

BSgenome.Tguttata.UCSC.taeGut1.masked

October 7, 2014

BSgenome.Tguttata.UCSC.taeGut1.masked

Full masked genome sequences for Taeniopygia guttata (UCSC version taeGut1)

Description

Full genome sequences for *Taeniopygia guttata* (Zebra finch) as provided by UCSC (taeGut1, Jul. 2008) and stored in Biostrings objects. The sequences are the same as in BSgenome.Tguttata.UCSC.taeGut1, 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). Both masks are "active" by default.

Note

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

AGAPS masks: all the chr*_gap.txt.gz files from <http://hgdownload.cse.ucsc.edu/goldenPath/taeGut1/data/>

See [?BSgenome.Tguttata.UCSC.taeGut1](#) in the **BSgenome.Tguttata.UCSC.taeGut1** 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.Tguttata.UCSC.taeGut1](#) in the **BSgenome.Tguttata.UCSC.taeGut1** 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.Tguttata.UCSC.taeGut1.masked
genome <- BSgenome.Tguttata.UCSC.taeGut1.masked
seqlengths(genome)
genome$chr1 # a MaskedDNAString object!
## To get rid of the masks altogether:
unmasked(genome$chr1) # same as BSgenome.Tguttata.UCSC.taeGut1$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

*Topic **data**

BSgenome.Tguttata.UCSC.taeGut1.masked,

[1](#)

*Topic **package**

BSgenome.Tguttata.UCSC.taeGut1.masked,

[1](#)

available.genomes, [2](#)

BSgenome, [2](#)

BSgenome.Tguttata.UCSC.taeGut1, [1](#), [2](#)

BSgenome.Tguttata.UCSC.taeGut1.masked,

[1](#)

BSgenome.Tguttata.UCSC.taeGut1.masked-package

(BSgenome.Tguttata.UCSC.taeGut1.masked),

[1](#)

BSgenomeForge, [1](#)

MaskedDNAString, [2](#)