of existing constraints by extending one or more existing constraints (see the examples below).

7. Testing: Try `constraint(x) <- "MyConstraint"`. It will or will not work depending on whether `x` satisfies the constraint or not. In the former case, trying to modify `x` in a way that breaks the constraint on it will also raise an error.

**Note**

WARNING: This note is not true anymore as the constraint slot has been temporarily removed from `GenomicRanges` objects (starting with package GenomicRanges >= 1.7.9).

Currently, only `GenomicRanges` objects can be constrained, that is:

- they have a constraint slot;
- they have `constraint()` accessors (getter and setter) for this slot;
- their validity method has been modified so it also returns the result of `checkConstraint(x, constraint(x))`.

More classes in the GenomicRanges and IRanges packages will support constraints in the near future.

**Author(s)**

H. Pagès

**See Also**

`setClass`, `is`, `setMethod`, `showMethods`, `validObject`, `GenomicRanges-class`

**Examples**

```r
## The examples below show how to define and set constraints on
## GenomicRanges objects. Note that this is how the constraint()
## setter is defined for GenomicRanges objects:
##setReplaceMethod("constraint", "GenomicRanges",
##  function(x, value)
##  {
##    if (isSingleString(value))
##      value <- new(value)
##    if (!is(value, "Constraint_OR_NULL"))
```
Constraints

```r
# The supplied 'constraint' must be a 
# "Constraint object, a single string, or NULL"
# x@constraint <- value
# validObject(x)
# x
# )
#
#selectMethod("constraint", "GenomicRanges") # the getter
#selectMethod("constraint<-", "GenomicRanges") # the setter

## We'll use the GRanges instance 'gr' created in the GRanges examples
## to test our constraints:
example(GRanges, echo=FALSE)
gr
#constraint(gr)

## EXAMPLE 1: The HasRangeTypeCol constraint.
## The HasRangeTypeCol constraint checks that the constrained object
## has a unique "rangeType" metadata column and that this column
## is a 'factor' Rle with no NAs and with the following levels
## (in this order): gene, transcript, exon, cds, 5utr, 3utr.

setClass("HasRangeTypeCol", contains="Constraint")

## Like validity methods, "checkConstraint" methods must return NULL or
## a character vector describing the problems found. They should never
## fail i.e. they should never raise an error.
setMethod("checkConstraint", c("GenomicRanges", "HasRangeTypeCol"),
  function(x, constraint, verbose=FALSE)
  {
    x_mcols <- mcols(x)
    idx <- match("rangeType", colnames(x_mcols))
    if (length(idx) != 1L || is.na(idx)) {
      msg <- c("mcols(x) must have exactly 1 column ",
                "named "rangeType\"")
      return(paste(msg, collapse=""))
    }
    rangeType <- x_mcols[[idx]]
    .LEVELS <- c("gene", "transcript", "exon", "cds", "5utr", "3utr")
    if (is(rangeType, "Rle") ||
        S4Vectors:::anyMissing(runValue(rangeType)) ||
        !identical(levels(rangeType), .LEVELS))
    {
      msg <- c("mcols(x)\$rangeType' must be a ",
                "'factor' Rle with no NAs and with levels: ",
                paste(.LEVELS, collapse=" "))
      return(paste(msg, collapse=""))
    }
    NULL
  } )

#dontrun{
#constraint(gr) <- "HasRangeTypeCol" # will fail
```
## Constraints

```r
#} checkConstraint(gr, new("HasRangeTypeCol"))) # with GenomicRanges >= 1.7.9

levels <- c("gene", "transcript", "exon", "cds", "5utr", "3utr")
rangeType <- Rle(factor(c("cds", "gene"), levels=levels), c(8, 2))
mcols(gr)$rangeType <- rangeType
#constraint(gr) <- "HasRangeTypeCol" # OK
checkConstraint(gr, new("HasRangeTypeCol")) # with GenomicRanges >= 1.7.9

## Use is() to check whether the object has a given constraint or not:
#is(constraint(gr), "HasRangeTypeCol") # TRUE

#} mcols(gr)$rangeType[3] <- NA # will fail
#
mcols(gr)$rangeType[3] <- NA
checkConstraint(gr, new("HasRangeTypeCol")) # with GenomicRanges >= 1.7.9

## EXAMPLE 2: The GeneRanges constraint.
## The GeneRanges constraint is defined on top of the HasRangeTypeCol constraint. It checks that all the ranges in the object are of type "gene".

setClass("GeneRanges", contains="HasRangeTypeCol")

## The checkConstraint() generic will check the HasRangeTypeCol constraint first, and, only if it's satisfied, it will then check the GeneRanges constraint.
setMethod("checkConstraint", c("GenomicRanges", "GeneRanges"),
  function(x, constraint, verbose=FALSE)
  {
    rangeType <- mcols(x)$rangeType
    if (!all(rangeType == "gene")) {
      msg <- c("all elements in "mcols(x)$rangeType", 
                "must be equal to \"gene\"")
      return(paste(msg, collapse=""))
    }
    NULL
  })

#} constraint(gr) <- "GeneRanges" # will fail
#
checkConstraint(gr, new("GeneRanges")) # with GenomicRanges >= 1.7.9

mcols(gr)$rangeType[] <- "gene"
# This replace the previous constraint (HasRangeTypeCol):
#constraint(gr) <- "GeneRanges" # OK
checkConstraint(gr, new("GeneRanges")) # with GenomicRanges >= 1.7.9

#is(constraint(gr), "GeneRanges") # TRUE
# However, 'gr' still indirectly has the HasRangeTypeCol constraint
# (because the GeneRanges constraint extends the HasRangeTypeCol constraint):
#is(constraint(gr), "HasRangeTypeCol") # TRUE
```
### Constraints

```r
#
dontrun{
#mcols(gr)$rangeType[] <- "exon" # will fail #}
mcols(gr)$rangeType[] <- "exon"
checkConstraint(gr, new("GeneRanges")) # with GenomicRanges >= 1.7.9
}

## EXAMPLE 3: The HasGCCol constraint.
## ---------------------------------------------------------------------
## The HasGCCol constraint checks that the constrained object has a
## unique "GC" metadata column, that this column is of type numeric,
## with no NAs, and that all the values in that column are >= 0 and <= 1.

setClass("HasGCCol", contains="Constraint")
setMethod("checkConstraint", c("GenomicRanges", "HasGCCol"),
function(x, constraint, verbose=FALSE)
{
  x_mcols <- mcols(x)
  idx <- match("GC", colnames(x_mcols))
  if (length(idx) != 1L || is.na(idx)) {
    msg <- c("mcols(x) must have exactly ",
             "one column named "GC")
    return(paste(msg, collapse=""))
  }
  GC <- x_mcols[[idx]]
  if (!is.numeric(GC) ||
      S4Vectors:::anyMissing(GC) ||
      any(GC < 0) || any(GC > 1))
  {
    msg <- c("mcols(x)$GC must be a numeric vector ",
             "with no NAs and with values between 0 and 1")
    return(paste(msg, collapse=""))
  }
  NULL
}
)

## This replace the previous constraint (GeneRanges):
#constraint(gr) <- "HasGCCol" # OK
ccheckConstraint(gr, new("HasGCCol")) # with GenomicRanges >= 1.7.9

#is(constraint(gr), "HasGCCol") # TRUE
#is(constraint(gr), "GeneRanges") # FALSE
#is(constraint(gr), "HasRangeTypeCol") # FALSE

## EXAMPLE 4: The HighGCRanges constraint.
## ---------------------------------------------------------------------
## The HighGCRanges constraint is defined on top of the HasGCCol
## constraint. It checks that all the ranges in the object have a GC
## content >= 0.5.

setClass("HighGCRanges", contains="HasGCCol")

# The checkConstraint() generic will check the HasGCCol constraint
# first, and, if it's satisfied, it will then check the HighGCRanges
```
## Constraint.

```r
setMethod("checkConstraint", c("GenomicRanges", "HighGCRanges"),

function(x, constraint, verbose=FALSE)
{
  GC <- mcols(x)$GC
  if (!all(GC >= 0.5)) {
    msg <- c("all elements in 'mcols(x)$GC' ",
             "must be => 0.5")
    return(paste(msg, collapse=""))
  }
  NULL
}
)
```

```r
#dontrun{
#constraint(gr) <- "HighGCRanges" # will fail
#}
checkConstraint(gr, new("HighGCRanges")) # with GenomicRanges >= 1.7.9
mcols(gr)$GC[6:10] <- 0.5
#constraint(gr) <- "HighGCRanges" # OK
checkConstraint(gr, new("HighGCRanges")) # with GenomicRanges >= 1.7.9

#is(constraint(gr), "HighGCRanges") # TRUE
#is(constraint(gr), "HasGCCol") # TRUE
```

```r
## EXAMPLE 5: The HighGCGeneRanges constraint.

## The HighGCGeneRanges constraint is the combination (AND) of the
## GeneRanges and HighGCRanges constraints.

setClass("HighGCGeneRanges", contains=c("GeneRanges", "HighGCRanges"))

## No need to define a method for this constraint: the checkConstraint()
## generic will automatically check the GeneRanges and HighGCRanges
## constraints.

#constraint(gr) <- "HighGCGeneRanges" # OK
checkConstraint(gr, new("HighGCGeneRanges")) # with GenomicRanges >= 1.7.9

#is(constraint(gr), "HighGCGeneRanges") # TRUE
#is(constraint(gr), "HighGCRanges") # TRUE
#is(constraint(gr), "GeneRanges") # TRUE
#is(constraint(gr), "HasRangeTypeCol") # TRUE

## See how all the individual constraints are checked (from less
## specific to more specific constraints):
#checkConstraint(gr, constraint(gr), verbose=TRUE)
checkConstraint(gr, new("HighGCGeneRanges"), verbose=TRUE) # with
  # GenomicRanges
  # >= 1.7.9

## See all the "checkConstraint" methods:
showMethods("checkConstraint")
```
coverage-methods

Coverage of a GRanges or GRangesList object

Description

coverage methods for GRanges and GRangesList objects.

NOTE: The coverage generic function and methods for Ranges and RangesList objects are defined and documented in the IRanges package. Methods for GAlignments and GAlignmentPairs objects are defined and documented in the GenomicAlignments package.

Usage

## S4 method for signature 'GenomicRanges'
coverage(x, shift=0L, width=NULL, weight=1L, method=c("auto", "sort", "hash"))

## S4 method for signature 'GRangesList'
coverage(x, shift=0L, width=NULL, weight=1L, method=c("auto", "sort", "hash"))

Arguments

x A GRanges or GRangesList object.

shift A numeric vector or a list-like object. If numeric, it must be parallel to x (recycled if necessary). If a list-like object, it must have 1 list element per seqlevel in x, and its names must be exactly seqlevels(x). Alternatively, shift can also be specified as a single string naming a metadata column in x (i.e. a column in mcols(x)) to be used as the shift vector. See ?coverage in the IRanges package for more information about this argument.

width Either NULL (the default), or an integer vector. If NULL, it is replaced with seqlengths(x). Otherwise, the vector must have the length and names of seqlengths(x) and contain NAs or non-negative integers. See ?coverage in the IRanges package for more information about this argument.

weight A numeric vector or a list-like object. If numeric, it must be parallel to x (recycled if necessary). If a list-like object, it must have 1 list element per seqlevel in x, and its names must be exactly seqlevels(x). Alternatively, weight can also be specified as a single string naming a metadata column in x (i.e. a column in mcols(x)) to be used as the weight vector. See ?coverage in the IRanges package for more information about this argument.

method See ?coverage in the IRanges package for a description of this argument.

Details

When x is a GRangesList object, coverage(x, ...) is equivalent to coverage(unlist(x), ...).
DelegatingGenomicRanges-class

Value

A named RleList object with one coverage vector per seqlevel in x.

Author(s)

H. Pagès and P. Aboyoun

See Also

- coverage in the IRanges package.
- coverage-methods in the GenomicAlignments package.
- RleList objects in the IRanges package.
- GRanges and GRangesList objects.

Examples

```r
## Coverage of a GRanges object:
gr <- GRanges(
  seqnames=Rle(c("chr1", "chr2", "chr1", "chr3"), c(1, 3, 2, 4)),
  ranges=IRanges(1:10, end=10),
  strand=Rle(strand(c("-", "+", "+", "+", "-")), c(1, 2, 2, 3, 2)),
  seqlengths=c(chr1=11, chr2=12, chr3=13))
cvg <- coverage(gr)
pcvg <- coverage(gr[strand(gr) == "+"])
mcvg <- coverage(gr[strand(gr) == "-"])
scvg <- coverage(gr[strand(gr) == "+"])
stopifnot(identical(pcvg + mcvg + scvg, cvg))

## Coverage of a GRangesList object:
gr1 <- GRanges(seqnames="chr2",
  ranges=IRanges(3, 6),
  strand="+"
) gr2 <- GRanges(seqnames=c("chr1", "chr1"),
  ranges=IRanges(c(7,13), width=3),
  strand=c("+", "+")
) gr3 <- GRanges(seqnames=c("chr1", "chr2"),
  ranges=IRanges(c(1, 4), c(3, 9)),
  strand=c("-", "+")
) grl <- GRangesList(gr1=gr1, gr2=gr2, gr3=gr3)
stopifnot(identical(coverage(grl), coverage(unlist(grl))))
```

DelegatingGenomicRanges-class

DelegatingGenomicRanges objects

Description

The DelegatingGenomicRanges class implements the virtual GenomicRanges class using a delegate GenomicRanges. This enables developers to create GenomicRanges subclasses that add specialized columns or other structure, while remaining agnostic to how the data are actually stored. See the Extending GenomicRanges vignette.
findOverlaps-methods

Description

Various methods for finding/counting overlaps between objects containing genomic ranges. This man page describes the methods that operate on GenomicRanges and GRangesList objects.

NOTE: The `findOverlaps` generic function and methods for Ranges and RangesList objects are defined and documented in the IRanges package. The methods for GAlignments, GAlignmentPairs, and GAlignmentsList objects are defined and documented in the GenomicAlignments package.

GenomicRanges and GRangesList objects also support countOverlaps, overlapsAny, and subsetByOverlaps thanks to the default methods defined in the IRanges package and to the findOverlaps and countOverlaps methods defined in this package and documented below.

Usage

```r
## S4 method for signature 'GenomicRanges,GenomicRanges'
findOverlaps(query, subject,
             maxgap=0L, minoverlap=1L,
             type=c("any", "start", "end", "within", "equal"),
             select=c("all", "first", "last", "arbitrary"),
             ignore.strand=FALSE)

## S4 method for signature 'GenomicRanges,GenomicRanges'
countOverlaps(query, subject,
               maxgap=0L, minoverlap=1L,
               type=c("any", "start", "end", "within", "equal"),
               ignore.strand=FALSE)
```

Arguments

- `query, subject` A GRanges or GRangesList object.
- `maxgap, minoverlap`  See `findOverlaps` in the IRanges package for a description of these arguments.
- `type`  When select is "all" (the default), the results are returned as a Hits object. Otherwise the returned value is an integer vector parallel to query (i.e. same length) containing the first, last, or arbitrary overlapping interval in subject, with NA indicating intervals that did not overlap any intervals in subject.
- `select`  When set to TRUE, the strand information is ignored in the overlap calculations.

Details

When the query and the subject are GRanges or GRangesList objects, findOverlaps uses the triplet (sequence name, range, strand) to determine which features (see paragraph below for the definition of feature) from the query overlap which features in the subject, where a strand value of "*" is treated as occurring on both the "+" and "-" strand. An overlap is recorded when a
feature in the query and a feature in the subject have the same sequence name, have a compatible pairing of strands (e.g. "+"/"-", "+"/"+", "-"/"-", "*"/"+", "*"/"-", etc.), and satisfy the interval overlap requirements.

In the context of `findOverlaps`, a feature is a collection of ranges that are treated as a single entity. For `GRanges` objects, a feature is a single range; while for `GRangesList` objects, a feature is a list element containing a set of ranges. In the results, the features are referred to by number, which run from 1 to `length(query)/length(subject)`.

**Value**

For `findOverlaps` either a `Hits` object when `select="all"` or an integer vector otherwise.

For `countOverlaps` an integer vector containing the tabulated query overlap hits.

**Author(s)**

P. Aboyoun, S. Falcon, M. Lawrence, and H. Pagès

**See Also**

- The `Hits` class for representing a set of hits between 2 vector-like objects.
- The `findOverlaps` generic function defined in the `IRanges` package.
- The `GNCList` constructor and class for preprocessing and representing a GenomicRanges or object as a data structure based on Nested Containment Lists.
- The `GRanges` and `GRangesList` classes.

