

Package ‘restez’

November 26, 2018

Type Package

Title Create and Query a Local Copy of 'GenBank' in R

Version 1.0.0

Maintainer Dom Bennett <dominic.john.bennett@gmail.com>

Description Download large sections of 'GenBank' <<https://www.ncbi.nlm.nih.gov/genbank/>> and generate a local SQL-based database. A user can then query this database using 'restez' functions or through 'rentrez' <<https://CRAN.R-project.org/package=rentrez>> wrappers.

URL <https://github.com/ropensci/restez#readme>

BugReports <https://github.com/ropensci/restez/issues>

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Depends R (>= 3.3.0)

Imports utils, rentrez, MonetDBLite (>= 0.6.0), DBI (>= 1.0.0), downloader, RCurl, cli, crayon, callr (>= 3.0.0), devtools

Suggests testthat, knitr, R.utils, rmarkdown

RoxygenNote 6.1.1

NeedsCompilation no

Author Dom Bennett [aut, cre] (<<https://orcid.org/0000-0003-2722-1359>>)

Repository CRAN

Date/Publication 2018-11-26 20:00:12 UTC

R topics documented:

add_rcrd_log	4
cat_line	4
char	5

check_connection	6
cleanup	6
connected	7
connection_get	8
count_db_ids	8
custom_download	9
custom_download2	10
db_create	10
db_delete	11
db_download	12
db_sqlngths_get	13
db_sqlngths_log	14
demo_db_create	14
dir_size	15
dwncld_path_get	16
dwncld_rcrd_log	16
entrez_fasta_get	17
entrez_fetch	18
entrez_gb_get	19
extract_accession	20
extract_by_patterns	21
extract_clean_sequence	22
extract_definition	23
extract_features	23
extract_inforecpart	24
extract_keywords	25
extract_locus	26
extract_organism	27
extract_seqrecpart	27
extract_sequence	28
extract_version	29
filename_log	30
file_download	30
flatfile_read	31
gbrelease_check	32
gbrelease_get	32
gbrelease_log	33
gb_build	34
gb_build2	35
gb_definition_get	36
gb_df_create	36
gb_df_generate	37
gb_extract	38
gb_fasta_get	39
gb_organism_get	40
gb_record_get	41
gb_sequence_get	41
gb_sql_add	42

gb_sql_query	43
gb_version_get	44
has_data	44
identify_downloadable_files	45
is_in_db	46
last_add_get	47
last_dwnld_get	47
last_entry_get	48
latest_genbank_release	49
latest_genbank_release_notes	50
list_db_ids	50
message_missing	51
mock_def	52
mock_gb_df_generate	53
mock_org	53
mock_rec	54
mock_seq	55
predict_datasizes	56
print.status	56
quiet_connect	57
readme_log	58
record	58
restez	59
restez_connect	60
restez_disconnect	60
restez_path_check	61
restez_path_get	62
restez_path_set	62
restez_path_unset	63
restez_ready	63
restez_rl	64
restez_status	65
seshinfo_log	66
setup	66
slctn_get	67
slctn_log	68
sql_path_get	68
stat	69
status_class	70
testdatadir_get	70

add_rcrd_log	<i>Log files added to the SQL database in the restez path</i>
--------------	---

Description

This function is called whenever sequence files have been successfully added to the nucleotide SQL database. Row entries are added to 'add_lot.tsv' in the user's restez path containing the filename, GB release numbers and the time of successful adding. The log is to help users keep track of when sequences have been added.

