

BSgenome.Mmulatta.UCSC.rheMac2

April 6, 2012

Mmulatta

Macaca mulatta (Rhesus) full genome (UCSC version rheMac2)

Description

Macaca mulatta (Rhesus) full genome as provided by UCSC (rheMac2, Jan. 2006) and stored in Biostrings objects. NOTE: In most assemblies available at UCSC, Tandem Repeats Finder repeats were filtered to retain only the repeats with period ≤ 12 . However, the filtering was omitted for this assembly, so the TRF masks contain all Tandem Repeats Finder results.

Note

This BSgenome data package was made from the following source data files:

sequences: chromFa.tar.gz, upstream1000.fa.gz, upstream2000.fa.gz, upstream5000.fa.gz
from <http://hgdownload.cse.ucsc.edu/goldenPath/rheMac2/bigZips/>
AGAPS masks: gap.txt.gz from <http://hgdownload.cse.ucsc.edu/goldenPath/rheMac2/database/>
RM and TRF masks: chromOut.tar.gz and chromTrf.tar.gz
from <http://hgdownload.cse.ucsc.edu/goldenPath/rheMac2/bigZips/>

See [?BSgenomeForge](#) and the BSgenomeForge vignette (vignette("BSgenomeForge")) in the BSgenome software package for how to make a BSgenome data package.

Author(s)

The Bioconductor Dev Team

See Also

[BSgenome-class](#), [DNAString-class](#), [available.genomes](#), [BSgenomeForge](#)

Examples

```

Mmulatta
seqlengths(Mmulatta)
Mmulatta$chr1 # same as Mmulatta[["chr1"]]
## NOTE: In most assemblies available at UCSC, Tandem Repeats
## Finder repeats were filtered to retain only the repeats
## with period <= 12. However, the filtering was omitted for
## this assembly, so, despite the description being displayed
## for this mask, it contains all the Tandem Repeats Finder
## results.
masks(Mmulatta$chr1)$TRF
## To get rid of the masks altogether:
unmasked(Mmulatta$chr1)

if ("AGAPS" %in% masknames(Mmulatta)) {

  ## Check that the assembly gaps contain only Ns:
  checkOnlyNsInGaps <- function(seq)
  {
    ## Replace all masks by the inverted AGAPS mask
    masks(seq) <- gaps(masks(seq)["AGAPS"])
    unique_letters <- uniqueLetters(seq)
    if (any(unique_letters != "N"))
      stop("assembly gaps contain more than just Ns")
  }

  ## A message will be printed each time a sequence is removed
  ## from the cache:
  options(verbose=TRUE)

  for (seqname in seqnames(Mmulatta)) {
    cat("Checking sequence", seqname, "... ")
    seq <- Mmulatta[[seqname]]
    checkOnlyNsInGaps(seq)
    cat("OK\n")
  }
}

## See the GenomeSearching vignette in the BSgenome software
## package for some examples of genome-wide motif searching using
## Biostrings and the BSgenome data packages:
if (interactive())
  vignette("GenomeSearching", package="BSgenome")

```

Index

*Topic **data**

Mmulatta, [1](#)

*Topic **package**

Mmulatta, [1](#)

available.genomes, [1](#)

BSgenome-class, [1](#)

BSgenome.Mmulatta.UCSC.rheMac2
(Mmulatta), [1](#)

BSgenome.Mmulatta.UCSC.rheMac2-package
(Mmulatta), [1](#)

BSgenomeForge, [1](#)

DNASTring-class, [1](#)

Mmulatta, [1](#)