**Examples**

```r
## BASIC EXAMPLES

## GRanges object:
gr <- GRanges(
  seqnames=Rle(c("chr1", "chr2", "chr3"), c(1, 3, 2, 4)),
  ranges=IRanges(1:10, width=10:1, names=head(letters,10)),
  strand=Rle(strand(c("-", "+", "#", "#", "-")), c(1, 2, 3, 2)),
  score=1:10,
  GC=seq(1, 0, length=10)
)
gr

## GRangesList object:
gr1 <- GRanges(seqnames="chr2", ranges=IRanges(4:3, 6),
  strand="+", score=5:4, GC=0.45)
gr2 <- GRanges(seqnames=c("chr1", "chr1"),
  ranges=IRanges(c(7,13), width=3),
  strand=c("+", "-"), score=3:4, GC=c(0.3, 0.5))
gr3 <- GRanges(seqnames=c("chr1", "chr2"),
  ranges=IRanges(c(1, 4), c(3, 9)),
  strand=c("-", "-"), score=c(6L, 2L), GC=c(0.4, 0.1))
gr1 <- GRangesList("gr1"=gr1, "gr2"=gr2, "gr3"=gr3)

## Overlapping two GRanges objects:
table(!is.na(findOverlaps(gr, gr1, select="arbitrary")))
```
countOverlaps(gr, gr1)
findOverlaps(gr, gr1)
subsetByOverlaps(gr, gr1)

countOverlaps(gr, gr1, type="start")
findOverlaps(gr, gr1, type="start")
subsetByOverlaps(gr, gr1, type="start")

findOverlaps(gr, gr1, select="first")
findOverlaps(gr, gr1, select="last")

findOverlaps(gr1, gr)
findOverlaps(gr1, gr, type="start")
findOverlaps(gr1, gr, type="within")
findOverlaps(gr1, gr, type="equal")

## MORE EXAMPLES
##

table(!is.na(findOverlaps(gr, gr1, select="arbitrary")))
countOverlaps(gr, gr1)
findOverlaps(gr, gr1)
subsetByOverlaps(gr, gr1)

## Overlaps between a GRanges and a GRangesList object:

table(!is.na(findOverlaps(grl, gr, select="first")))
countOverlaps(grl, gr)
findOverlaps(grl, gr)
subsetByOverlaps(grl, gr)
countOverlaps(grl, gr, type="start")
findOverlaps(grl, gr, type="start")
subsetByOverlaps(grl, gr, type="start")
findOverlaps(grl, gr, select="first")

table(!is.na(findOverlaps(grl, gr1, select="first")))
countOverlaps(grl, gr1)
findOverlaps(grl, gr1)
subsetByOverlaps(grl, gr1)
countOverlaps(grl, gr1, type="start")
findOverlaps(grl, gr1, type="start")
subsetByOverlaps(grl, gr1, type="start")
findOverlaps(grl, gr1, select="first")

## Overlaps between two GRangesList objects:
countOverlaps(grl, rev(grl))
findOverlaps(grl, rev(grl))
subsetByOverlaps(grl, rev(grl))

GenomicRanges-comparison

Comparing and ordering genomic ranges
Description

Methods for comparing and ordering the elements in one or more GenomicRanges objects.

Usage

```r
## duplicated()
## ------------
## S4 method for signature 'GenomicRanges'
duplicated(x, incomparables=FALSE, fromLast=FALSE,
          nmax=NA, method=c("auto", "quick", "hash"))

## match() & selfmatch()
## ---------------------
## S4 method for signature 'GenomicRanges,GenomicRanges'
match(x, table, nomatch=NA_integer_, incomparables=NULL,
      method=c("auto", "quick", "hash"), ignore.strand=FALSE)

## S4 method for signature 'GenomicRanges'
selfmatch(x, method=c("auto", "quick", "hash"), ignore.strand=FALSE)

## order() and related methods
## ----------------------------
## S4 method for signature 'GenomicRanges'
is.unsorted(x, na.rm=FALSE, strictly=FALSE, ignore.strand=FALSE)

## S4 method for signature 'GenomicRanges'
order(..., na.last=TRUE, decreasing=FALSE, method=c("shell", "radix"))

## S4 method for signature 'GenomicRanges'
sort(x, decreasing=FALSE, ignore.strand=FALSE, by)

## S4 method for signature 'GenomicRanges'
rank(x, na.last=TRUE,
     ties.method=c("average", "first", "last", "random", "max", "min"),
     ignore.strand=FALSE)

## Generalized parallel comparison of 2 GenomicRanges objects
## ----------------------------------------------------------
## S4 method for signature 'GenomicRanges,GenomicRanges'
pcompare(x, y)
```

Arguments

- `x`, `table`, `y` GenomicRanges objects.
- `incomparables` Not supported.
- `fromLast`, `method`, `nomatch`, `nmax`
  See ?'Ranges-comparison' in the IRanges package for a description of these arguments.
ignore.strand  Whether or not the strand should be ignored when comparing 2 genomic ranges.
na.rm              Ignored.
strictly           Logical indicating if the check should be for strictly increasing values.
...                One or more GenomicRanges objects. The GenomicRanges objects after the first one are used to break ties.
na.last             Ignored.
decreasing          TRUE or FALSE.
ties.method         A character string specifying how ties are treated. Only "first" is supported for now.
by                  An optional formula that is resolved against as.env(x); the resulting variables are passed to order to generate the ordering permutation.

Details
Two elements of a GenomicRanges object (i.e. two genomic ranges) are considered equal iff they are on the same underlying sequence and strand, and have the same start and width. duplicated() and unique() on a GenomicRanges object are conforming to this.

The "natural order" for the elements of a GenomicRanges object is to order them (a) first by sequence level, (b) then by strand, (c) then by start, (d) and finally by width. This way, the space of genomic ranges is totally ordered. Note that the reduce method for GenomicRanges uses this "natural order" implicitly. Also, note that, because we already do (c) and (d) for regular ranges (see ?Ranges-comparison), genomic ranges that belong to the same underlying sequence and strand are ordered like regular ranges.

is.unsorted(), order(), sort(), and rank() on a GenomicRanges object behave accordingly to this "natural order".

Author(s)
H. Pagès, is.unsorted contributed by Pete Hickey

See Also
• The GenomicRanges class.
• Ranges-comparison in the IRanges package for comparing and ordering genomic ranges.
• findOverlaps-methods for finding overlapping genomic ranges.
• intra-range-methods and inter-range-methods for intra range and inter range transformations of a GRanges object.
• setops-methods for set operations on GenomicRanges objects.

Examples
```
gr0 <- GRanges(
  Rle(c("chr1", "chr2", "chr1", "chr3"), c(1, 3, 2, 4)),
  IRanges(c(1:9L), end=10),
  strand=Rle(strand(c("-", "+", "*", "*", "-")), c(1, 2, 2, 3, 2)),
  seqlengths=c(chr1=11, chr2=12, chr3=13)
  )
gr <- c(gr0, gr0[7:3])
names(gr) <- LETTERS[seq_along(gr)]
```
## A. ELEMENT-WISE (AKA "PARALLEL") COMPARISON OF 2 GenomicRanges OBJECTS

gr == gr[4]
gr >= gr[3]

## B. duplicated(), unique()

duplicated(gr)
unique(gr)

## C. match(), %in%

table <- gr[1:7]
match(gr, table)
match(gr, table, ignore.strand=TRUE)
gr %in% table

## D. findMatches(), countMatches()

findMatches(gr, table)
countMatches(gr, table)

findMatches(gr, table, ignore.strand=TRUE)
countMatches(gr, table, ignore.strand=TRUE)
gr_levels <- unique(gr)
countMatches(gr_levels, gr)

## E. order() AND RELATED METHODS

is.unsorted(gr)
order(gr)
sort(gr)
is.unsorted(sort(gr))

is.unsorted(gr, ignore.strand=TRUE)
gr2 <- sort(gr, ignore.strand=TRUE)
is.unsorted(gr2)  # TRUE
is.unsorted(gr2, ignore.strand=TRUE)  # FALSE

## TODO: Broken. Please fix!
#sort(gr, by = ~ seqnames + start + end)  # equivalent to (but slower than) above

cscore(gr) <- rev(seq_len(length(gr)))

## TODO: Broken. Please fix!
#sort(gr, by = ~ score)
GenomicRangesList-class

GenomicRangesList objects

Description

A GenomicRangesList is a List of GenomicRanges. It is a virtual class: SimpleGenomicRangesList is the basic implementation. The subclass GRangesList provides special behavior and is particularly efficient for storing a large number of elements.

Constructor

GenomicRangesList(...): Constructs a SimpleGenomicRangesList with elements taken from the arguments in .... If the only argument is a list, the elements are taken from that list.

Author(s)

Michael Lawrence

See Also

GRangesList, which differs from SimpleGenomicRangesList in that the GRangesList treats its elements as single, compound ranges, particularly in overlap operations. SimpleGenomicRangesList is just a barebones list for now, without that compound semantic.

genomicvars

Manipulating genomic variables

Description

A genomic variable is a variable defined along a genome. Here are 2 ways a genomic variable is generally represented in Bioconductor:

1. as a named RleList object with one list element per chromosome;
2. as a metadata column on a disjoint GRanges object.

This man page documents tools for switching from one form to the other.
Usage

bindAsGRanges(...)  
mcolAsRleList(x, varname)  
binnedAverage(bins, numvar, varname, na.rm=FALSE)

Arguments

... One or more genomic variables in the form of named RleList objects.

x A disjoint GRanges object with metadata columns on it. A GRanges object is said to be disjoint if it contains ranges that do not overlap with each other. This can be tested with isDisjoint. See ?isDisjoint,GenomicRanges-method for more information about the isDisjoint method for GRanges objects.

varname The name of the genomic variable. For mcolAsRleList this must be the name of the metadata column on x to be turned into an RleList object. For binnedAverage this will be the name of the metadata column that contains the binned average in the returned object.

bins A GRanges object representing the genomic bins. Typically obtained by calling tileGenome with cut.last.tile.in.chrom=TRUE.

numvar A named RleList object representing a numerical variable defined along the genome covered by bins (which is the genome described by seqinfo(bins)).

na.rm A logical value indicating whether NA values should be stripped before the average is computed.

Details

bindAsGRanges allows to switch the representation of one or more genomic variables from the named RleList form to the metadata column on a disjoint GRanges object form by binding the supplied named RleList objects together and putting them on the same GRanges object. This transformation is lossless.

mcolAsRleList performs the opposite transformation and is also lossless (however the circularity flags and genome information in seqinfo(x) won’t propagate). It works for any metadata column on x that can be put in Rle form i.e. that is an atomic vector or a factor.

binnedAverage computes the binned average of a numerical variable defined along a genome.

Value

For bindAsGRanges: a GRanges object with 1 metadata column per supplied genomic variable.

For mcolAsRleList: a named RleList object with 1 list element per seqlevel in x.

For binnedAverage: input GRanges object bins with an additional metadata column named varname containing the binned average.

Author(s)

H. Pagès
See Also

- `RleList` objects in the `IRanges` package.
- `coverage.GenomicRanges-method` for computing the coverage of a `GRanges` object.
- The `tileGenome` function for putting tiles on a genome.
- `GRanges` objects and `isDisjoint.GenomicRanges-method` for the `isDisjoint` method for `GenomicRanges` objects.

Examples

```r
## ---------------------------------------------------------------------
## A. TWO WAYS TO REPRESENT A GENOMIC VARIABLE
## -----------------------------------------------------------------
##
## 1) As a named RleList object
##
## Let's create a genomic variable in the "named RleList" form:
library(BSgenome.Scerevisiae.UCSC.sacCer2)
set.seed(55)
my_var <- RleList(
  lapply(seqlengths(Scerevisiae),
    function(seqlen) {
      tmp <- sample(50L, seqlen, replace=TRUE)
      Rle(cumsum(tmp - rev(tmp)))
    }
  ),
  compress=FALSE)
my_var

## 2) As a metadata column on a disjoint GRanges object
##
## Going back to the "named RleList" form:
mcolAsRleList(gr1, "my_var")

## Bind 'my_var', 'score', and 'id' together:
gr3 <- bindAsGRanges(my_var=my_var, score=score, id=id)
```
stopifnot(identical(score, mcolAsRleList(gr3, "score")))
stopifnot(identical(id, mcolAsRleList(gr3, "id")))
gr2b <- bindAsGRanges(score=score, id=id)
seqinfo(gr2b) <- seqinfo(gr2)
stopifnot(identical(gr2, gr2b))

## B. BIND TOGETHER THE COVERAGES OF SEVERAL BAM FILES

library(pasillaBamSubset)
library(GenomicAlignments)
untreated1_cvg <- coverage(BamFile(untreated1_chr4()))
untreated3_cvg <- coverage(BamFile(untreated3_chr4()))
all_cvg <- bindAsGRanges(untreated1=untreated1_cvg,
                         untreated3=untreated3_cvg)

## Keep regions with coverage:
all_cvg[with(mcols(all_cvg), untreated1 + untreated3 >= 1)]

## Plot the coverage profiles with the Gviz package:
library(Gviz)
plotNumvars <- function(numvars, region, name="numvars", ...) {
  stopifnot(is(numvars, "GRanges"))
  stopifnot(is(region, "GRanges"), length(region) == 1L)
  gtrack <- GenomeAxisTrack()
  dtrack <- DataTrack(numvars,
                       chromosome=as.character(seqnames(region)),
                       name=name,
                       groups=colnames(mcols(numvars)), type="l", ...)
  plotTracks(list(gtrack, dtrack), from=start(region), to=end(region))
}
plotNumvars(all_cvg, GRanges("chr4:1-25000"),
            "coverage", col=c("red", "blue"))
plotNumvars(all_cvg, GRanges("chr4:1.03e6-1.08e6"),
            "coverage", col=c("red", "blue"))

## Sanity checks:
stopifnot(identical(untreated1_cvg, mcolAsRleList(all_cvg, "untreated1")))
stopifnot(identical(untreated3_cvg, mcolAsRleList(all_cvg, "untreated3")))

## C. COMPUTE THE BINNED AVERAGE OF A NUMERICAL VARIABLE DEFINED ALONG A
## GENOME

## In some applications (e.g. visualization), there is the need to compute
## the average of a genomic variable for a set of predefined fixed-width
## regions (sometimes called "bins").
## Let’s use tileGenome() to create such a set of bins:
bins1 <- tileGenome(seqinfo(Scerevisiae), tilewidth=100,
                     cut.last.tile.in.chrom=TRUE)

## Compute the binned average for 'my_var' and 'score':
bins1 <- binnedAverage(bins1, my_var, "binned_var")
bins1
bins1 <- binnedAverage(bins1, score, "binned_score")
bins1

## Binned average in "named RleList" form:
binned_var1 <- mcolAsRleList(bins1, "binned_var")
binned_var1
stopifnot(all.equal(mean(my_var), mean(binned_var1))) # sanity check

mcolAsRleList(bins1, "binned_score")

## With bigger bins:
bins2 <- tileGenome(seqinfo(Scerevisiae), tilewidth=50000,
cut.last.tile.in.chrom=TRUE)
bins2 <- binnedAverage(bins2, my_var, "binned_var")
bins2 <- binnedAverage(bins2, score, "binned_score")
bins2
binned_var2 <- mcolAsRleList(bins2, "binned_var")
binned_var2
stopifnot(all.equal(mean(my_var), mean(binned_var2))) # sanity check

mcolAsRleList(bins2, "binned_score")

## Not surprisingly, the "binned" variables are much more compact in
## memory than the original variables (they contain much less runs):
object.size(my_var)
object.size(binned_var1)
object.size(binned_var2)

## D. SANITY CHECKS

### GNCList-class

**GNCList objects**

**Description**

The GNCList class is a container for storing the Nested Containment List representation of a vector of genomic ranges (typically represented as a GRanges object). To preprocess a GRanges object,
simply call the GNCList constructor function on it. The resulting GNCList object can then be used for efficient overlap-based operations on the genomic ranges.

Usage

GNCList(x)

Arguments

x The GRanges (or more generally GenomicRanges) object to preprocess.

Details

The IRanges package also defines the NList and NLists constructors and classes for preprocessing and representing a Ranges or RangesList object as a data structure based on Nested Containment Lists.

Note that GNCList objects (introduced in BioC 3.1) are replacements for GIntervalTree objects (BioC < 3.1).

See ?NCList in the IRanges package for some important differences between the new algorithm based on Nested Containment Lists and the old algorithm based on interval trees. In particular, the new algorithm supports preprocessing of a GenomicRanges object with ranges defined on circular sequences (e.g. on the mitochondrial chromosome). See below for some examples.

Value

A GNCList object.

Author(s)

H. Pagès

References


See Also

• The NList and NLists constructors and classs defined in the IRanges package.
• findOverlaps for finding/counting interval overlaps between two range-based objects.
• GRanges objects.

Examples

## The examples below are for illustration purpose only and do NOT
## reflect typical usage. This is because, for a one time use, it is
## NOT advised to explicitly preprocess the input for findOverlaps()
## or countOverlaps(). These functions will take care of it and do a
## better job at it (by preprocessing only what's needed when it's
## needed, and release memory as they go).

