

# BSgenome.Mmulatta.UCSC.rheMac2.masked

May 8, 2024

---

BSgenome.Mmulatta.UCSC.rheMac2.masked

*Full masked genome sequences for Macaca mulatta (UCSC version rheMac2)*

---

## Description

Full genome sequences for *Macaca mulatta* (Rhesus) as provided by UCSC (rheMac2, Jan. 2006) and stored in Biostrings objects. The sequences are the same as in BSgenome.Mmulatta.UCSC.rheMac2, 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: In most assemblies available at UCSC, Tandem Repeats Finder repeats were filtered to retain only the repeats with period  $\leq 12$ . However, the filtering was omitted for this assembly, so the TRF masks contain all Tandem Repeats Finder results.

## 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/rheMac2/database/>  
RM and TRF masks: chromOut.tar.gz and chromTrf.tar.gz  
from <http://hgdownload.cse.ucsc.edu/goldenPath/rheMac2/bigZips/>

See [?BSgenome.Mmulatta.UCSC.rheMac2](#) in the **BSgenome.Mmulatta.UCSC.rheMac2** 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.Mmulatta.UCSC.rheMac2](#) in the **BSgenome.Mmulatta.UCSC.rheMac2** 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.Mmulatta.UCSC.rheMac2.masked
genome <- BSgenome.Mmulatta.UCSC.rheMac2.masked
seqlengths(genome)
genome$chr1 # a MaskedDNAString object!
## NOTE: In most assemblies available at UCSC, Tandem Repeats
## Finder repeats were filtered to retain only the repeats
## with period <= 12. However, the filtering was omitted for
## this assembly, so, despite the description being displayed
## for this mask, it contains all the Tandem Repeats Finder
## results.
masks(genome$chr1)$TRF
## To get rid of the masks altogether:
unmasked(genome$chr1) # same as BSgenome.Mmulatta.UCSC.rheMac2$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.Mmulatta.UCSC.rheMac2.masked,  
[1](#)

\* **package**

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

available.genomes, [2](#)

BSgenome, [2](#)

BSgenome.Mmulatta.UCSC.rheMac2, [1](#), [2](#)

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

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

BSgenomeForge, [1](#)

MaskedDNAString, [2](#)