BSgenome.Hsapiens.UCSC.hg38

January 21, 2026

BSgenome.Hsapiens.UCSC.hg38

Full genomic sequences for Homo sapiens (UCSC genome hg38)

Description

Full genomic sequences for Homo sapiens as provided by UCSC (genome hg38, based on assembly GRCh38.p14 since 2023/01/31). The sequences are stored in DNAString objects.

Note

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

hg38.p14.2bit, downloaded from https://hgdownload.soe.ucsc.edu/goldenPath/hg38/bigZips/p14/ on Feb

See ?BSgenomeForge and the BSgenomeForge vignette (vignette("BSgenomeForge")) in the **BSgenome** software package for how to create a BSgenome data package.

Author(s)

The Bioconductor Dev Team

See Also

- BSgenome objects in the **BSgenome** software package.
- The seqinfo getter and Seqinfo objects in the **GenomeInfoDb** package.
- The seqlevelsStyle getter and setter in the **GenomeInfoDb** package.
- DNAString objects in the **Biostrings** package.
- The available.genomes function in the **BSgenome** software package.
- The BSgenomeForge vignette (vignette("BSgenomeForge")) in the **BSgenome** software package for how to create a BSgenome data package.

Examples

```
BSgenome.Hsapiens.UCSC.hg38
bsg <- BSgenome.Hsapiens.UCSC.hg38
head(seqlengths(bsg))
seqinfo(bsg)

## Access individual sequences:

bsg$chr1  # same as bsg[["chr1"]]
alphabetFrequency(bsg[["chr1"]])

bsg[["chrM"]]  # same as bsg$chrM
reverseComplement(bsg$chrM)

## ---------------------------------------------------------------------
## Switch the sequence names back and forth between UCSC and NCBI
## ---------------------------------------------------------------------

## IMPORTANT NOTE: Even though hg38 is officially based on the
## GRCh38.p14 assembly (this is as of Jan 31, 2023, hg38 was based on
## GRCh38.p13 before that), it contains 2 sequences that do not belong
## to GRCh38.p14: chr11_KQ759759v1_fix and chr22_KQ759762v1_fix
## These 2 foreign sequences belong to GRCh38.p13 (they are named
## HG107_PATCH and HG1311_PATCH there), but they've been replaced with
## sequences HG107_HG2565_PATCH and HG1311_HG2539_PATCH in GRCh38.p14.

seqinfo(bsg)
seqlevelsStyle(bsg)  # UCSC

## --- switch to NCBI names ---

bsg0 <- bsg
seqlevelsStyle(bsg) <- "NCBI"
bsg
seqinfo(bsg)
bsg[["1"]]

## Surprise!
table(genome(bsg))   # 2 sequences belong to GRCh38.p13!
foreign_idx <- which(genome(bsg) == "GRCh38.p13")
seqinfo(bsg)[seqnames(bsg)[foreign_idx]]
seqlevelsStyle(bsg)  # NCBI

## --- switch back ---

seqlevelsStyle(bsg) <- "UCSC"
bsg
stopifnot(identical(bsg0, bsg))
seqinfo(bsg)[seqnames(bsg)[foreign_idx]]

## ---------------------------------------------------------------------
## Genome-wide motif searching
## ---------------------------------------------------------------------

## 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