## ---------------------------------------------------------------------
## PREPROCESS QUERY OR SUBJECT
## ---------------------------------------------------------------------
query <- GRanges(Rle(c("chrM", "chr1", "chrM", "chr1"), 4:1),
  IRanges(1:10, width=5), strand=rep(c("+", "-"), 5))
subject <- GRanges(Rle(c("chr1", "chr2", "chrM"), 3:1),
  IRanges(6:1, width=5), strand="+")

## Either the query or the subject of findOverlaps() can be preprocessed:
ppsubject <- GNCList(subject)
hits1a <- findOverlaps(query, ppsubject)
hits1b <- findOverlaps(query, ppsubject, ignore.strand=TRUE)

ppquery <- GNCList(query)
hits2a <- findOverlaps(ppquery, subject)
hits2b <- findOverlaps(ppquery, subject, ignore.strand=TRUE)

## Note that `hits1a` and `hits2a` contain the same hits but not
## necessarily in the same order.
stopifnot(identical(sort(hits1a), sort(hits2a)))
## Same for `hits1b` and `hits2b`.
stopifnot(identical(sort(hits1b), sort(hits2b)))

## WITH CIRCULAR SEQUENCES
## ---------------------------------------------------------------------
seqinfo <- Seqinfo(c("chr1", "chr2", "chrM"),
  seqlengths=c(100, 50, 10),
  isCircular=c(FALSE, FALSE, TRUE))
seqinfo(query) <- seqinfo[seqlevels(query)]
seqinfo(subject) <- seqinfo[seqlevels(subject)]

ppsubject <- GNCList(subject)
hits3 <- findOverlaps(query, ppsubject)

## Circularity introduces more hits:
stopifnot(all(hits1a %in% hits3))
new_hits <- setdiff(hits3, hits1a)
new_hits # 1 new hit
query[queryHits(new_hits)]
subject[subjectHits(new_hits)] # positions 11:13 on chrM are the same
# as positions 1:3

## Sanity checks:
stopifnot(identical(new_hits, Hits(9, 6, 10, 6, sort.by=query=TRUE)))
ppquery <- GNCList(query)
hits4 <- findOverlaps(ppquery, subject)
stopifnot(identical(sort(hits3), sort(hits4)))
GPos-class

Description
The GPos class is a container for storing a set of genomic *positions*, that is, genomic ranges of width 1. Even though a GRanges object can be used for that, using a GPos object can be much more memory-efficient, especially when the object contains long runs of adjacent positions.

Usage
```
GPos(pos_runs)  # constructor function
```

Arguments
- `pos_runs` A GRanges object (or any other GenomicRanges derivative) where each range is interpreted as a run of adjacent genomic positions. If `pos_runs` is not a GenomicRanges object, GPos() first tries to coerce it to one with `as(pos_runs, "GenomicRanges", strict=FALSE)`.

Value
A GPos object.

Accessors
**Getters:** GPos objects support the same set of getters as GRanges objects (i.e. `seqnames()`, `start()`, `end()`, `ranges()`, `strand()`, `mcols()`, `seqinfo()`, etc...), plus the `pos()` getter which is equivalent to `start()` or `end()`. See ?GRanges for the list of getters supported by GRanges objects.

Note that a GPos object cannot hold names i.e. `names()` always returns NULL on it.

**Setters:** Like GRanges objects, GPos objects support the following setters:
- The `mcols()` and `metadata()` setters.
- The family of setters that operate on the seqinfo component of an object: `seqlevels()`, `seqlevelsStyle()`, `seqlengths()`, `isCircular()`, `genome()`, and `seqinfo()`. These setters are defined and documented in the GenomeInfoDb package.

However, there is no `seqnames()`, `pos()`, or `strand()` setter for GPos objects at the moment (although they might be added in the future).

Coercion
From GenomicRanges to GPos: A GenomicRanges object `x` in which all the ranges have a width of 1 can be coerced to a GPos object with `as(x, "GPos")`. The names on `x` are not propagated (a warning is issued if `x` has names on it).

From GPos to GRanges: A GPos object `x` can be coerced to a GRanges object with `as(x, "GRanges")`. However be aware that the resulting object can use thousands times (or more) memory than `x`! See "MEMORY USAGE" in the Examples section below.

From GPos to ordinary R objects: Like with a GRanges object, `as.character()`, `as.factor()`, and `as.data.frame()` work with a GPos object `x`. Note that `as.data.frame(x)` returns a data frame with a `pos` column (containing `pos(x)`) instead of the `start`, `end`, and `width` columns that one gets when `x` is a GRanges object.
Subsetting

A GPos object can be subsetted exactly like a GRanges object.

Combining

GPos objects can be combined (a.k.a. appended) with c() or append().

Splitting and Relisting

Like with a GRanges object, split() and relist() work with a GPos object x. Note that they return a GenomicRangesList object instead of a GRangesList object.

Note

Like for any Vector derivative, the length of a GPos object cannot exceed .Machine$integer.max (i.e. 2^31 on most platforms). GPos() will return an error if pos_runs contains too many genomic positions.

Author(s)

Hervé Pagès; based on ideas borrowed from Georg Stricker <georg.stricker@in.tum.de> and Julien Gagneur <gagneur@in.tum.de>

See Also

• GRanges objects.
• The seqinfo accessor and family in the GenomeInfoDb package for accessing/modifying the seqinfo component of an object.
• GenomicRanges-comparison for comparing and ordering genomic positions.
• findOverlaps-methods for finding overlapping genomic ranges and/or positions.
• nearest-methods for finding the nearest genomic range/position neighbor.
• The snpsBySeqname, snpsByOverlaps, and snpsById methods for SNPlocs objects defined in the BSgenome package for extractors that return a GPos object.
• SummarizedExperiment objects in the SummarizedExperiment package.

Examples

```r
## Example 1:
gpos1 <- GPos(c("chr1:44-53", "chr1:5-10", "chr2:2-5"))
gpos1

length(gpos1)
seqnames(gpos1)
pos(gpos1) # same as 'start(gpos1)' and 'end(gpos1)'
strand(gpos1)
as.character(gpos1)
as.data.frame(gpos1)
as(gpos1, "GRanges")
as.data.frame(as(gpos1, "GRanges"))
```
Example 2:

```r
gpos2 <- GPos(pos_runs)
gpos2
```

## Example 3:

```r
gpos3A <- gpos3B <- GPos(c("chrI:1-1000", "chrI:1005-2000"))
npos <- length(gpos3A)
mcols(gpos3A)$sample <- Rle("sA")
sA_counts <- sample(10, npos, replace=TRUE)
mcols(gpos3A)$counts <- sA_counts
mcols(gpos3B)$sample <- Rle("sB")
sB_counts <- sample(10, npos, replace=TRUE)
mcols(gpos3B)$counts <- sB_counts
gpos3 <- c(gpos3A, gpos3B)
gpos3
```

## Example 4:

```r
library(BSgenome.Scerevisiae.UCSC.sacCer2)
genome <- BSgenome.Scerevisiae.UCSC.sacCer2
gpos4 <- GPos(seqinfo(genome))
gpos4 # all the positions along the genome are represented
mcols(gpos4)$dna <- do.call("c", unname(as.list(genome)))
gpos4
```

## Note however that, like for any Vector derivative, the length of a GPos object cannot exceed ".Machine$integer.max" (i.e. 2^31 on most platforms) so the above only works with a "small" genome.

## For example it doesn't work with the Human genome:

```r
library(TxDb.Hsapiens.UCSC.hg19.knownGene)
## Not run:
GPos(seqinfo(TxDb.Hsapiens.UCSC.hg19.knownGene)) # error!
```

## End(Not run)

## You can use isSmallGenome() to check upfront whether the genome is "small" or not.

```r
isSmallGenome(genome)
isSmallGenome(TxDb.Hsapiens.UCSC.hg19.knownGene)
```

## MEMORY USAGE

```r
# Coercion to GRanges works...
gr4 <- as(gpos4, "GRanges")
gr4
```

## ... but is generally not a good idea:

```r
object.size(gpos4)
```
## Shuffling the order of the positions impacts memory usage:

gpos4r <- rev(gpos4)
optimize.size(gpos4r)  # significantly

gpos4s <- sample(gpos4)
optimize.size(gpos4s)  # even worse!

## AN IMPORTANT NOTE: In the worst situations, GPos still performs as
## good as a GRanges object.

object.size(as(gpos4r, "GRanges"))  # same size as 'gpos4r'
optimize.size(as(gpos4s, "GRanges"))  # same size as 'gpos4s'

## Best case scenario is when the object is strictly sorted (i.e.
## positions are in strict ascending order).
## This can be checked with:

is.unsorted(gpos4, strict=TRUE)  # 'gpos4' is strictly sorted

## USING MEMORY-EFFICIENT METADATA COLUMNS

## In order to keep memory usage as low as possible, it is recommended
## to use a memory-efficient representation of the metadata columns that
## we want to set on the object. Rle's are particularly well suited for
## this, especially if the metadata columns contain long runs of
## identical values. This is the case for example if we want to use a
## GPos object to represent the coverage of sequencing reads along a
## genome.

## Example 5:

library(pasillaBamSubset)
library(Rsamtools)  # for the BamFile() constructor function
bamfile1 <- BamFile(untreated1_chr4())
bamfile2 <- BamFile(untreated3_chr4())
gpos5 <- GPos(seqinfo(bamfile1))

library(GenomicAlignments)  # for "coverage" method for BamFile objects
cov1 <- unlist(coverage(bamfile1), use.names=FALSE)
cov2 <- unlist(coverage(bamfile2), use.names=FALSE)
mcols(gpos5) <- DataFrame(cov1, cov2)
gpos5

object.size(gpos5)  # lightweight

## Keep only the positions where coverage is at least 10 in one of the
## 2 samples:
gpos5[mcols(gpos5)$cov1 >= 10 | mcols(gpos5)$cov2 >= 10]

## USING A GPos OBJECT IN A SummarizedExperiment OBJECT

## Because the GPos class extends the GenomicRanges virtual class, a GPos
## object can be used as the rowRanges component of a SummarizedExperiment
## object.

## As a 1st example, we show how the counts for samples sA and sB in
## 'gpos3' can be stored in a SummarizedExperiment object where the rows
## correspond to unique genomic positions and the columns to samples:
library(SummarizedExperiment)
counts <- cbind(sA=sA_counts, sB=sB_counts)
mcols(gpos3A) <- NULL
rse3 <- SummarizedExperiment(list(counts=counts), rowRanges=gpos3A)
rse3
rowRanges(rse3)
head(assay(rse3))

## Finally we show how the coverage data from Example 5 can be easily
## stored in a lightweight SummarizedExperiment object:
cov <- mcols(gpos5)
mcols(gpos5) <- NULL
rse5 <- SummarizedExperiment(list(cov=cov), rowRanges=gpos5)
rse5
rowRanges(rse5)
assay(rse5)

## Keep only the positions where coverage is at least 10 in one of the
## 2 samples:
rse5[assay(rse5)$cov1 >= 10 | assay(rse5)$cov2 >= 10]

---

**GRanges-class**

**GRanges objects**

---

**Description**

The GRanges class is a container for the genomic locations and their associated annotations.

**Details**

GRanges is a vector of genomic locations and associated annotations. Each element in the vector is comprised of a sequence name, an interval, a strand, and optional metadata columns (e.g. score, GC content, etc.). This information is stored in four components:

- **seqnames** a 'factor' Rle object containing the sequence names.
- **ranges** an IRanges object containing the ranges.
- **strand** a 'factor' Rle object containing the strand information.
- **mcols** a DataFrame object containing the metadata columns. Columns cannot be named "seqnames", "ranges", "strand", "seqlevels", "seqlengths", "isCircular", "start", "end", "width", or "element".
- **seqinfo** a Seqinfo object containing information about the set of genomic sequences present in the GRanges object.

**Constructor**

```
GRanges(seqnames=NULL, ranges=NULL, strand=NULL, ...)
```

 Creates a GRanges object.

**seqnames** NULL, or an Rle object, character vector, or factor containing the sequence names.

**ranges** NULL, or an IRanges object containing the ranges.

**strand** NULL, or an Rle object, character vector, or factor containing the strand information.
Optional metadata columns. These columns cannot be named "start", "end", "width", or "element".

seqlengths NULL, or an integer vector named with levels(seqnames) and containing the lengths (or NA) for each level in levels(seqnames).

seqinfo NULL, or a Seqinfo object containing allowed sequence names, lengths (or NA), and circularity flag, for each level in levels(seqnames).

If ranges is not supplied and/or NULL then the constructor proceeds in 2 steps:
1. An initial GRanges object is created with as(seqnames, "GRanges").
2. Then this GRanges object is updated according to whatever non-NULL remaining arguments were passed to the call to GRanges().

As a consequence of this behavior, GRanges(x) is equivalent to as(x, "GRanges").

Coercion

In the code snippets below, x is a GRanges object.

as(from, "GRanges"): Creates a GRanges object from a character vector, a factor, or a RangedData, or RangesList object.
When from is a character vector (or a factor), each element in it must represent a genomic range in format chr1:2501-2800 (unstranded range) or chr1:2501-2800:+ (stranded range). .. is also supported as a separator between the start and end positions. Strand can be +, -, *, or missing. The names on from are propagated to the returned GRanges object. See as.character() and as.factor() below for the reverse transformations.

Coercing a data.frame or DataFrame into a GRanges object is also supported. See makeGRangesFromDataFrame for the details.

as(from, "RangedData"): Creates a RangedData object from a GRanges object. The strand and metadata columns become columns in the result. The seqlengths(from), isCircular(from), and genome(from) vectors are stored in the metadata columns of ranges(rd).

as(from, "RangesList"): Creates a RangesList object from a GRanges object. The strand and metadata columns become inner metadata columns (i.e. metadata columns on the ranges). The seqlengths(from), isCircular(from), and genome(from) vectors become the metadata columns.

as.character(x, ignore.strand=FALSE): Turn GRanges object x into a character vector where each range in x is represented by a string in format chr1:2501-2800:+. If ignore.strand is TRUE or if all the ranges in x are unstranded (i.e. their strand is set to *), then all the strings in the output are in format chr1:2501-2800.
The names on x are propagated to the returned character vector. Its metadata (metadata(x)) and metadata columns (mcols(x)) are ignored.
See as(from, "GRanges") above for the reverse transformation.

as.factor(x): Equivalent to

factor(as.character(x), levels=as.character(sort(unique(x))))

See as(from, "GRanges") above for the reverse transformation.
Note that table(x) is supported on a GRanges object. It is equivalent to, but much faster than, table(as.factor(x)).

as.data.frame(x, row.names = NULL, optional = FALSE, ...): Creates a data.frame with columns seqnames (factor), start (integer), end (integer), width (integer), strand (factor), as well as the additional metadata columns stored in mcols(x). Pass an explicit stringsAsFactors=TRUE/FALSE argument via ... to override the default conversions for the metadata columns in mcols(x).
as(from, "Grouping"): Creates a ManyToOneGrouping object that groups from by seqname, strand, start and end (same as the default sort order). This makes it convenient, for example, to aggregate a GenomicRanges object by range.

In the code snippets below, x is a Seqinfo object.

as(x, "GRanges"), as(x, "GenomicRanges"), as(x, "RangesList"): Turns Seqinfo object x (with no NA lengths) into a GRanges or RangesList.