Usage

```
add_rcrd_log(f1)
```

Arguments

f1	filename, character
----	---------------------

See Also

Other private: [cat_line](#), [char](#), [check_connection](#), [cleanup](#), [connected](#), [connection_get](#), [custom_download2](#), [custom_download](#), [db_sqlngths_get](#), [db_sqlngths_log](#), [dir_size](#), [dwnld_path_get](#), [dwnld_rcrd_log](#), [entrez_fasta_get](#), [entrez_gb_get](#), [extract_accession](#), [extract_by_patterns](#), [extract_clean_sequence](#), [extract_definition](#), [extract_features](#), [extract_inforecpart](#), [extract_keywords](#), [extract_locus](#), [extract_organism](#), [extract_seqrecpart](#), [extract_sequence](#), [extract_version](#), [file_download](#), [filename_log](#), [flatfile_read](#), [gb_build2](#), [gb_build](#), [gb_df_create](#), [gb_df_generate](#), [gb_sql_add](#), [gb_sql_query](#), [gbrelease_check](#), [gbrelease_get](#), [gbrelease_log](#), [has_data](#), [identify_downloadable_files](#), [last_add_get](#), [last_dwnld_get](#), [last_entry_get](#), [latest_genbank_release_notes](#), [latest_genbank_release](#), [message_missing](#), [mock_def](#), [mock_gb_df_generate](#), [mock_org](#), [mock_rec](#), [mock_seq](#), [predict_datasizes](#), [print.status](#), [quiet_connect](#), [readme_log](#), [restez_path_check](#), [restez_rl](#), [seshinfo_log](#), [setup](#), [slctn_get](#), [slctn_log](#), [sql_path_get](#), [status_class](#), [stat](#), [testdatadir_get](#)

cat_line	<i>Cat lines</i>
----------	------------------

Description

Helper function for printing lines to console. Automatically formats lines by adding newlines.

Usage

```
cat_line(...)
```

Arguments

...	Text to print, character
-----	--------------------------

See Also

Other private: `add_rcrd_log`, `char`, `check_connection`, `cleanup`, `connected`, `connection_get`, `custom_download2`, `custom_download`, `db_sqlngths_get`, `db_sqlngths_log`, `dir_size`, `dnwld_path_get`, `dnwld_rcrd_log`, `entrez_fasta_get`, `entrez_gb_get`, `extract_accession`, `extract_by_patterns`, `extract_clean_sequence`, `extract_definition`, `extract_features`, `extract_inforecpart`, `extract_keywords`, `extract_locus`, `extract_organism`, `extract_seqrecpart`, `extract_sequence`, `extract_version`, `file_download`, `filename_log`, `flatfile_read`, `gb_build2`, `gb_build`, `gb_df_create`, `gb_df_generate`, `gb_sql_add`, `gb_sql_query`, `gbrelease_check`, `gbrelease_get`, `gbrelease_log`, `has_data`, `identify_downloadable_files`, `last_add_get`, `last_dnwld_get`, `last_entry_get`, `latest_genbank_release_notes`, `latest_genbank_release`, `message_missing`, `mock_def`, `mock_gb_df_generate`, `mock_org`, `mock_rec`, `mock_seq`, `predict_datasizes`, `print.status`, `quiet_connect`, `readme_log`, `restez_path_check`, `restez_rl`, `seshinfo_log`, `setup`, `slctn_get`, `slctn_log`, `sql_path_get`, `status_class`, `stat`, `testdatadir_get`

 char

Print green

Description

Print to console green text to indicate a name/filepath/text

Usage

`char(x)`

Arguments

x Text to print, character

Value

coloured character encoding, character

See Also

Other private: `add_rcrd_log`, `cat_line`, `check_connection`, `cleanup`, `connected`, `connection_get`, `custom_download2`, `custom_download`, `db_sqlngths_get`, `db_sqlngths_log`, `dir_size`, `dnwld_path_get`, `dnwld_rcrd_log`, `entrez_fasta_get`, `entrez_gb_get`, `extract_accession`, `extract_by_patterns`, `extract_clean_sequence`, `extract_definition`, `extract_features`, `extract_inforecpart`, `extract_keywords`, `extract_locus`, `extract_organism`, `extract_seqrecpart`, `extract_sequence`, `extract_version`, `file_download`, `filename_log`, `flatfile_read`, `gb_build2`, `gb_build`, `gb_df_create`, `gb_df_generate`, `gb_sql_add`, `gb_sql_query`, `gbrelease_check`, `gbrelease_get`, `gbrelease_log`, `has_data`, `identify_downloadable_files`, `last_add_get`, `last_dnwld_get`, `last_entry_get`, `latest_genbank_release_notes`, `latest_genbank_release`, `message_missing`, `mock_def`, `mock_gb_df_generate`, `mock_org`, `mock_rec`, `mock_seq`, `predict_datasizes`, `print.status`, `quiet_connect`, `readme_log`, `restez_path_check`, `restez_rl`, `seshinfo_log`, `setup`, `slctn_get`, `slctn_log`, `sql_path_get`, `status_class`, `stat`, `testdatadir_get`

