

# Package ‘read.gb’

September 6, 2021

**Title** Open GenBank Files

**Version** 2.2

**Description** Opens complete record(s) with .gb extension from the NCBI/GenBank Nucleotide database and returns a list containing shaped record(s). These kind of files contains detailed records of DNA samples (locus, organism, type of sequence, source of the sequence...). An example of record can be found at <<https://www.ncbi.nlm.nih.gov/nuccore/HE799070>>.

**License** GPL-3

**Encoding** UTF-8

**RoxygenNote** 7.1.1

**Collate** 'CDS.F.R' 'C\_region.F.R' 'DNA.treatment.R' 'D\_loop.F.R' 'D\_segment.F.R' 'F\_UTR.F.R' 'Feature.sep.R' 'J\_segment.F.R' 'N\_region.F.R' 'Reference.sep.R' 'Reorganize.report.R' 'STS.F.R' 'S\_region.F.R' 'T\_UTR.F.R' 'V\_region.F.R' 'V\_segment.F.R' 'assembly\_gap.F.R' 'centromere.F.R' 'exon.F.R' 'extract.gb.R' 'gap.F.R' 'gene.F.R' 'iDNA.F.R' 'intron.F.R' 'mRNA.F.R' 'mat\_peptide.F.R' 'misc\_RNA.F.R' 'misc\_binding.F.R' 'misc\_difference.F.R' 'misc\_feature.F.R' 'misc\_recomb.F.R' 'misc\_structure.F.R' 'mobile\_element.F.R' 'modified\_base.F.R' 'ncRNA.F.R' 'old\_sequence.F.R' 'operon.F.R' 'oriT.F.R' 'polyA\_site.F.R' 'precursor\_RNA.F.R' 'prim\_transcript.F.R' 'primer\_bind.F.R' 'propeptide.F.R' 'protein\_bind.F.R' 'rRNA.F.R' 'read.gb.R' 'regulatory.F.R' 'rep\_origin.F.R' 'repeat\_region.F.R' 'sig\_peptide.F.R' 'source.F.R' 'stem\_loop.F.R' 'tRNA.F.R' 'telomere.F.R' 'tmRNA.F.R' 'transit\_peptide.F.R' 'unsure.F.R' 'variation.F.R'

**Imports** rentrez

**NeedsCompilation** no

**Author** Robin Mercier [aut, cre]

**Maintainer** Robin Mercier <[robin.largon.mercier@hotmail.fr](mailto:robin.largon.mercier@hotmail.fr)>

**Repository** CRAN

**Date/Publication** 2021-09-06 15:00:02 UTC

## R topics documented:

extract.gb . . . . .	2
read.gb . . . . .	3
<b>Index</b>	<b>4</b>

---

extract.gb	<i>Extracts and returns a specific item from .gb records</i>
------------	--

---

### Description

This function regroup every items found in a list if they correspond to the character string specified. Works for general parts (ACCESSION, SOURCE...) and parts of FEATURES (source, gene, CDS...)

### Usage

```
extract.gb(x, Item)
```

### Arguments

x	The name of the list containing the records.
Item	character. The part to extract. It have to match case.

### Value

Returns a list containing the data

### Examples

```
## Not run:
extract.gb(Example, "gene")

## End(Not run)
```

---

read.gb	<i>Opens files with .gb extensions</i>
---------	--

---

### Description

This function opens complete record(s) with .gb extension from the NCBI/GenBank Nucleotide database and returns a list containing shaped record(s). These kind of files contains detailed records of DNA samples (locus, organism, type of sequence, source of the sequence...). An example of record can be found at <https://www.ncbi.nlm.nih.gov/nuccore/HE799070>. Records with > 200'000 bp may experience long processing times, especially if they have numerous FEATURES items. Also works for reports obtained with rentrez package.

### Usage

```
read.gb(x, DNA = TRUE, Type = "full", Source = "File")
```

### Arguments

x	character. The name of the file which the data are to be read from, or the character string containing the data. It can contains several records
DNA	logical. If TRUE, the DNA sequence in the ORIGIN part will be merged in one character string. If FALSE, the default layout will be kept. Default if TRUE
Type	character. Should the output contain FEATURES and REFERENCE parts ? Possible values are "full" for a full record, "nofeat" to ignore FEATURES part, "noref" to ignore REFERENCE part and "nfnr" to ignore both parts. Default is "full"
Source	character. If x is a character string, use "Char". If x is a filename use "File". Default is "File"

### Value

Returns a table containing the data

### Examples

```
## Not run:
read.gb(File = "sequence.gb", DNA = TRUE, Type = "full", Source = "File")

## End(Not run)

## Not run:
require(rentrez)
data <- rentrez::entrez_fetch(db = "Nucleotide", id = "508082122", rettype = "gb")
read.gb(File = data, DNA = TRUE, Type = "full", Source = "Char")

## End(Not run)
```

# Index

`extract.gb`, 2

`read.gb`, 3