Accessors

In the following code snippets, x is a GRanges object.

length(x): Get the number of elements.

seqnames(x), seqnames(x) <- value: Get or set the sequence names. value can be an Rle object, a character vector, or a factor.

ranges(x), ranges(x) <- value: Get or set the ranges. value can be a Ranges object.

names(x), names(x) <- value: Get or set the names of the elements.

strand(x), strand(x) <- value: Get or set the strand. value can be an Rle object, character vector, or factor.

mcols(x, use.names=FALSE), mcols(x) <- value: Get or set the metadata columns. If use.names=TRUE and the metadata columns are not NULL, then the names of x are propagated as the row names of the returned DataFrame object. When setting the metadata columns, the supplied value must be NULL or a data.frame-like object (i.e. DataTable or data.frame) object holding element-wise metadata.

elementMetadata(x), elementMetadata(x) <- value, values(x), values(x) <- value: Alternatives to mcols functions. Their use is discouraged.

seqinfo(x), seqinfo(x) <- value: Get or set the information about the underlying sequences. value must be a Seqinfo object.

seqlevels(x), seqlevels(x, pruning.mode=c("error", "coarse", "fine", "tidy")) <- value: Get or set the sequence levels. seqlevels(x) is equivalent to seqlevels(seqinfo(x)) or to levels(seqnames(x)), those 2 expressions being guaranteed to return identical character vectors on a GRanges object. value must be a character vector with no NAs. See ?seqlevels for more information.

seqlengths(x), seqlengths(x) <- value: Get or set the sequence lengths. seqlengths(x) is equivalent to seqlengths(seqinfo(x)). value can be a named non-negative integer or numeric vector eventually with NAs.

isCircular(x), isCircular(x) <- value: Get or set the circularity flags. isCircular(x) is equivalent to isCircular(seqinfo(x)). value must be a named logical vector eventually with NAs.

gene(x), genome(x) <- value: Get or set the genome identifier or assembly name for each sequence. genome(x) is equivalent to genome(seqinfo(x)). value must be a named character vector eventually with NAs.

seqlevelsStyle(x), seqlevelsStyle(x) <- value: Get or set the seqname style for x. See the seqlevelsStyle generic getter and setter in the GenomeInfoDb package for more information.

score(x), score(x) <- value: Get or set the “score” column from the element metadata.

granges(x, use.mcols=FALSE): Gets a GRanges with only the range information from x, unless use.mcols is TRUE, in which case the metadata columns are also returned. Those columns will include any "extra column slots" if x is a specialized GenomicRanges derivative.
**Ranges methods**

In the following code snippets, \( x \) is a GRanges object.

\[
\begin{align*}
\text{start}(x), & \quad \text{start}(x) \leftarrow \text{value}: \text{Get or set start(ranges}(x)). \\
\text{end}(x), & \quad \text{end}(x) \leftarrow \text{value}: \text{Get or set end(ranges}(x)). \\
\text{width}(x), & \quad \text{width}(x) \leftarrow \text{value}: \text{Get or set width(ranges}(x)).
\end{align*}
\]

**Splitting and Combining**

In the code snippets below, \( x \) is a GRanges object.

\[
\begin{align*}
\text{append}(x, \text{values}, \text{after} = \text{length}(x)): \text{Inserts the values into } x \text{ at the position given by after, where } x \text{ and values are of the same class.} \\
\text{c}(x, \ldots): \text{Combines } x \text{ and the GRanges objects in } \ldots \text{ together. Any object in } \ldots \text{ must belong to the same class as } x, \text{ or to one of its subclasses, or must be NULL. The result is an object of the same class as } x. \\
\text{c}(x, \ldots, \text{ignore.mcols}=\text{FALSE}) \quad \text{If the GRanges objects have metadata columns (represented as one DataFrame per object), each such DataFrame must have the same columns in order to combine successfully. In order to circumvent this restraint, you can pass in an } \text{ignore.mcols}=\text{TRUE} \text{ argument which will combine all the objects into one and drop all of their metadata columns.} \\
\text{split}(x, f, \text{drop}=\text{FALSE}): \text{Splits } x \text{ according to } f \text{ to create a GRangesList object. If } f \text{ is a list-like object then drop is ignored and } f \text{ is treated as if it was } \text{rep(seq_len(length}(f)), \text{sapply}(f, \text{length})), \text{ so the returned object has the same shape as } f \text{ (it also receives the names of } f). \text{ Otherwise, if } f \text{ is not a list-like object, empty list elements are removed from the returned object if drop is TRUE.}
\end{align*}
\]

**Subsetting**

In the code snippets below, \( x \) is a GRanges object.

\[
\begin{align*}
x[i, j], & \quad x[i, j] \leftarrow \text{value}: \text{Get or set elements } i \text{ with optional metadata columns } \text{mcols}(x)[,j], \text{ where } i \text{ can be missing; an NA-free logical, numeric, or character vector; or a 'logical' Rle object.} \\
x[i, j] \leftarrow \text{value}: \text{Replaces elements } i \text{ and optional metadata columns } j \text{ with value.} \\
\text{head}(x, n = 6L): \text{If } n \text{ is non-negative, returns the first } n \text{ elements of the GRanges object. If } n \text{ is negative, returns all but the last } \text{abs}(n) \text{ elements of the GRanges object.} \\
\text{rep}(x, \text{times}, \text{length.out}, \text{each}): \text{Repeats the values in } x \text{ through one of the following conventions:} \\
\text{times} & \quad \text{Vector giving the number of times to repeat each element if of length length}(x), \text{ or to repeat the whole vector if of length l.} \\
\text{length.out} & \quad \text{Non-negative integer. The desired length of the output vector.} \\
\text{each} & \quad \text{Non-negative integer. Each element of } x \text{ is repeated each times.} \\
\text{subset}(x, \text{subset}): \text{Returns a new object of the same class as } x \text{ made of the subset using logical vector subset, where missing values are taken as } \text{FALSE.} \\
\text{tail}(x, n = 6L): \text{If } n \text{ is non-negative, returns the last } n \text{ elements of the GRanges object. If } n \text{ is negative, returns all but the first } \text{abs}(n) \text{ elements of the GRanges object.} \\
\text{window}(x, \text{start} = \text{NA}, \text{end} = \text{NA}, \text{width} = \text{NA}, \text{frequency} = \text{NULL}, \text{delta} = \text{NULL}, \ldots): \text{Extracts the subsequence window from the GRanges object using:} \\
\text{start, end, width} & \quad \text{The start, end, or width of the window. Two of the three are required.}
\end{align*}
\]
frequency, delta  Optional arguments that specify the sampling frequency and increment within the window.

In general, this is more efficient than using "[" operator.

window(x, start = NA, end = NA, width = NA, keepLength = TRUE) <- value: Replaces the subsequence window specified on the left (i.e. the subsequence in x specified by start, end and width) by value. value must either be of class class(x), belong to a subclass of class(x), be coercible to class(x), or be NULL. If keepLength is TRUE, the elements of value are repeated to create a GRanges object with the same number of elements as the width of the subsequence window it is replacing. If keepLength is FALSE, this replacement method can modify the length of x, depending on how the length of the left subsequence window compares to the length of value.

x$name, x$name <- value: Shortcuts for mcols(x)$name and mcols(x)$name <- value, respectively. Provided as a convenience, for GRanges objects *only*, and as the result of strong popular demand. Note that those methods are not consistent with the other $ and <$- methods in the IRanges/GenomicRanges infrastructure, and might confuse some users by making them believe that a GRanges object can be manipulated as a data.frame-like object. Therefore we recommend using them only interactively, and we discourage their use in scripts or packages. For the latter, use mcols(x)$name and mcols(x)$name <- value, instead of x$name and x$name <- value, respectively.

Note that a GRanges object can be used to as a subscript to subset a list-like object that has names on it. In that case, the names on the list-like object are interpreted as sequence names. In the code snippets below, x is a list or List object with names on it, and the subscript gr is a GRanges object with all its seqnames being valid x names.

x[gr]: Return an object of the same class as x and parallel to gr. More precisely, it’s conceptually doing:

\[ \text{lapply(gr, function(gr1) x[[seqnames(gr1)]][ranges(gr1)])} \]

Other methods

show(x): By default the show method displays 5 head and 5 tail elements. This can be changed by setting the global options showHeadLines and showTailLines. If the object length is less than (or equal to) the sum of these 2 options plus 1, then the full object is displayed. Note that these options also affect the display of GAlignments and GAlignmentPairs objects (defined in the GenomicAlignments package), as well as other objects defined in the IRanges and Biostrings packages (e.g. IRanges and DNAStringSet objects).

Author(s)

P. Aboyoun and H. Pagès

See Also

- makeGRangesFromDataFrame for making a GRanges object from a data.frame or DataFrame object.
- seqinfo for accessing/modifying information about the underlying sequences of a GRanges object.
- The GPos class, a memory-efficient container for storing genomic positions, that is, genomic ranges of width 1.
- GenomicRanges-comparison for comparing and ordering genomic ranges.
• findOverlaps-methods for finding/counting overlapping genomic ranges.
• intra-range-methods and inter-range-methods for intra range and inter range transformations of a GRanges object.
• coverage-methods for computing the coverage of a GRanges object.
• setops-methods for set operations on GRanges objects.
• nearest-methods for finding the nearest genomic range neighbor.
• absoluteRanges for transforming genomic ranges into absolute ranges (i.e. into ranges on the sequence obtained by virtually concatenating all the sequences in a genome).
• tileGenome for putting tiles on a genome.
• genomicvars for manipulating genomic variables.
• GRangesList objects.
• Ranges objects in the IRanges package.
• Vector, Rle, and DataFrame objects in the S4Vectors package.

Examples

```r
## CONSTRUCTION

## Specifying the bare minimum i.e. seqnames and ranges only. The GRanges object will have no names, no strand information, and no metadata columns:
gr0 <- GRanges(Rle(c("chr2", "chr2", "chr1", "chr3"), c(1, 3, 2, 4)),
               IRanges(1:10, width=10:1))
gr0

## Specifying names, strand, metadata columns. They can be set on an existing object:
names(gr0) <- head(letters, 10)
strand(gr0) <- Rle(strand(c("-", "+", "x", "+", "-")), c(1, 2, 2, 3, 2))
mcols(gr0)$score <- 1:10
mcols(gr0)$GC <- seq(1, 0, length=10)
gr0

## ... or specified at construction time:
gr <- GRanges(Rle(c("chr2", "chr2", "chr1", "chr3"), c(1, 3, 2, 4)),
               IRanges(1:10, width=10:1, names=head(letters, 10)),
               Rle(strand(c("-", "+", "x", "+", "-")), c(1, 2, 2, 3, 2)),
               score=1:10, GC=seq(1, 0, length=10))
stopifnot(identical(gr0, gr))

## Specifying the seqinfo. It can be set on an existing object:
seqinfo <- Seqinfo(paste0("chr", 1:3), c(1000, 2000, 1500), NA, "mock1")
seqinfo(gr0) <- merge(seqinfo(gr0), seqinfo)
seqlevels(gr0) <- seqlevels(seqinfo)

## ... or specified at construction time:
gr <- GRanges(Rle(c("chr2", "chr2", "chr1", "chr3"), c(1, 3, 2, 4)),
               IRanges(1:10, width=10:1, names=head(letters, 10)),
               Rle(strand(c("-", "+", "x", "+", "-")), c(1, 2, 2, 3, 2)),
               score=1:10, GC=seq(1, 0, length=10),
               seqinfo=seqinfo)
stopifnot(identical(gr0, gr))
```
## COERCION

## From GRanges:
- `as.character(gr)`
- `as.factor(gr)`
- `as.data.frame(gr)`

## From character to GRanges:
- `x1 <- "chr2:56-125"`
- `as(x1, "GRanges")`
- `as(rep(x1, 4), "GRanges")`
- `x2 <- c(A=x1, B="chr1:25-30:-")`
- `as(x2, "GRanges")`

## From data.frame to GRanges:
- `df <- data.frame(chrom="chr2", start=11:15, end=20:24)`
- `gr3 <- as(df, "GRanges")`

## Alternatively, coercion to GRanges can be done by just calling the
## `GRanges()` constructor on the object to coerce:
- `gr1 <- GRanges(x1)`  # same as `as(x1, "GRanges")`
- `gr2 <- GRanges(x2)`  # same as `as(x2, "GRanges")`
- `gr3 <- GRanges(df)`  # same as `as(df, "GRanges")`

## Sanity checks:
- `stopifnot(identical(as(x1, "GRanges"), gr1))`
- `stopifnot(identical(as(x2, "GRanges"), gr2))`
- `stopifnot(identical(as(df, "GRanges"), gr3))`

## SUMMARIZING ELEMENTS

## Table:
- `table(seqnames(gr))`
- `table(strand(gr))`
- `sum(width(gr))`
- `table(gr)`
- `summary(mcols(gr)[,"score"])`

## The number of lines displayed in the 'show' method are controlled
## with two global options:
- `longGR <- sample(gr, 25, replace=TRUE)`
- `longGR <- options(showHeadLines=7)`
- `options(showTailLines=2)`

## Revert to default values
- `options(showHeadLines=NULL)`
- `options(showTailLines=NULL)`

## INVERTING THE STRAND

## `invertStrand(gr)`
## RENAMING THE UNDERLYING SEQUENCES
```
seqlevels(gr) <- sub("chr", "Chrom", seqlevels(gr))
gr
seqlevels(gr) <- sub("Chrom", "chr", seqlevels(gr)) # revert
```

## COMBINING OBJECTS
```
gr2 <- GRanges(seqnames=Rle(c("chr1", "chr2", "chr3"), c(3, 3, 4)),
IRanges(1:10, width=5),
strand="-",
score=101:110, GC=runif(10),
seqinfo=seqinfo)
gr3 <- GRanges(seqnames=Rle(c("chr1", "chr2", "chr3"), c(3, 4, 3)),
IRanges(101:110, width=10),
strand="-",
score=21:30,
seqinfo=seqinfo)
some.gr <- c(gr, gr2)
```

## USING A GRANGES OBJECT AS A SUBSCRIPT TO SUBSET ANOTHER OBJECT
```
x <- RleList(chr1=101:120, chr2=2:-8, chr3=31:40)
x[gr]
```
length(x): Get the number of list elements.
names(x), names(x) <- value: Get or set the names on x.

elementNROWS(x): Get a vector of the length of each of the list element.
isEmpty(x): Returns a logical indicating either if the GRangesList has no elements or if all its elements are empty.

seqnames(x), seqnames(x) <- value: Get or set the sequence names in the form of an RleList. value can be an RleList or CharacterList object.
ranges(x, use.mcols=FALSE), ranges(x) <- value: Get or set the ranges in the form of a CompressedIRangesList. value can be a RangesList object.

start(x), start(x) <- value: Get or set start(ranges(x)).
end(x), end(x) <- value: Get or set end(ranges(x)).
width(x), width(x) <- value: Get or set width(ranges(x)).

strand(x), strand(x) <- value: Get or set the strand in the form of an RleList. value can be an RleList, CharacterList or single character. value as a single character converts all ranges in x to the same value; for selective strand conversion (i.e., mixed “+” and “-”) use RleList or CharacterList.
mcols(x, use.names=FALSE), mcols(x) <- value: Get or set the metadata columns. value can be NULL, or a data.frame-like object (i.e. DataFrame or data.frame) holding element-wise metadata.
elementMetadata(x), elementMetadata(x) <- value, values(x), values(x) <- value: Alternatives to mcols functions. Their use is discouraged.

seqinfo(x), seqinfo(x) <- value: Get or set the information about the underlying sequences. value must be a Seqinfo object.

seqlevels(x), seqlevels(x) <- value, seqlevels(x, pruning.mode=c("error", "coarse", "fine", "tidy")) <- value: Get or set the sequence levels. seqlevels(x) is equivalent to seqlevels(seqinfo(x)) or to levels(seqnames(x)), those 2 expressions being guaranteed to return identical character vectors on a GRangesList object. value must be a character vector with no NAs. See ?seqlevels for more information.

seqlengths(x), seqlengths(x) <- value: Get or set the sequence lengths. seqlengths(x) is equivalent to seqlengths(seqinfo(x)). value can be a named non-negative integer or numeric vector eventually with NAs.
isCircular(x), isCircular(x) <- value: Get or set the circularity flags. isCircular(x) is equivalent to isCircular(seqinfo(x)). value must be a named logical vector eventually with NAs.

genome(x), genome(x) <- value: Get or set the genome identifier or assembly name for each sequence. genome(x) is equivalent to genome(seqinfo(x)). value must be a named character vector eventually with NAs.

seqlevelsStyle(x), seqlevelsStyle(x) <- value: Get or set the seqname style for x. See the seqlevelsStyle generic getter and setter in the GenomeInfoDb package for more information.
score(x), score(x) <- value: Get or set the “score” metadata column.

**Coercion**

In the code snippets below, x is a GRangesList object.

as.data.frame(x, row.names = NULL, optional = FALSE, ..., value.name = "value", use.outer.mcols = FALSE)
Coerces x to a data.frame. See as.data.frame on the List man page for details (?List).
as.list(x, use.names = TRUE): Creates a list containing the elements of x.
as(x, "IRangesList"): Turns x into an IRangesList object.

When x is a list of GRanges, it can be coerced to a GRangesList.
as(x, "GRangesList"): Turns x into a GRangesList.

Subsetting

In the following code snippets, x is a GRangesList object.

x[i, j], x[i, j] <- value: Get or set elements i with optional metadata columns mcols(x)[i, j], where i can be missing; an NA-free logical, numeric, or character vector; a 'logical' Rle object, or an AtomicList object.
x[[i]], x[[i]] <- value: Get or set element i, where i is a numeric or character vector of length 1.
x$name, x$name <- value: Get or set element name, where name is a name or character vector of length 1.

head(x, n = 6L): If n is non-negative, returns the first n elements of the GRangesList object. If n is negative, returns all but the last abs(n) elements of the GRangesList object.

rep(x, times, length.out, each): Repeats the values in x through one of the following conventions:
taxes Vector giving the number of times to repeat each element if of length length(x), or to repeat the whole vector if of length 1.
times Vector giving the number of times to repeat each element if of length length(x), or to repeat the whole vector if of length 1.
length.out Non-negative integer. The desired length of the output vector.
each Non-negative integer. Each element of x is repeated each times.

subset(x, subset): Returns a new object of the same class as x made of the subset using logical vector subset, where missing values are taken as FALSE.

tail(x, n = 6L): If n is non-negative, returns the last n elements of the GRanges object. If n is negative, returns all but the first abs(n) elements of the GRanges object.

Combining

In the code snippets below, x is a GRangesList object.

c(x, ...): Combines x and the GRangesList objects in ... together. Any object in ... must belong to the same class as x, or to one of its subclasses, or must be NULL. The result is an object of the same class as x.

append(x, values, after = length(x)): Inserts the values into x at the position given by after, where x and values are of the same class.

unlist(x, recursive = TRUE, use.names = TRUE): Concatenates the elements of x into a single GRanges object.

Looping

In the code snippets below, x is a GRangesList object.

endoapply(X, FUN, ...): Similar to lapply, but performs an endomorphism, i.e. returns an object of class(X).
lapply(X, FUN, ...): Like the standard `lapply` function defined in the base package, the `lapply` method for GRangesList objects returns a list of the same length as X, with each element being the result of applying FUN to the corresponding element of X.