check_connection	<i>Helper function to test if a stable internet connection can be established.</i>
------------------	--

Description

All retrieval functions need a stable internet connection to work properly. This internal function pings the google homepage and throws an error if it cannot be reached.

Usage

```
check_connection()
```

Author(s)

Hajk-Georg Drost

See Also

Other private: [add_rcrd_log](#), [cat_line](#), [char](#), [cleanup](#), [connected](#), [connection_get](#), [custom_download2](#), [custom_download](#), [db_sqlngths_get](#), [db_sqlngths_log](#), [dir_size](#), [dwnld_path_get](#), [dwnld_rcrd_log](#), [entrez_fasta_get](#), [entrez_gb_get](#), [extract_accession](#), [extract_by_patterns](#), [extract_clean_sequence](#), [extract_definition](#), [extract_features](#), [extract_inforecpart](#), [extract_keywords](#), [extract_locus](#), [extract_organism](#), [extract_seqrecpart](#), [extract_sequence](#), [extract_version](#), [file_download](#), [filename_log](#), [flatfile_read](#), [gb_build2](#), [gb_build](#), [gb_df_create](#), [gb_df_generate](#), [gb_sql_add](#), [gb_sql_query](#), [gbrelease_check](#), [gbrelease_get](#), [gbrelease_log](#), [has_data](#), [identify_downloadable_files](#), [last_add_get](#), [last_dwnld_get](#), [last_entry_get](#), [latest_genbank_release_notes](#), [latest_genbank_release](#), [message_missing](#), [mock_def](#), [mock_gb_df_generate](#), [mock_org](#), [mock_rec](#), [mock_seq](#), [predict_datasizes](#), [print.status](#), [quiet_connect](#), [readme_log](#), [restez_path_check](#), [restez_rl](#), [seshinfo_log](#), [setup](#), [slctn_get](#), [slctn_log](#), [sql_path_get](#), [status_class](#), [stat](#), [testdatadir_get](#)

cleanup	<i>Clean up test data</i>
---------	---------------------------

Description

Removes all temporary test data created.

Usage

```
cleanup()
```

See Also

Other private: `add_rcrd_log`, `cat_line`, `char`, `check_connection`, `connected`, `connection_get`, `custom_download2`, `custom_download`, `db_sqlngths_get`, `db_sqlngths_log`, `dir_size`, `dwncld_path_get`, `dwncld_rcrd_log`, `entrez_fasta_get`, `entrez_gb_get`, `extract_accession`, `extract_by_patterns`, `extract_clean_sequence`, `extract_definition`, `extract_features`, `extract_inforecpart`, `extract_keywords`, `extract_locus`, `extract_organism`, `extract_seqrecpart`, `extract_sequence`, `extract_version`, `file_download`, `filename_log`, `flatfile_read`, `gb_build2`, `gb_build`, `gb_df_create`, `gb_df_generate`, `gb_sql_add`, `gb_sql_query`, `gbrelease_check`, `gbrelease_get`, `gbrelease_log`, `has_data`, `identify_downloadable_files`, `last_add_get`, `last_dwncld_get`, `last_entry_get`, `latest_genbank_release_notes`, `latest_genbank_release`, `message_missing`, `mock_def`, `mock_gb_df_generate`, `mock_org`, `mock_rec`, `mock_seq`, `predict_datasizes`, `print.status`, `quiet_connect`, `readme_log`, `restez_path_check`, `restez_rl`, `seshinfo_log`, `setup`, `slctn_get`, `slctn_log`, `sql_path_get`, `status_class`, `stat`, `testdatadir_get`

connected

Is restez connected?

