

BSgenome.Mmusculus.UCSC.mm10.masked

December 11, 2024

BSgenome.Mmusculus.UCSC.mm10.masked

Full masked genome sequences for Mus musculus (UCSC genome mm10, based on GRCm38.p6)

Description

Full genome sequences for *Mus musculus* (Mouse) as provided by UCSC (genome mm10, based on GRCm38.p6) and stored in Biostrings objects. The sequences are the same as in BSgenome.Mmusculus.UCSC.mm10, except that each of them has the 2 following masks on top: (1) the mask of assembly gaps (AGAPS mask), and (2) the mask of intra-contig ambiguities (AMB mask).

Note

The masks in this BSgenome data package were made from the following source data files:

AGAPS masks: gap.txt.gz from <http://hgdownload.cse.ucsc.edu/goldenPath/mm10/database/>

See [?BSgenome.Mmusculus.UCSC.mm10](#) in the **BSgenome.Mmusculus.UCSC.mm10** package for information about how the sequences were obtained.

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.Mmusculus.UCSC.mm10](#) in the **BSgenome.Mmusculus.UCSC.mm10** package for information about how the sequences were obtained.
- [BSgenome](#) objects and the [available.genomes](#) function in the **BSgenome** software package.
- [MaskedDNAString](#) objects in the **Biostrings** package.
- The BSgenomeForge vignette (`vignette("BSgenomeForge")`) in the **BSgenome** software package for how to make a BSgenome data package.

Examples

```
BSgenome.Mmusculus.UCSC.mm10.masked
genome <- BSgenome.Mmusculus.UCSC.mm10.masked
head(seqlengths(genome))
genome$chr1 # a MaskedDNAString object!
## To get rid of the masks altogether:
unmasked(genome$chr1) # same as BSgenome.Mmusculus.UCSC.mm10$chr1

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

  ## 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(genome)) {
    cat("Checking sequence", seqname, "... ")
    seq <- genome[[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

* **data**

BSgenome.Mmusculus.UCSC.mm10.masked,
[1](#)

* **package**

BSgenome.Mmusculus.UCSC.mm10.masked,
[1](#)

available.genomes, [1](#)

BSgenome, [1](#)

BSgenome.Mmusculus.UCSC.mm10, [1](#)

BSgenome.Mmusculus.UCSC.mm10.masked, [1](#)

BSgenome.Mmusculus.UCSC.mm10.masked-package
(BSgenome.Mmusculus.UCSC.mm10.masked),
[1](#)

BSgenomeForge, [1](#)

MaskedDNAString, [1](#)