

# BSgenome.Drerio.UCSC.danRer5.masked

August 15, 2024

---

BSgenome.Drerio.UCSC.danRer5.masked

*Full masked genome sequences for Danio rerio (UCSC version danRer5)*

---

## Description

Full genome sequences for *Danio rerio* (Zebrafish) as provided by UCSC (danRer5, Jul. 2007) and stored in Biostrings objects. The sequences are the same as in BSgenome.Drerio.UCSC.danRer5, except that each of them has the 4 following masks on top: (1) the mask of assembly gaps (AGAPS mask), (2) the mask of intra-contig ambiguities (AMB mask), (3) the mask of repeats from Repeat-Masker (RM mask), and (4) the mask of repeats from Tandem Repeats Finder (TRF mask). Only the AGAPS and AMB masks are "active" by default.

## Note

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

AGAPS masks: <http://hgdownload.cse.ucsc.edu/goldenPath/danRer5/database/gap.txt.gz>

RM masks: <http://hgdownload.cse.ucsc.edu/goldenPath/danRer5/bigZips/danRer5.fa.out.gz>

TRF masks: <http://hgdownload.cse.ucsc.edu/goldenPath/danRer5/bigZips/danRer5.trf.bed.gz>

See [?BSgenome.Drerio.UCSC.danRer5](#) in the **BSgenome.Drerio.UCSC.danRer5** 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.Drerio.UCSC.danRer5](#) in the **BSgenome.Drerio.UCSC.danRer5** 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.Drerio.UCSC.danRer5.masked
genome <- BSgenome.Drerio.UCSC.danRer5.masked
seqlengths(genome)
genome$chr1 # a MaskedDNAString object!
## To get rid of the masks altogether:
unmasked(genome$chr1) # same as BSgenome.Drerio.UCSC.danRer5$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.Drerio.UCSC.danRer5.masked,  
[1](#)

\* **package**

BSgenome.Drerio.UCSC.danRer5.masked,  
[1](#)

available.genomes, [2](#)

BSgenome, [2](#)

BSgenome.Drerio.UCSC.danRer5, [1](#), [2](#)

BSgenome.Drerio.UCSC.danRer5.masked, [1](#)

BSgenome.Drerio.UCSC.danRer5.masked-package  
(BSgenome.Drerio.UCSC.danRer5.masked),  
[1](#)

BSgenomeForge, [1](#)

MaskedDNAString, [2](#)