Map(f, ...): Applies a function to the corresponding elements of given GRangesList objects.

mapply(FUN, ..., MoreArgs = NULL, SIMPLIFY = TRUE, USE.NAMES = TRUE): Like the standard `mapply` function defined in the base package, the `mapply` method for GRangesList objects is a multivariate version of `sapply`.

mendoapply(FUN, ..., MoreArgs = NULL): Similar to `mapply`, but performs an endomorphism across multiple objects, i.e. returns an object of class(list(...)[[1]])

Reduce(f, x, init, right = FALSE, accumulate = FALSE): Uses a binary function to successively combine the elements of x and a possibly given initial value.

f A binary argument function.
init An R object of the same kind as the elements of x.
right A logical indicating whether to proceed from left to right (default) or from right to left.
nomatch The value to be returned in the case when "no match" (no element satisfying the predicate) is found.

sapply(X, FUN, ..., simplify=TRUE, USE.NAMES=TRUE): Like the standard `sapply` function defined in the base package, the `sapply` method for GRangesList objects is a user-friendly version of `lapply` by default returning a vector or matrix if appropriate.

Author(s)
P. Aboyoun & H. Pagès

See Also

GRanges-class, seqinfo, Vector-class, RangesList-class, RleList-class, DataFrameList-class, intra-range-methods, inter-range-methods, coverage-methods, setops-methods, findOverlaps-methods

Examples

```r
## Construction with GRangesList():
gr1 <-
  GRanges(seqnames = "chr2", ranges = IRanges(3, 6),
   strand = "+", score = 5L, GC = 0.45)
gr2 <-
  GRanges(seqnames = c("chr1", "chr2"),
   ranges = IRanges(c(7,13), width = 3),
   strand = c("+", "-"), score = 3:4, GC = c(0.3, 0.5))
gr3 <-
  GRanges(seqnames = c("chr2"),
   ranges = IRanges(c(1, 4), c(3, 9)),
   strand = c("-", "-"), score = c(6L, 2L), GC = c(0.4, 0.1))
gr1 <- GRangesList("gr1" = gr1, "gr2" = gr2, "gr3" = gr3)
grl

## Summarizing elements:
elementNROWS(grl)
table(seqnames(grl))

## Extracting subsets:
grl[seqnames(grl) == "chr1", ]
grl[seqnames(grl) == "chr1" & strand(grl) == "+", ]
```

## Renaming the underlying sequences:
seqlevels(grl)
seqlevels(grl) <- sub("chr", "Chrom", seqlevels(grl))
grl

## Coerce to IRangesList (seqnames and strand information is lost):
as(grl, "IRangesList")

## isDisjoint():
isDisjoint(grl)

## disjoint():
disjoint(grl) # metadata columns and order NOT preserved

## Construction with makeGRangesListFromFeatureFragments():
filepath <- system.file("extdata", "feature_frags.txt", 
package="GenomicRanges")
featfrags <- read.table(filepath, header=TRUE, stringsAsFactors=FALSE)
grl2 <- with(featfrags,
makeGRangesListFromFeatureFragments(seqnames=targetName,
fragmentStarts=targetStart,
fragmentWidths=blockSizes,
strand=strand))

names(grl2) <- featfrags$RefSeqID
grl2

---

**inter-range-methods**

*Inter range transformations of a GRanges or GRangesList object*

**Description**

This man page documents *inter range transformations* of a *GenomicRanges* object (i.e. of an object that belongs to the *GenomicRanges* class or one of its subclasses, this includes for example *GRanges* objects), or a *GRangesList* object.

See ?`intra-range-methods` and ?`inter-range-methods` in the *IRanges* package for a quick introduction to *intra range* and *inter range transformations*.

See ?`intra-range-methods` for *intra range transformations* of a *GenomicRanges* object or *GRangesList* object.

**Usage**

```
## S4 method for signature 'GenomicRanges'
range(x, ..., with.revmap=FALSE, ignore.strand=FALSE, na.rm=FALSE)
## S4 method for signature 'GRangesList'
range(x, ..., with.revmap=FALSE, ignore.strand=FALSE, na.rm=FALSE)
## S4 method for signature 'GenomicRangesList'
range(x, ..., with.revmap=FALSE, ignore.strand=FALSE, na.rm=FALSE)
## S4 method for signature 'GenomicRanges'
reduce(x, drop.empty.ranges=FALSE, min.gapwidth=1L, with.revmap=FALSE, 
       with.inframe.attrib=FALSE, ignore.strand=FALSE)
```
## S4 method for signature 'GRangesList'
reduce(x, drop.empty.ranges=FALSE, min.gapwidth=1L, with.revmap=FALSE, with.inframe.attrib=FALSE, ignore.strand=FALSE)

## S4 method for signature 'GenomicRangesList'
reduce(x, drop.empty.ranges=FALSE, min.gapwidth=1L, with.inframe.attrib=FALSE, ignore.strand=FALSE)

## S4 method for signature 'GenomicRanges'
gaps(x, start=1L, end=seqlengths(x))

## S4 method for signature 'GenomicRangesList'
gaps(x, start=1L, end=seqlengths(x))

## S4 method for signature 'GenomicRanges'
disjoin(x, with.revmap=FALSE, ignore.strand=FALSE)

## S4 method for signature 'GRangesList'
disjoin(x, with.revmap=FALSE, ignore.strand=FALSE)

## S4 method for signature 'GenomicRangesList'
disjoin(x, with.revmap=FALSE, ignore.strand=FALSE)

## S4 method for signature 'GenomicRanges'
isDisjoint(x, ignore.strand=FALSE)

## S4 method for signature 'GRangesList'
isDisjoint(x, ignore.strand=FALSE)

## S4 method for signature 'GenomicRangesList'
isDisjoint(x, ignore.strand=FALSE)

## S4 method for signature 'GenomicRanges'
disjointBins(x, ignore.strand=FALSE)

Arguments

- **x**
  - A `GenomicRanges` or `GenomicRangesList` object.
- **drop.empty.ranges**, **min.gapwidth**, **with.revmap**, **with.inframe.attrib**, **start**, **end**
  - See `?inter-range-methods` in the `IRanges` package.
- **ignore.strand**
  - TRUE or FALSE. Whether the strand of the input ranges should be ignored or not. See details below.
- ... For `range`, additional `GenomicRanges` objects to consider. Ignored otherwise.
- **na.rm**
  - Ignored.

Details

**On a GRanges object:** `range` returns an object of the same type as `x` containing range bounds for each distinct (seqname, strand) pairing. The names (names(x)) and the metadata columns in `x` are dropped.

`reduce` returns an object of the same type as `x` containing reduced ranges for each distinct (seqname, strand) pairing. The names (names(x)) and the metadata columns in `x` are dropped. See `?reduce` for more information about range reduction and for a description of the optional arguments.

`gaps` returns an object of the same type as `x` containing complemented ranges for each distinct (seqname, strand) pairing. The names (names(x)) and the columns in `x` are dropped. For the start and end arguments of this gaps method, it is expected that the user will supply a named integer vector (where the names correspond to the appropriate seqlevels). See `?gaps` for more information about range complements and for a description of the optional arguments.
disjoin returns an object of the same type as \( x \) containing disjoint ranges for each distinct (seqname, strand) pairing. The names (\( \text{names}(x) \)) and the metadata columns in \( x \) are dropped. If \( \text{with.revmap=TRUE} \), a metadata column that maps the output ranges to the input ranges is added to the returned object. See \(?\text{disjoin}\) for more information.

\( \text{isDisjoint} \) returns a logical value indicating whether the ranges in \( x \) are disjoint (i.e. non-overlapping).

\( \text{disjointBins} \) returns bin indexes for the ranges in \( x \), such that ranges in the same bin do not overlap. If \( \text{ignore.strand=FALSE} \), the two features cannot overlap if they are on different strands.

**On a GRangesList/GenomicRangesList object:** When they are supported on GRangesList object \( x \), the above inter range transformations will apply the transformation to each of the list elements in \( x \) and return a list-like object parallel to \( x \) (i.e. with 1 list element per list element in \( x \)). If \( x \) has names on it, they’re propagated to the returned object.

**Author(s)**

H. Pagès and P. Aboyoun

**See Also**

- The GenomicRanges and GRanges classes.
- The Ranges class in the IRanges package.
- The inter-range-methods man page in the IRanges package.
- GenomicRanges-comparison for comparing and ordering genomic ranges.
- endoapply in the S4Vectors package.

**Examples**

```r
gr <- GRanges(
    seqnames=Rle(paste("chr", c(1, 2, 1, 3), sep=""), c(1, 3, 2, 4)),
    ranges=IRanges(1:10, width=10:1, names=letters[1:10]),
    strand=Rle(strand(c("-", "+", "+", "+", "+")), c(1, 2, 2, 3, 2)),
    score=1:10,
    GC=seq(1, 0, length=10)
)
```

```r
g1 <- GRanges(seqnames="chr2", ranges=IRanges(3, 6),
    strand="+", score=5L, GC=0.45)
g2 <- GRanges(seqnames="chr1",
    ranges=IRanges(c(10, 7, 19), width=5),
    strand=c("-", "-", "+"), score=3:5, GC=c(0.3, 0.5, 0.66))
g3 <- GRanges(seqnames=c("chr1", "chr2"),
    ranges=IRanges(c(1, 4), c(3, 9)),
    strand=c("-", "-"), score=c(6L, 2L), GC=c(0.4, 0.1))
g1 <- GRangesList(g1=g1, g2=g2, g3=g3)
g1

## --------------------------------------------------------------------------------
## range()
## --------------------------------------------------------------------------------
```

```r
## On a GRanges object:
range(gr)
```
range(gr, with.revmap=TRUE)

## On a GRangesList object:
range(grl)
range(grl, ignore.strand=TRUE)
range(grl, with.revmap=TRUE, ignore.strand=TRUE)

# ---------------------------------------------------------------------
# reduce()
# ---------------------------------------------------------------------
reduce(gr)

gr2 <- reduce(gr, with.revmap=TRUE)
revmap <- mcols(gr2)$revmap  # an IntegerList

## Use the mapping from reduced to original ranges to group the original
## ranges by reduced range:
relist(gr[unlist(revmap)], revmap)

## Or use it to split the DataFrame of original metadata columns by
## reduced range:
relist(mcols(gr)[unlist(revmap), ], revmap)  # a SplitDataFrameList

## [For advanced users] Use this reverse mapping to compare the reduced
## ranges with the ranges they originate from:
expanded_gr2 <- rep(gr2, elementNROWS(revmap))
reordered_gr <- gr[unlist(revmap)]
codes <- pcompare(expanded_gr2, reordered_gr)
## All the codes should translate to "d", "e", "g", or "h" (the 4 letters
## indicating that the range on the left contains the range on the right):
alphacodes <- rangeComparisonCodeToLetter(pcompare(expanded_gr2, reordered_gr))
stopifnot(all(alphacodes %in% c("d", "e", "g", "h")))

## On a big GRanges object with a lot of seqlevels:
mcols(gr) <- NULL
biggr <- c(gr, GRanges("chr1", IRanges(c(4, 1), c(5, 2)), strand="+"))
seqlevels(biggr) <- paste0("chr", 1:2000)
biggr <- rep(biggr, 25000)
set.seed(33)
seqnames(biggr) <- sample(factor(seqlevels(biggr), levels=seqlevels(biggr)), length(biggr), replace=TRUE)

biggr2 <- reduce(biggr, with.revmap=TRUE)
revmap <- mcols(biggr2)$revmap
expanded_biggr2 <- rep(biggr2, elementNROWS(revmap))
reordered_biggr <- biggr[unlist(revmap)]
codes <- pcompare(expanded_biggr2, reordered_biggr)
alphacodes <- rangeComparisonCodeToLetter(pcompare(expanded_biggr2, reordered_biggr))
stopifnot(all(alphacodes %in% c("d", "e", "g", "h")))
table(alphacodes)

## On a GRangesList object:
reduce(grl)  # Doesn't really reduce anything but note the reordering
# of the inner elements in the 2nd and 3rd list elements:
# the ranges are reordered by sequence name first (which
# should appear in the same order as in 'seqlevels(grl)'),

# and then by strand.
reduce(grl, ignore.strand=TRUE) # 2nd list element got reduced

## gaps()
## gaps(gr, start=1, end=10)

## disjoin(), isDisjoint(), disjointBins()

## On a GRangesList object:
disjoin(grl) # doesn't really disjoin anything but note the reordering
disjoin(grl, with.revmap=TRUE)

---

**intra-range-methods**  
*Intra range transformations of a GRanges or GRangesList object*

### Description

This man page documents *intra range transformations* of a GenomicRanges object (i.e. of an object that belongs to the GenomicRanges class or one of its subclasses, this includes for example GRanges objects), or a GRangesList object.

See ?`intra-range-methods` and ?`inter-range-methods` in the IRanges package for a quick introduction to intra range and inter range transformations.

Intra range methods for GAlignments and GAlignmentsList objects are defined and documented in the GenomicAlignments package.

See ?`inter-range-methods` for inter range transformations of a GenomicRanges or GRangesList object.

### Usage

#### S4 method for signature 'GenomicRanges'

```r
shift(x, shift=0L, use.names=TRUE)
```

#### S4 method for signature 'GRangesList'

```r
shift(x, shift=0L, use.names=TRUE)
```

#### S4 method for signature 'GenomicRangesList'

```r
shift(x, shift=0L, use.names=TRUE)
```

#### S4 method for signature 'GenomicRanges'

```r
narrow(x, start=NA, end=NA, width=NA, use.names=TRUE)
```

#### S4 method for signature 'GRangesList'

```r
narrow(x, start=NA, end=NA, width=NA, use.names=TRUE)
```

#### S4 method for signature 'GenomicRangesList'

```r
narrow(x, start=NA, end=NA, width=NA, use.names=TRUE)
```
Arguments

x  A GenomicRanges or GRangesList object.

shift, use.names, start, end, width, both, fix, keep.all.ranges, upstream, downstream  
See ?'intra-range-methods'.

ignore.strand  TRUE or FALSE. Whether the strand of the input ranges should be ignored or not. 
See details below.

...  Additional arguments to methods.
Details

- `narrow` on a `GenomicRanges` object behaves like on a `Ranges` object. See `?intra-range-methods` for the details.

A major difference though is that it returns a `GenomicRanges` object instead of a `Ranges` object. The returned object is parallel (i.e. same length and names) to the original object `x`.
- `resize` returns an object of the same type and length as `x` containing intervals that have been resized to width `width` based on the `strand(x)` values. Elements where `strand(x) == "+"` or `strand(x) == "*"` are anchored at `start(x)` and elements where `strand(x) == "-"` are anchored at the `end(x)`. The `use.names` argument determines whether or not to keep the names on the ranges.
- `flank` returns an object of the same type and length as `x` containing intervals of width `width` that flank the intervals in `x`. The start argument takes a logical indicating whether `x` should be flanked at the "start" (TRUE) or the "end" (FALSE), which for `strand(x) != "-"` is `start(x)` and `end(x)` respectively and for `strand(x) == "-"` is `end(x)` and `start(x)` respectively. The both argument takes a single logical value indicating whether the flanking region width positions extends into the range. If both=TRUE, the resulting range thus straddles the end point, with width positions on either side.
- `promoters` returns an object of the same type and length as `x` containing promoter ranges. Promoter ranges extend around the transcription start site (TSS) which is defined as `start(x)`. The upstream and downstream arguments define the number of nucleotides in the 5’ and 3’ direction, respectively. The full range is defined as,

\[(start(x) - \text{upstream}) \text{ to } (start(x) + \text{downstream} - 1)\]

Ranges on the * strand are treated the same as those on the + strand. When no seqlengths are present in `x`, it is possible to have non-positive start values in the promoter ranges. This occurs when `(TSS - \text{upstream}) < 1`. In the equal but opposite case, the end values of the ranges may extend beyond the chromosome end when `(TSS + \text{downstream} + 1) > \text{chromosome end}'. When seqlengths are not NA the promoter ranges are kept within the bounds of the defined seqlengths.
- `restrict` returns an object of the same type and length as `x` containing restricted ranges for distinct seqnames. The start and end arguments can be a named numeric vector of seqnames for the ranges to be restricted or a numeric vector or length 1 if the restriction operation is to be applied to all the sequences in `x`. See `?intra-range-methods` for more information about range restriction and for a description of the optional arguments.
- `trim` trims out-of-bound ranges located on non-circular sequences whose length is not NA.

Author(s)

P. Aboyoun and V. Obenchain <vobencha@fhcrc.org>

See Also

- `GenomicRanges`, `GRanges`, and `GRangesList` objects.
- The `intra-range-methods` man page in the `IRanges` package.
- The `Ranges` class in the `IRanges` package.
- `endoapply` in the `S4Vectors` package.
## makeGRangesFromDataFrame

Make a GRanges object from a data.frame or DataFrame

### Description

makeGRangesFromDataFrame takes a data-frame-like object as input and tries to automatically find the columns that describe genomic ranges. It returns them as a GRanges object.

makeGRangesFromDataFrame is also the workhorse behind the coercion method from data.frame (or DataFrame) to GRanges.

