BSgenome.Hsapiens.UCSC.hg18

September 24, 2013

BSgenome.Hsapiens.UCSC.hg18

Homo sapiens (Human) full genome (UCSC version hg18)

Description

Homo sapiens (Human) full genome as provided by UCSC (hg18, Mar. 2006) and stored in Biostrings objects.

Note

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

```
sequences: chromFa.zip, upstream1000.zip, upstream2000.zip, upstream5000.zip
from http://hgdownload.cse.ucsc.edu/goldenPath/hg18/bigZips/
AGAPS masks: all the chr*_gap.txt.gz files from ftp://hgdownload.cse.ucsc.edu/goldenPath/hg18/database
RM masks: http://hgdownload.cse.ucsc.edu/goldenPath/hg18/bigZips/chromOut.zip
TRF masks: http://hgdownload.cse.ucsc.edu/goldenPath/hg18/bigZips/chromTrf.zip
```

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

```
BSgenome.Hsapiens.UCSC.hg18
genome <- BSgenome.Hsapiens.UCSC.hg18
seqlengths(genome)
genome$chr1  # same as genome[["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