### Usage

```r
makeGRangesFromDataFrame(df,
              keep.extra.columns=FALSE,
              ignore.strand=FALSE,
```
makeGRangesFromDataFrame

seqinfo=NULL,
seqnames.field=c("seqnames", "seqname",
"chromosome", "chrom",
"chr", "chromosome_name",
"seqid"),
start.field="start",
end.field=c("end", "stop"),
strand.field="strand",
starts.in.df.are.0based=FALSE)

Arguments

df A data.frame or DataFrame object. If not, then the function first tries to turn df into a data frame with as.data.frame(df).

keep.extra.columns TRUE or FALSE (the default). If TRUE, the columns in df that are not used to form the genomic ranges of the returned GRanges object are then returned as metadata columns on the object. Otherwise, they are ignored. If df has a width column, then it's always ignored.

ignore.strand TRUE or FALSE (the default). If TRUE, then the strand of the returned GRanges object is set to "*".

seqinfo Either NULL, or a Seqinfo object, or a character vector of seqlevels, or a named numeric vector of sequence lengths. When not NULL, it must be compatible with the genomic ranges in df i.e. it must include at least the sequence levels represented in df.

seqnames.field A character vector of recognized names for the column in df that contains the chromosome name (a.k.a. sequence name) associated with each genomic range. Only the first name in seqnames.field that is found in colnames(df) is used. If no one is found, then an error is raised.

start.field A character vector of recognized names for the column in df that contains the start positions of the genomic ranges. Only the first name in start.field that is found in colnames(df) is used. If no one is found, then an error is raised.

end.field A character vector of recognized names for the column in df that contains the end positions of the genomic ranges. Only the first name in start.field that is found in colnames(df) is used. If no one is found, then an error is raised.

strand.field A character vector of recognized names for the column in df that contains the strand associated with each genomic range. Only the first name in strand.field that is found in colnames(df) is used. If no one is found or if ignore.strand is TRUE, then the strand of the returned GRanges object is set to "*".

starts.in.df.are.0based TRUE or FALSE (the default). If TRUE, then the start positions of the genomic ranges in df are considered to be 0-based and are converted to 1-based in the returned GRanges object. This feature is intended to make it more convenient to handle input that contains data obtained from resources using the "0-based start" convention. A notorious example of such resource is the UCSC Table Browser (http://genome.ucsc.edu/cgi-bin/hgTables).

Value

A GRanges object with one element per row in the input.
If the seqinfo argument was supplied, the returned object will have exactly the seqlevels specified in seqinfo and in the same order. Otherwise, the seqlevels are ordered according to the output of the rankSeqlevels function (except if df contains the seqnames in the form of a factor-Rle, in which case the levels of the factor-Rle become the seqlevels of the returned object and with no re-ordering).

If df has non-automatic row names (i.e. rownames(df) is not NULL and is not seq_len(nrow(df))), then they will be used to set names on the returned GRanges object.

Note

Coercing data.frame or DataFrame df into a GRanges object (with as(df, "GRanges")), or calling GRanges(df), are both equivalent to calling makeGRangesFromDataFrame(df, keep.extra.columns=TRUE).

Author(s)

H. Pagès, based on a proposal by Kasper Daniel Hansen

See Also

- GRanges objects.
- Seqinfo objects and the rankSeqlevels function in the GenomeInfoDb package.
- The makeGRangesListFromFeatureFragments function for making a GRangesList object from a list of fragmented features.
- The getTable function in the rtracklayer package for an R interface to the UCSC Table Browser.
- DataFrame objects in the S4Vectors package.

Examples

```r
## BASIC EXAMPLES
##
## df <- data.frame(chr="chr1", start=11:15, end=12:16, strand=c("+","-","+","-", "."), score=1:5)
df
makeGRangesFromDataFrame(df) # strand value "." is replaced with "*"

## The strand column is optional:
df <- data.frame(chr="chr1", start=11:15, end=12:16, score=1:5)
makeGRangesFromDataFrame(df)

## The seqinfo argument is optional:
gr <- makeGRangesFromDataFrame(df, keep.extra.columns=TRUE)
gr2 <- as(df, "GRanges") # equivalent to the above
stopifnot(identical(gr, gr2))
gr2 <- GRanges(df) # equivalent to the above
stopifnot(identical(gr, gr2))

## With ignore.strand=TRUE:
makeGRangesFromDataFrame(df, ignore.strand=TRUE)
makeGRangesFromDataFrame(df, keep.extra.columns=TRUE, ignore.strand=TRUE)

## With custom seqinfo:
makeGRangesFromDataFrame(df, seqinfo=paste0("chr", 4:1))
makeGRangesFromDataFrame(df, seqinfo=c(chrM=NA, chr1=500, chrX=100))
```
makeGRangesListFromDataFrame

makeGRangesListFromDataFrame(df, seqinfo=Seqinfo(paste0("chr", 4:1)))

## ABOUT AUTOMATIC DETECTION OF THE seqnames/start/end/strand COLUMNS
## ---------------------------------------------------------------------
## Automatic detection of the seqnames/start/end/strand columns is
## case insensitive:
## df <- data.frame(ChRoM="chr1", StarT=11:15, stoP=12:16, STRAND=c("\"+\", \"-\", \"\"+\", \"\"\", \"\"\"\")}, score=1:5)
makeGRangesFromDataFrame(df)

## It also ignores a common prefix between the start and end columns:
## df <- data.frame(seqnames="chr1", tx_start=11:15, tx_end=12:16, strand=c("\"+\", \"-\", \"\"+\", \"\"\", \"\"\"\")}, score=1:5)
makeGRangesFromDataFrame(df)

## The common prefix between the start and end columns is used to
## disambiguate between more than one seqnames column:
## df <- data.frame(chrom="chr1", tx_start=11:15, tx_end=12:16, tx_chr="chr2", score=1:5)
makeGRangesFromDataFrame(df)

## 0-BASED VS 1-BASED START POSITIONS
## ---------------------------------------------------------------------
if (require(rtracklayer)) {
  session <- browserSession()
genome(session) <- "sacCer2"
query <- ucscTableQuery(session, "Most Conserved")
df <- getTable(query)

## A common pitfall is to forget that the UCSC Table Browser uses the
## 0-based start" convention:
gr0 <- makeGRangesFromDataFrame(df, keep.extra.columns=TRUE)
head(gr0)
min(start(gr0))

## The start positions need to be converted into 1-based positions,
## to adhere to the convention used in Bioconductor:
gr1 <- makeGRangesFromDataFrame(df, keep.extra.columns=TRUE,
                                  starts.in.df.are.0based=TRUE)
head(gr1)
}

makeGRangesListFromDataFrame

Make a GRangesList object from a data.frame or DataFrame

Description

makeGRangesListFromDataFrame extends the makeGRangesFromDataFrame functionality from GenomicRanges. It can take a data-frame-like object as input and tries to automatically find the
columns that describe the genomic ranges. It returns a GRangesList object. This is different from
the makeGRangesFromDataFrame function by requiring a split.field. The split.field acts
like the "f" argument in the split function. This factor must be of the same length as the number
of rows in the DataFrame argument. The split.field may also be a character vector.

Usage

makeGRangesListFromDataFrame(df,  
    split.field = NULL,  
    names.field = NULL,  
    ...)  

Arguments

df  A DataFrame or data.frame class object

split.field  A character string of a recognized column name in df that contains the grouping.
This column defines how the rows of df are split and is typically a factor or character vector. When split.field is not provided the df will be split by
the number of rows.

names.field  An optional single character string indicating the name of the column in df that designates the names for the ranges in the elements of the GRangesList.

...  Additional arguments passed on to makeGRangesFromDataFrame

Value

A GRangesList of the same length as the number of levels or unique character strings in the df column indicated by split.field. When split.field is not provided the df is split by row and the resulting GRangesList has the same length as nrow(df).

Names on the individual ranges are taken from the names.field argument. Names on the outer list elements of the GRangesList are propagated from split.field.

Author(s)

M. Ramos

See Also

• makeGRangesFromDataFrame

Examples

## BASIC EXAMPLES

```r
df <- data.frame(chr="chr1", start=11:15, end=12:16,  
    strand=c("+","-","+","*","-"), score=1:5,  
    specimen = c("a", "a", "b", "b", "c"),  
    gene_symbols = paste0("GENE", letters[1:5]))

df

grl <- makeGRangesListFromDataFrame(df, split.field = "specimen",  
    names.field = "gene_symbols")

grl
```

## Keep metadata columns

```r
makeGRangesListFromDataFrame(df, split.field = "specimen",
               keep.extra.columns = TRUE)
```

---

### Finding the nearest genomic range neighbor

#### Description

The nearest, precede, follow, distance and distanceToNearest methods for `GenomicRanges` objects and subclasses.

#### Usage

```r
## S4 method for signature 'GenomicRanges,GenomicRanges'
precede(x, subject,
select=c("first", "all"), ignore.strand=FALSE)
## S4 method for signature 'GenomicRanges,missing'
precede(x, subject,
select=c("first", "all"), ignore.strand=FALSE)

## S4 method for signature 'GenomicRanges,GenomicRanges'
follow(x, subject,
select=c("last", "all"), ignore.strand=FALSE)
## S4 method for signature 'GenomicRanges,missing'
follow(x, subject,
select=c("last", "all"), ignore.strand=FALSE)

## S4 method for signature 'GenomicRanges,GenomicRanges'
nearest(x, subject,
select=c("arbitrary", "all"), ignore.strand=FALSE)
## S4 method for signature 'GenomicRanges,missing'
nearest(x, subject,
select=c("arbitrary", "all"), ignore.strand=FALSE)

## S4 method for signature 'GenomicRanges,GenomicRanges'
distanceToNearest(x, subject,
ignore.strand=FALSE, ...)
## S4 method for signature 'GenomicRanges,missing'
distanceToNearest(x, subject,
ignore.strand=FALSE, ...)

## S4 method for signature 'GenomicRanges,GenomicRanges'
distance(x, y,
ignore.strand=FALSE, ...)
```

#### Arguments

- **x**
  - The query `GenomicRanges` instance.
The subject GenomicRanges instance within which the nearest neighbors are found. Can be missing, in which case x is also the subject.

For the distance method, a GRanges instance. Cannot be missing. If x and y are not the same length, the shortest will be recycled to match the length of the longest.

Logic for handling ties. By default, all methods select a single interval (arbitrary for nearest, the first by order in subject for precede, and the last for follow). When select="all" a Hits object is returned with all matches for x.

A logical indicating if the strand of the input ranges should be ignored. When TRUE, strand is set to ' + '.

Additional arguments for methods.

Details

• nearest: Performs conventional nearest neighbor finding. Returns an integer vector containing the index of the nearest neighbor range in subject for each range in x. If there is no nearest neighbor NA is returned. For details of the algorithm see the man page in IRanges, ?nearest.

• precede: For each range in x, precede returns the index of the range in subject that is directly preceded by the range in x. Overlapping ranges are excluded. NA is returned when there are no qualifying ranges in subject.

• follow: The opposite of precede, follow returns the index of the range in subject that is directly followed by the range in x. Overlapping ranges are excluded. NA is returned when there are no qualifying ranges in subject.

• Orientation and strand for precede and follow: Orientation is 5' to 3', consistent with the direction of translation. Because positional numbering along a chromosome is from left to right and transcription takes place from 5' to 3', precede and follow can appear to have 'opposite' behavior on the + and - strand. Using positions 5 and 6 as an example, 5 precedes 6 on the + strand but follows 6 on the - strand.

The table below outlines the orientation when ranges on different strands are compared. In general, a feature on * is considered to belong to both strands. The single exception is when both x and subject are * in which case both are treated as +.

<table>
<thead>
<tr>
<th>x</th>
<th>subject</th>
<th>orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>a)</td>
<td>+</td>
<td>--- &gt;</td>
</tr>
<tr>
<td>b)</td>
<td>+</td>
<td>NA</td>
</tr>
<tr>
<td>c)</td>
<td>+</td>
<td>--- &gt;</td>
</tr>
<tr>
<td>d)</td>
<td>-</td>
<td>NA</td>
</tr>
<tr>
<td>e)</td>
<td>-</td>
<td>&lt; ---</td>
</tr>
<tr>
<td>f)</td>
<td>-</td>
<td>&lt; ---</td>
</tr>
<tr>
<td>g)</td>
<td>*</td>
<td>--- &gt;</td>
</tr>
<tr>
<td>h)</td>
<td>*</td>
<td>&lt; ---</td>
</tr>
<tr>
<td>i)</td>
<td>*</td>
<td>--- &gt; (the only situation where * arbitrarily means +)</td>
</tr>
</tbody>
</table>

• distanceToNearest: Returns the distance for each range in x to its nearest neighbor in the subject.

• distance: Returns the distance for each range in x to the range in y. The behavior of distance has changed in Bioconductor 2.12. See the man page ?distance in IRanges for details.
Value

For nearest, precede and follow, an integer vector of indices in subject, or a Hits if select="all".
For distanceToNearest, a Hits object with a column for the query index (queryHits), subject index (subjectHits) and the distance between the pair.
For distance, an integer vector of distances between the ranges in x and y.

Author(s)

P. Aboyoun and V. Obenchain <vobencha@fhcrc.org>

See Also

• The GenomicRanges and GRanges classes.
• The Ranges class in the IRanges package.
• The Hits class in the S4Vectors package.
• The nearest-methods man page in the IRanges package.
• findOverlaps-methods for finding just the overlapping ranges.
• The nearest-methods man page in the GenomicFeatures package.

Examples

```r
## -----------------------------------------------------------
## precede() and follow()
## -----------------------------------------------------------
query <- GRanges("A", IRanges(c(5, 20), width=1), strand="+")
subject <- GRanges("A", IRanges(rep(c(10, 15), 2), width=1),
                     strand=c("+", "+", "-", ")
precede(query, subject)
follow(query, subject)

strand(query) <- "-"
precede(query, subject)
follow(query, subject)

## ties choose first in order
query <- GRanges("A", IRanges(10, width=1), c("+", "-", "*")
subject <- GRanges("A", IRanges(c(5, 5, 15, 15, 15), width=1),
                     rep(c("+", ",", "*"), 2))
precede(query, subject)
precede(query, rev(subject))

## ignore.strand=TRUE treats all ranges as '+'
precede(query[1], subject[4:6], select="all", ignore.strand=FALSE)
precede(query[1], subject[4:6], select="all", ignore.strand=TRUE)

## when multiple ranges overlap an "arbitrary" range is chosen
query <- GRanges("A", IRanges(5, 15))
subject <- GRanges("A", IRanges(c(1, 15), c(5, 19)))
nearest(query, subject)
```

Calculate the "phi coefficient" between two binary variables

Description

The `phicoef` function calculates the "phi coefficient" between two binary variables.

Usage

`phicoef(x, y=NULL)`

Arguments

- `x, y` Two logical vectors of the same length. If `y` is not supplied, `x` must be a 2x2 integer matrix (or an integer vector of length 4) representing the contingency table of two binary variables.

Value

The "phi coefficient" between the two binary variables. This is a single numeric value ranging from -1 to +1.
range-squeezer

Author(s)
H. Pagès

References
http://en.wikipedia.org/wiki/Phi_coefficient

Examples
```r
set.seed(33)
x <- sample(c(TRUE, FALSE), 100, replace=TRUE)
y <- sample(c(TRUE, FALSE), 100, replace=TRUE)
phicoef(x, y)
phicoef(rep(x, 10), c(rep(x, 9), y))
stopifnot(phicoef(table(x, y)) == phicoef(x, y))
stopifnot(phicoef(y, x) == phicoef(x, y))
stopifnot(phicoef(x, !y) == - phicoef(x, y))
stopifnot(phicoef(x, x) == 1)
```

Description
S4 generic functions for squeezing the ranges out of a range-based object. granges returns them in a GRanges object, grglist in a GRangesList object, and rglist in a RangesList object.

Usage
```r
granges(x, use.names=TRUE, use.mcols=FALSE, ...)
grglist(x, use.names=TRUE, use.mcols=FALSE, ...)
rglist(x, use.names=TRUE, use.mcols=FALSE, ...)
```

Arguments
- `x` A range-based object e.g. a RangedSummarizedExperiment, GAlignments, GAlignmentPairs, GAlignmentsList or a Pairs object containing ranges.
- `use.names` TRUE (the default) or FALSE. Whether or not the names on x (accessible with names(x)) should be propagated to the returned object.
- `use.mcols` TRUE or FALSE (the default). Whether or not the metadata columns on x (accessible with mcols(x)) should be propagated to the returned object.
- `...` Additional arguments, for use in specific methods.

Details
The GenomicRanges, SummarizedExperiment, and GenomicAlignments packages define and document methods for various types of range-based objects (e.g. for RangedSummarizedExperiment, GAlignments, GAlignmentPairs, and GAlignmentsList objects). Other Bioconductor packages might as well.
Note that these functions can be seen as a specific kind of *object getters* as well as functions performing coercion.

For some objects (e.g. `GAlignments` and `GAlignmentPairs` objects defined in the `GenomicAlignments` package), `as(x, "GRanges"), as(x, "GRangesList"), and as(x, "RangesList"),` are equivalent to `granges(x, use.names=TRUE, use.mcols=TRUE), grglist(x, use.names=TRUE, use.mcols=TRUE),` and `rglist(x, use.names=TRUE, use.mcols=TRUE),` respectively.

**Value**

A `GRanges` object for `granges`.

A `GRangesList` object for grglist.

A `RangesList` object for rglist.

If `x` is a vector-like object (e.g. `GAlignments`), the returned object is expected to be *parallel* to `x`, that is, the i-th element in the output corresponds to the i-th element in the input.

If `use.names` is TRUE, then the names on `x` (if any) are propagated to the returned object. If `use.mcols` is TRUE, then the metadata columns on `x` (if any) are propagated to the returned object.

**Author(s)**

H. Pagès

**See Also**

- `GRanges` and `GRangesList` objects.
- `RangesList` objects in the `IRanges` package.
- `RangedSummarizedExperiment` objects in the `SummarizedExperiment` packages.
- `GAlignments`, `GAlignmentPairs`, and `GAlignmentsList` objects in the `GenomicAlignments` package.

**Examples**

```r
## See ?GAlignments in the GenomicAlignments package for some
## examples.
```

---

**Description**

Performs set operations on `GRanges` and `GRangesList` objects.

NOTE: The `punion`, `pintersect`, `psetdiff`, and `pgap` generic functions and methods for `Ranges` objects are defined and documented in the `IRanges` package.
Usage

## Vector-wise set operations
## --------------------------

## S4 method for signature 'GenomicRanges,GenomicRanges'
union(x, y, ignore.strand=FALSE)

## S4 method for signature 'GenomicRanges,GenomicRanges'
intersect(x, y, ignore.strand=FALSE)

## S4 method for signature 'GenomicRanges,GenomicRanges'
setdiff(x, y, ignore.strand=FALSE)

## Element-wise (aka "parallel") set operations
## --------------------------------------------

## S4 method for signature 'GRanges,GRanges'
punion(x, y, fill.gap=FALSE, ignore.strand=FALSE)

## S4 method for signature 'GRanges,GRanges'
pintersect(x, y, drop.nohit.ranges=FALSE, ignore.strand=FALSE, strict.strand=FALSE)

## S4 method for signature 'GRanges,GRanges'
psetdiff(x, y, ignore.strand=FALSE)

Arguments

x, y For union, intersect, and setdiff: 2 GenomicRanges objects or 2 GRanges-List objects.
For punion and pintersect: 2 GRanges objects, or 1 GRanges object and 1 GRangesList object.
For psetdiff: x must be a GRanges object and y can be a GRanges or GRangesList object.
For pgap: 2 GRanges objects.

In addition, for the parallel operations, x and y must be of equal length (i.e. length(x) == length(y)).

fill.gap Logical indicating whether or not to force a union by using the rule start = min(start(x), start(y)),

ignore.strand For set operations: If set to TRUE, the strand of x and y is set to "*" prior to any computation.
For parallel set operations: If set to TRUE, the strand information is ignored in the computation and the result has the strand information of x.

drop.nohit.ranges 

If TRUE then elements in x that don’t intersect with their corresponding element in y are removed from the result (so the returned object is no more parallel to the input).
If FALSE (the default) then nothing is removed and a hit metadata column is added to the returned object to indicate elements in x that intersect with the corresponding element in y. For those that don’t, the reported intersection is a zero-width range that has the same start as x.
strict.strand  If set to FALSE (the default), features on the "*" strand are treated as occurring on both the "+" and "-" strand. If set to TRUE, the strand of intersecting elements must be strictly the same.

Details

The pintersect methods involving GRanges and/or GRangesList objects use the triplet (sequence name, range, strand) to determine the element by element intersection of features, where a strand value of "*" is treated as occurring on both the "+" and "-" strand (unless strict.strand is set to TRUE, in which case the strand of intersecting elements must be strictly the same).

The psetdiff methods involving GRanges and/or GRangesList objects use the triplet (sequence name, range, strand) to determine the element by element set difference of features, where a strand value of "*" is treated as occurring on both the "+" and "-" strand.

Value

For union, intersect, and setdiff: a GRanges object if x and y are GenomicRanges objects, and a GRangesList object if they are GRangesList objects.

For punion and pintersect: when x or y is not a GRanges object, an object of the same class as this non-GRanges object. Otherwise, a GRanges object.

For psetdiff: either a GRanges object when both x and y are GRanges objects, or a GRangesList object when y is a GRangesList object.

For pgap: a GRanges object.

Author(s)

P. Aboyoun and H. Pagès

See Also

• setops-methods in the IRanges package for set operations on Ranges and RangesList objects.
• findOverlaps-methods for finding/counting overlapping genomic ranges.
• intra-range-methods and inter-range-methods for intra range and inter range transformations of a GRanges object.
• GRanges and GRangesList objects.
• mendoapply in the S4Vectors package.

Examples

```r
## ---------------------------------------------------------------------
## A. SET OPERATIONS
## ---------------------------------------------------------------------

x <- GRanges("chr1", IRanges(c(2, 9) , c(7, 19)), strand=c("+", "-"))
y <- GRanges("chr1", IRanges(5, 10), strand="-")
union(x, y)
union(x, y, ignore.strand=TRUE)

intersect(x, y)
intersect(x, y, ignore.strand=TRUE)
```
setdiff(x, y)
setdiff(x, y, ignore.strand=TRUE)

## With 2 GRangesList objects:
gr1 <- GRanges(seqnames="chr2",
ranges=IRanges(3, 6))
gr2 <- GRanges(seqnames=c("chr1", "chr1"),
ranges=IRanges(c(7,13), width = 3),
strand=c("+", "-"))
gr3 <- GRanges(seqnames=c("chr1", "chr2"),
ranges=IRanges(c(4, 9)),
strand=c("-", "-"))
grlist <- GRangesList(gr1=gr1, gr2=gr2, gr3=gr3)

union(grlist, shift(grlist, 3))
intersect(grlist, shift(grlist, 3))
setdiff(grlist, shift(grlist, 3))

## Sanity checks:
grlist2 <- shift(grlist, 3)
stopifnot(identical(
  union(grlist, grlist2),
  mendoapply(union, grlist, grlist2)
))
stopifnot(identical(
  intersect(grlist, grlist2),
  mendoapply(intersect, grlist, grlist2)
))
stopifnot(identical(
  setdiff(grlist, grlist2),
  mendoapply(setdiff, grlist, grlist2)
))

## B. PARALLEL SET OPERATIONS

punion(x, shift(x, 6))
## Not run:
punion(x, shift(x, 7)) # will fail

## End(Not run)
punion(x, shift(x, 7), fill.gap=TRUE)

pintersect(x, shift(x, 6))
pintersect(x, shift(x, 7))
psetdiff(x, shift(x, 7))

## C. MORE EXAMPLES

## GRanges object:
gr <- GRanges(seqnames=c("chr2", "chr1", "chr1"),
ranges=IRanges(1:3, width = 12),
strand=Rle(strand(c("-", "*", "-"))))
## Parallel intersection of a GRanges and a GRangesList object

\[
\text{piintersect(gr, grlist)} \quad \text{piintersect(grlist, gr)}
\]

## Parallel set difference of a GRanges and a GRangesList object

\[
\text{psetdiff(gr, grlist)}
\]

---

Description

A bunch of useful strand and invertStrand methods.

Usage

```r
## S4 method for signature 'missing'
strand(x)
## S4 method for signature 'character'
strand(x)
## S4 method for signature 'factor'
strand(x)
## S4 method for signature 'integer'
strand(x)
## S4 method for signature 'logical'
strand(x)
## S4 method for signature 'Rle'
strand(x)
## S4 method for signature 'RleList'
strand(x)
## S4 method for signature 'DataTable'
strand(x)
## S4 replacement method for signature 'DataTable,ANY'
strand(x) <- value
## S4 method for signature 'character'
invertStrand(x)
## S4 method for signature 'factor'
invertStrand(x)
## S4 method for signature 'integer'
invertStrand(x)
## S4 method for signature 'logical'
invertStrand(x)
## S4 method for signature 'Rle'
invertStrand(x)
```
### S4 method for signature 'RleList'

**invertStrand(x)**

**Arguments**

- **x**
  - The object from which to obtain a *strand factor*, *strand factor Rle*, or *strand factor RleList* object. Can be missing. See Details and Value sections below for more information.

- **value**
  - Replacement value for the strand.

**Details**

All the *strand* and *invertStrand* methods documented here return either a *strand factor*, *strand factor Rle*, or *strand factor RleList* object. These are factor, factor-Rle, or factor-RleList objects containing the "standard strand levels" (i.e. +, -, and *) and no NAs.

**Value**

All the *strand* and *invertStrand* methods documented here return an object that is _parallel_ to input object x when x is a character, factor, integer, logical, Rle, or RleList object.

For the *strand* methods:

- If x is missing, returns an empty factor with the "standard strand levels" i.e. +, -, and *.
- If x is a character vector or factor, it is coerced to a factor with the levels listed above. NA values in x are not accepted.
- If x is an integer vector, it is coerced to a factor with the levels listed above. 1, -1, and NA values in x are mapped to the +, -, and * levels respectively.
- If x is a logical vector, it is coerced to a factor with the levels listed above. FALSE, TRUE, and NA values in x are mapped to the +, -, and * levels respectively.
- If x is a character-, factor-, integer-, or logical-Rle, it is transformed with runValue(x) <- strand(runValue(x)) and returned.
- If x is an RleList object, each list element in x is transformed by calling strand() on it and the resulting RleList object is returned. More precisely the returned object is endoapply(x, strand). Note that in addition to being _parallel_ to x, this object also has the same _shape_ as x (i.e. its list elements have the same lengths as in x).
- If x inherits from DataTable, the "strand" column is passed thru strand() and returned. If x has no "strand" column, this return value is populated with *s.

Each *invertStrand* method returns the same object as its corresponding *strand* method but with "+" and "-" switched.

**Author(s)**

M. Lawrence and H. Pagès

**See Also**

strand
Examples

```r
strand()

x1 <- c("-", "+", "+", "+-", "+-")
x2 <- factor(c("-", "+", "+-", "+"))
x3 <- c(-1L, NA, NA, 1L, -1L, NA)
x4 <- c(TRUE, NA, NA, FALSE, TRUE, NA)

strand(x1)
invertStrand(x1)
strand(x2)
invertStrand(x2)
strand(x3)
invertStrand(x3)
strand(x4)
invertStrand(x4)

strand(Rle(x1))
invertStrand(Rle(x1))
strand(Rle(x2))
invertStrand(Rle(x2))
strand(Rle(x3))
invertStrand(Rle(x3))
strand(Rle(x4))
invertStrand(Rle(x4))

x5 <- RleList(x1, character(0), as.character(x2))
strand(x5)
invertStrand(x5)

strand(DataFrame(score=2:-3))
strand(DataFrame(score=2:-3, strand=x3))
strand(DataFrame(score=2:-3, strand=Rle(x3)))

## Sanity checks:
target <- strand(x1)
stopifnot(identical(target, strand(x3)))
stopifnot(identical(target, strand(x4)))

stopifnot(identical(Rle(strand(x1)), strand(Rle(x1))))
stopifnot(identical(Rle(strand(x2)), strand(Rle(x2))))
stopifnot(identical(Rle(strand(x3)), strand(Rle(x3))))
stopifnot(identical(Rle(strand(x4)), strand(Rle(x4))))
```

tile-methods

Generate windows for a GenomicRanges

Description

tile and slidingWindows methods for GenomicRanges. tile partitions each range into a set of tiles, which are defined in terms of their number or width. slidingWindows generates sliding windows of a specified width and frequency.
## Usage

```r
## S4 method for signature 'GenomicRanges'
tile(x, n, width)
## S4 method for signature 'GenomicRanges'
slidingWindows(x, width, step=1L)
```

### Arguments

- **x**
  A `GenomicRanges` object, like a `GRanges`.

- **n**
  The number of tiles to generate. See `tile` in the `IRanges` package for more information about this argument.

- **width**
  The (maximum) width of each tile. See `tile` in the `IRanges` package for more information about this argument.

- **step**
  The distance between the start positions of the sliding windows.

### Details

The `tile` function splits `x` into a `GRangesList`, each element of which corresponds to a tile, or partition, of `x`. Specify the tile geometry with either `n` or `width` (not both). Passing `n` creates `n` tiles of approximately equal width, truncated by sequence end, while passing `width` tiles the region with ranges of the given width, again truncated by sequence end.

The `slidingWindows` function generates sliding windows within each range of `x`, according to `width` and `step`, returning a `GRangesList`. If the sliding windows do not exactly cover a range in `x`, the last window is partial.

### Value

A `GRangesList` object, each element of which corresponds to a window.

### Author(s)

M. Lawrence

### See Also

`tile` in the `IRanges` package.

### Examples

```r
gr <- GRanges(
  seqnames=Rle(c("chr1", "chr2", "chr1", "chr3"), c(1, 3, 2, 4)),
  ranges=IRanges(1:10, end=11),
  strand=Rle(strand(c("-", "+", "x", "+-", "-")), c(1, 2, 2, 3, 2)),
  seqlengths=c(chr1=11, chr2=12, chr3=13))
# split every range in half
tiles <- tile(gr, n = 2L)
stopifnot(all(elementNROWS(tiles) == 2L))
# split ranges into subranges of width 2
# odd width ranges must contain one subrange of width 1
tiles <- tile(gr, width = 2L)
stopifnot(all(all(width(tiles) %in% c(1L, 2L))))
```
tileGenome <- slidingWindows(gr, width=3L, step=2L)
width(windows[[1L]]) # last range is truncated

---

**tileGenome**

*Put (virtual) tiles on a given genome*

**Description**

tileGenome returns a set of genomic regions that form a partitioning of the specified genome. Each region is called a "tile".

**Usage**

tileGenome(seqlengths, ntile, tilewidth, cut.last.tile.in.chrom=FALSE)

**Arguments**

- `seqlengths`: Either a named numeric vector of chromosome lengths or a Seqinfo object. More precisely, if a named numeric vector, it must have a length \( \geq 1 \), cannot contain NAs or negative values, and cannot have duplicated names. If a Seqinfo object, then it’s first replaced with the vector of sequence lengths stored in the object (extracted from the object with the `seqlengths` getter), then the restrictions described previously apply to this vector.

- `ntile`: The number of tiles to generate.

- `tilewidth`: The desired tile width. The effective tile width might be slightly different but is guaranteed to never be more than the desired width.

- `cut.last.tile.in.chrom`: Whether or not to cut the last tile in each chromosome. This is set to FALSE by default. Can be set to TRUE only when `tilewidth` is specified. In that case, a tile will never overlap with more than 1 chromosome and a GRanges object is returned with one element (i.e. one genomic range) per tile.

**Value**

If `cut.last.tile.in.chrom` is FALSE (the default), a GRangesList object with one list element per tile, each of them containing a number of genomic ranges equal to the number of chromosomes it overlaps with. Note that when the tiles are small (i.e. much smaller than the chromosomes), most of them only overlap with a single chromosome.

If `cut.last.tile.in.chrom` is TRUE, a GRanges object with one element (i.e. one genomic range) per tile.

**Author(s)**

H. Pagès, based on a proposal by M. Morgan
See Also

- `genomicvars` for an example of how to compute the binned average of a numerical variable defined along a genome.
- `GRangesList` and `GRanges` objects.
- `Seqinfo` objects and the `seqlengths` getter.
- `IntegerList` objects.
- `Views` objects.

Examples

```r
## A. WITH A TOY GENOME
##
seqlengths <- c(chr1=60, chr2=20, chr3=25)

## Create 5 tiles:
tiles <- tileGenome(seqlengths, ntile=5)
tiles
elementNROWS(tiles) # tiles 3 and 4 contain 2 ranges
width(tiles)
## Use sum() on this IntegerList object to get the effective tile widths:
sum(width(tiles)) # each tile covers exactly 21 genomic positions

## Create 9 tiles:
tiles <- tileGenome(seqlengths, ntile=9)
elementNROWS(tiles) # tiles 6 and 7 contain 2 ranges

table(sum(width(tiles))) # some tiles cover 12 genomic positions, others 11

## Specify the tile width:
tiles <- tileGenome(seqlengths, tilewidth=20)
length(tiles) # 6 tiles

table(sum(width(tiles))) # effective tile width is <= specified

## Specify the tile width and cut the last tile in each chromosome:
tiles <- tileGenome(seqlengths, tilewidth=24,
  cut.last.tile.in.chrom=TRUE)
tiles

table(sum(width(tiles))) # each tile covers exactly 24 genomic positions, except
  # the last tile in each chromosome

## Partition a genome by chromosome ("natural partitioning"):
tiles <- tileGenome(seqlengths, tilewidth=max(seqlengths),
  cut.last.tile.in.chrom=TRUE)
tiles # one tile per chromosome

## sanity check
stopifnot(all.equal(setNames(end(tiles), seqnames(tiles)), seqlengths))
```

## B. WITH A REAL GENOME
```
### C. AN APPLICATION: COMPUTE THE BINNED AVERAGE OF A NUMERICAL VARIABLE DEFINED ALONG A GENOME

See `?genomicvars` for an example of how to compute the binned average of a numerical variable defined along a genome.
Index

*Topic **classes**
  Constraints, 5
  GNCLList-class, 23
  GPos-class, 26

*Topic **manip**
  absoluteRanges, 2
  genomicvars, 19
  makeGRangesFromDataframe, 48
  phicof, 56
  tileGenome, 66

*Topic **methods**
  Constraints, 5
  coverage-methods, 11
  findOverlaps-methods, 13
  GenomicRanges-comparison, 15
  GNCLList-class, 23
  GPos-class, 26
  intra-range-methods, 45
  range-squeezers, 57
  setops-methods, 58
  strand-utils, 62
  tile-methods, 64

*Topic **utilities**
  coverage-methods, 11
  findOverlaps-methods, 13
  inter-range-methods, 41
  intra-range-methods, 45
  nearest-methods, 53
  setops-methods, 58
  tile-methods, 64

[, GRangesList, ANY-method
  (GRangesList-class), 37
[, GenomicRanges, ANY-method
  (GRanges-class), 30
[, List, GenomicRanges-method
  (GRanges-class), 30
[, list, GenomicRanges-method
  (GRanges-class), 30
[<-, GRanges-method (GRanges-class), 30
[<-, GRangesList, ANY-method
  (GRanges-class), 30
[<-, GenomicRanges, ANY, ANY, ANY-method
  (GRanges-class), 30
[<-, GRangesList, ANY, ANY-method
  (GRangesList-class), 37
[<-, GRangesList, ANY-method
  (GRangesList-class), 37
$-, GenomicRanges-method (GRangesList-class), 37

absoluleRanges, 2, 35
as.character, GenomicRanges-method
  (GRanges-class), 30
as.data.frame, GenomicRanges-method
  (GRanges-class), 30
as.data.frame, GPos-method (GPos-class), 26
as.factor, GenomicRanges-method
  (GRanges-class), 30

bindAsGRanges (genomicvars), 19
binnedAverage (genomicvars), 19
c, GenomicRanges-method (GRanges-class), 30
c, GPos-method (GPos-class), 26
checkConstraint (Constraints), 5
class: Constraint (Constraints), 5
class: Constraint.OR_NULL (Constraints), 5
class: DelegatingGenomicRanges
  (DelegatingGenomicRanges-class), 12
class: GenomicRanges (GRanges-class), 30
class: GenomicRangesList
  (GenomicRangesList-class), 19
class: GNCLList (GNCLList-class), 23
class: GPos (GPos-class), 26
class: GRanges (GRanges-class), 30
class:GRangesList (GRangesList-class), 37
class:SimpleGenomicRangesList (GenomicRangesList-class), 19
coe.re, ANY, GenomicRanges-method (GRanges-class), 30
coe.re, character, GenomicRanges-method (GRanges-class), 30
coe.re, data.frame, GenomicRanges-method (makeGRangesFromDataFrame), 48
coe.re, factor, GenomicRanges-method (GRanges-class), 30
coe.re, GenomicRanges, GNCList-method (GNCList-class), 23
coe.re, GenomicRanges, GPos-method (GPos-class), 26
coe.re, GenomicRanges, GRanges-method (GRanges-class), 30
coe.re, GenomicRanges, Grouping-method (GRanges-class), 30
coe.re, GenomicRanges, RangedData-method (GRanges-class), 30
coe.re, GenomicRanges, RangesList-method (GRanges-class), 30
coe.re, GNCList, GRanges-method (GNCList-class), 23
coe.re, GRanges, GRangesList-method (GRangesList-class), 37
coe.re, GRangesList, CompressedIRangesList-method (GRangesList-class), 37
coe.re, GRangesList, IRangesList-method (GRangesList-class), 37
coe.re, GRangesList, RangesList-method (GRangesList-class), 37
coe.re, list, GRangesList-method (GRangesList-class), 37
coe.re, RangedData, GRanges-method (GRanges-class), 30
coe.re, RangesList, GRanges-method (GRanges-class), 30
coe.re, RleList, GRanges-method (genomicvvars), 19
coe.re, RleViewsList, GRanges-method (genomicvvars), 19
coe.re, Seqinfo, GRanges-method (GRanges-class), 30
coe.re, Seqinfo, RangesList-method (GRanges-class), 30
Constraint (Constraints), 5
constraint (Constraints), 5
Constraint-class (Constraints), 5
constraint<- (Constraints), 5
Constraint_OR_NULL (Constraints), 5
Constraint_OR_NULL-class (Constraints), 5
Constraints, 5
countOverlaps (findOverlaps-methods), 13
countOverlaps, GenomicRanges, GenomicRanges-method (findOverlaps-methods), 13
coverage, 11, 12
coverage (coverage-methods), 11
coverage, GenomicRanges-method, 21
coverage, GenomicRanges-method (coverage-methods), 11
coverage, GRangesList-method (coverage-methods), 11
coverage-methods, 11, 12, 35, 40
DataFrame, 30, 32–35, 38, 48–50
DataFrameList-class, 40
DataTable, 32
DelegatingGenomicRanges-class, 12
disjoin, 43
disjoin (inter-range-methods), 41
disjoin, GenomicRanges-method (inter-range-methods), 41
disjoin, GenomicRangesList-method (inter-range-methods), 41
disjoin, GRangesList-method (inter-range-methods), 41
disjointBins (inter-range-methods), 41
disjointBins, GenomicRanges-method (inter-range-methods), 41
distance (nearest-methods), 53
distance, GenomicRanges-method (GenomicRanges-comparison), 15
distanceToNearest (nearest-methods), 53
distanceToNearest, GenomicRanges-method (nearest-methods), 53
distanceToNearest, GenomicRanges, GenomicRanges-method (nearest-methods), 53
distanceToNearest, GenomicRanges, missing-method (nearest-methods), 53
DNAStringSet, 34
duplicated, GenomicRanges-method (GenomicRanges-comparison), 15
duplicated, GenomicRanges (GenomicRanges-comparison), 15
elementMetadata, GRangesList-method (GRangesList-class), 37
elementMetadata<-, GRangesList-method (GRangesList-class), 37
diff, GenomicRanges-method (GRanges-class), 30
...
invertStrand, integer-method (strand-utils), 62
invertStrand, logical-method (strand-utils), 62
invertStrand, NULL-method (strand-utils), 62
invertStrand, Rle-method (strand-utils), 62
invertStrand, RleList-method (strand-utils), 62
IRanges, 3, 4, 30, 34
IRangesList, 39
is, 6
is.unsorted, GenomicRanges-method (GenomicRanges-comparison), 15
isCircular, 26
isDisjoint (intra-range-methods), 41
isDisjoint, GenomicRanges-method (intra-range-methods), 41
isDisjoint, GenomicRangesList-method (intra-range-methods), 41
isDisjoint, GPos-method (intra-range-methods), 41
isDisjoint, GRangesList-method (intra-range-methods), 41
isSmallGenome (absoluteRanges), 2
lapply, 39, 40
length, GenomicRanges-method (GRanges-class), 30
length, GNCList-method (GNCList-class), 23
length, GPos-method (GPos-class), 26
List, 19, 34
makeGRangesFromDataFrame, 31, 34, 48, 51, 52
makeGRangesListFromDataFrame, 51
makeGRangesListFromFeatureFragments, 50
makeGRangesListFromFeatureFragments (GRangesList-class), 37
ManyToOneGrouping, 32
mapply, 40
match, GenomicRanges, GenomicRanges-method (GenomicRanges-comparison), 15
mcolAsRleList (genomicvars), 19
mendoapply, 60
names, GenomicRanges-method (GRanges-class), 30
names, GNCList-method (GNCList-class), 23
names, GPos-method (GPos-class), 26
names<-, GenomicRanges-method (GRanges-class), 30
names<-, GPos-method (GPos-class), 26
narrow (intra-range-methods), 45
narrow, GenomicRanges-method (intra-range-methods), 45
narrow, GenomicRangesList-method (intra-range-methods), 45
narrow, GRangesList-method (intra-range-methods), 45
NCList, 24
NCLists, 24
nearest (nearest-methods), 53
nearest, GenomicRanges, GenomicRanges-method (nearest-methods), 53
nearest, GenomicRanges, missing-method (nearest-methods), 53
nearest-methods, 27, 35, 53, 55
Ops, GenomicRanges, numeric-method (intra-range-methods), 45
order, GenomicRanges-method (GenomicRanges-comparison), 15
overlapsAny (findOverlaps-methods), 13
Pairs, 57
pcompare (GenomicRanges-comparison), 15
pcompare, GenomicRanges, GenomicRanges-method (GenomicRanges-comparison), 15
pgap, 58
pgap (setops-methods), 58
pgap, GRanges, GRanges-method (setops-methods), 58
phicoef, 56
pintersect, 58
pintersect (setops-methods), 58
pintersect, GRanges, GRanges-method (setops-methods), 58
pintersect, GRangesList-method (setops-methods), 58
pos (GPos-class), 26
pos, GPos-method (GPos-class), 26
precede (nearest-methods), 53
precede, GenomicRanges, GenomicRanges-method (nearest-methods), 53
precede, GenomicRanges, missing-method (nearest-methods), 53
promoters (intra-range-methods), 45
promoters, GenomicRanges-method (intra-range-methods), 45
promoters, GenomicRangesList-method (intra-range-methods), 45
promoters, GRangesList-method (intra-range-methods), 45
psetdiff, 58
psetdiff (setops-methods), 58
psetdiff, GRanges, GRanges-method (setops-methods), 58
psetdiff, GRangesList, GRangesList-method (setops-methods), 58
punion, 58
punion (setops-methods), 58
punion, GRanges, GRanges-method (setops-methods), 58
punion, GRangesList, GRanges-method (setops-methods), 58
range (inter-range-methods), 41
range, GenomicRanges-method (inter-range-methods), 41
range, GenomicRangesList-method (inter-range-methods), 41
range, GPos-method (inter-range-methods), 41
range, GRangesList-method (inter-range-methods), 41
range-squeezer, 57
RangedSummarizedExperiment, 57, 58
Ranges, 3, 11, 13, 24, 35, 43, 47, 55, 58, 60
ranges, DelegatingGenomicRanges-method (DelegatingGenomicRanges-class), 12
ranges, GNCList-method (GNCList-class), 23
ranges, GPos-method (GPos-class), 26
ranges, GRanges-method (GRanges-class), 30
ranges, GRangesList-method (GRangesList-class), 37
Ranges-comparison, 17
ranges<-, GenomicRanges-method (GRanges-class), 30
ranges<-, GRangesList-method (GRangesList-class), 37
RangesList, 11, 13, 24, 57, 58, 60
RangesList-class, 40
rank, GenomicRanges-method (GenomicRanges-comparison), 15
rankSeqLevels, 50
reduce, 42
reduce (inter-range-methods), 41
reduce, GenomicRanges-method (inter-range-methods), 41
reduce, GenomicRangesList-method (inter-range-methods), 41
reduce, GRangesList-method (inter-range-methods), 41
relativeRanges (absoluteRanges), 2
relistToClass, GRanges-method (GRangesList-class), 37
resize (intra-range-methods), 45
resize, GenomicRanges-method (intra-range-methods), 45
resize, GenomicRangesList-method (intra-range-methods), 45
resize, GRangesList-method (intra-range-methods), 45
restrict (intra-range-methods), 45
restrict, GenomicRanges-method (intra-range-methods), 45
restrict, GRangesList-method (intra-range-methods), 45
rglist (range-squeezer), 57
rglist, Pairs-method (range-squeezer), 57
Rle, 20, 30, 32, 35, 63
RleList, 12, 19–21, 63
RleList-class, 40
sapply, 40
score, GenomicRanges-method (GRanges-class), 30
score, GRangesList-method (GRangesList-class), 37
score<-, GenomicRanges-method (GRanges-class), 30
score<-, GRangesList-method (GRangesList-class), 37
selfmatch, GenomicRanges-method (GenomicRanges-comparison), 15
Seqinfo, 4, 30–32, 38, 49, 50, 66, 67
seqinfo, 26, 27, 34, 40
seqinfo, DelegatingGenomicRanges-method (DelegatingGenomicRanges-class), 12
seqinfo, GNCList-method (GNCList-class), 23
seqinfo, GPos-method (GPos-class), 26
seqinfo, GRanges-method (GRanges-class), 30
seqinfo, GRangesList-method (GRangesList-class), 37
seqinfo, List-method (GRanges-class), 30
seqinfo, RangedData-method (GRanges-class), 30
seqinfo, RangesList-method (GRanges-class), 30
seqinfo<-, GenomicRanges-method (GRanges-class), 30
seqinfo<-, GPos-method (GPos-class), 26
seqinfo<-, GRangesList-method (GRangesList-class), 37
seqinfo<-, List-method (GRanges-class), 30
seqinfo<-, RangedData-method (GRanges-class), 30
seqlengths, 3, 4, 26, 66, 67
seqlevels, 26, 32, 38
seqlevelsStyle, 26, 32, 38
seqnames, DelegatingGenomicRanges-method (DelegatingGenomicRanges-class), 12
seqnames, GNCList-method (GNCList-class), 23
seqnames, GPos-method (GPos-class), 26
seqnames, GRanges-method (GRanges-class), 30
seqnames, GRangesList-method (GRangesList-class), 37
seqnames<-, RangedData-method (GRanges-class), 30
seqnames<-, GenomicRanges-method (GRanges-class), 30
seqnames<-, GRangesList-method (GRangesList-class), 37
setClass, 6
setdiff (setops-methods), 58
setdiff, GenomicRanges, GenomicRanges-method (setops-methods), 58
setdiff, GenomicRanges, Vector-method (setops-methods), 58
setdiff, GRangesList, GRangesList-method (setops-methods), 58
setdiff, Vector, GenomicRanges-method (setops-methods), 58
setMethod, 6
setops-methods, 17, 35, 40, 58, 60
shift (intra-range-methods), 45
shift, GenomicRanges-method (intra-range-methods), 45
shift, GenomicRangesList-method (intra-range-methods), 45
shift, GPos-method (intra-range-methods), 45
show, GenomicRanges-method (GRanges-class), 30
show, GPos-method (GPos-class), 26
show, GRangesList-method (GRangesList-class), 37
showMethods, 6
SimpleGenomicRangesList-class (GenomicRangesList-class), 19
slidingWindows, 64
slidingWindows (tile-methods), 64
slidingWindows, GenomicRanges-method (tile-methods), 64
SNPlocs, 27
snpsById, 27
snpsByOverlaps, 27
snpsBySeqname, 27
sort, GenomicRanges-method (GenomicRanges-comparison), 15
sort, GenomicRanges (GenomicRanges-comparison), 15
split, 52
start, GenomicRanges-method (GRanges-class), 30
start, GNCList-method (GNCList-class), 23
start, GPos-method (GPos-class), 26
start, GRanges-method (GRanges-class), 30
start, GRangesList-method (GRangesList-class), 37
start<-, GenomicRanges-method (GRanges-class), 30
start<-, GRangesList-method (GRangesList-class), 37
strand, 30, 63
strand, character-method (strand-utils), 62
strand, DataTable-method (strand-utils), 62
strand, DelegatingGenomicRanges-method (DelegatingGenomicRanges-class), 12
strand, factor-method (strand-utils), 62
strand, GNCList-method (GNCList-class), 23
strand, GPos-method (GPos-class), 26
strand, GRanges-method (GRanges-class), 30
strand, GRangesList-method (GRangesList-class), 37
strand, integer-method (strand-utils), 62
strand, logical-method (strand-utils), 62
strand, missing-method (strand-utils), 62
strand,NULL-method (strand-utils), 62
strand,Rle-method (strand-utils), 62
strand,RleList-method (strand-utils), 62
strand-utils, 62
strand<-,DataTable,ANY-method (strand-utils), 62
strand<-,GenomicRanges,ANY-method (GRanges-class), 30
strand<-,GRangesList,ANY-method (GRangesList-class), 37
subsetByOverlaps (findOverlaps-methods), 13
SummarizedExperiment, 27
summary,GenomicRanges-method (GRanges-class), 30
summary.GenomicRanges (GRanges-class), 30
tile, 64, 65
tile (tile-methods), 64
tile,GenomicRanges-method (tile-methods), 64
tile-methods, 64
tileGenome, 4, 20, 21, 35, 66
trim,GenomicRanges-method (intra-range-methods), 45
union (setops-methods), 58
union,GenomicRanges,GenomicRanges-method (setops-methods), 58
union,GenomicRanges,Vector-method (setops-methods), 58
union,GRangesList,GRangesList-method (setops-methods), 58
union,Vector,GenomicRanges-method (setops-methods), 58
updateObject,GRanges-method (GRanges-class), 30
updateObject,GRangesList-method (GRangesList-class), 37
validObject, 6
Vector, 27, 35
Vector-class, 40
Views, 67
width,GenomicRanges-method (GRanges-class), 30
width,GNCList-method (GNCList-class), 23
width,GPos-method (GPos-class), 26
width,GRangesList-method (GRangesList-class), 37