Description

Returns TRUE if a restez SQL database has been connected.

Usage

`connected()`

Value

Logical

See Also

Other private: `add_rcrd_log`, `cat_line`, `char`, `check_connection`, `cleanup`, `connection_get`, `custom_download2`, `custom_download`, `db_sqlngths_get`, `db_sqlngths_log`, `dir_size`, `dwncld_path_get`, `dwncld_rcrd_log`, `entrez_fasta_get`, `entrez_gb_get`, `extract_accession`, `extract_by_patterns`, `extract_clean_sequence`, `extract_definition`, `extract_features`, `extract_inforecpart`, `extract_keywords`, `extract_locus`, `extract_organism`, `extract_seqrecpart`, `extract_sequence`, `extract_version`, `file_download`, `filename_log`, `flatfile_read`, `gb_build2`, `gb_build`, `gb_df_create`, `gb_df_generate`, `gb_sql_add`, `gb_sql_query`, `gbrelease_check`, `gbrelease_get`, `gbrelease_log`, `has_data`, `identify_downloadable_files`, `last_add_get`, `last_dwncld_get`, `last_entry_get`, `latest_genbank_release_notes`, `latest_genbank_release`, `message_missing`, `mock_def`, `mock_gb_df_generate`, `mock_org`, `mock_rec`, `mock_seq`, `predict_datasizes`, `print.status`, `quiet_connect`, `readme_log`, `restez_path_check`, `restez_rl`, `seshinfo_log`, `setup`, `slctn_get`, `slctn_log`, `sql_path_get`, `status_class`, `stat`, `testdatadir_get`

connection_get	<i>Retrieve restez connection</i>
----------------	-----------------------------------

Description

Safely acquire the restez connection. Raises error if no connection set.

Usage

```
connection_get()
```

Value

connection

See Also

Other private: `add_rcrd_log`, `cat_line`, `char`, `check_connection`, `cleanup`, `connected`, `custom_download2`, `custom_download`, `db_sqlngths_get`, `db_sqlngths_log`, `dir_size`, `dwncld_path_get`, `dwncld_rcrd_log`, `entrez_fasta_get`, `entrez_gb_get`, `extract_accession`, `extract_by_patterns`, `extract_clean_sequence`, `extract_definition`, `extract_features`, `extract_infocpart`, `extract_keywords`, `extract_locus`, `extract_organism`, `extract_seqrcpart`, `extract_sequence`, `extract_version`, `file_download`, `filename_log`, `flatfile_read`, `gb_build2`, `gb_build`, `gb_df_create`, `gb_df_generate`, `gb_sql_add`, `gb_sql_query`, `gbrelease_check`, `gbrelease_get`, `gbrelease_log`, `has_data`, `identify_downloadable_files`, `last_add_get`, `last_dwncld_get`, `last_entry_get`, `latest_genbank_release_notes`, `latest_genbank_release`, `message_missing`, `mock_def`, `mock_gb_df_generate`, `mock_org`, `mock_rec`, `mock_seq`, `predict_datasizes`, `print.status`, `quiet_connect`, `readme_log`, `restez_path_check`, `restez_rl`, `seshinfo_log`, `setup`, `slctn_get`, `slctn_log`, `sql_path_get`, `status_class`, `stat`, `testdatadir_get`

count_db_ids	<i>Return the number of ids</i>
--------------	---------------------------------

Description

Return the number of ids in a user's restez database.

Usage

```
count_db_ids(db = "nucleotide")
```

Arguments

db	character, database name
----	--------------------------

Details

Requires an open connection. If no connection or db 0 is returned.

Value

integer

See Also

Other database: [db_create](#), [db_delete](#), [db_download](#), [demo_db_create](#), [is_in_db](#), [list_db_ids](#)

Examples

```
library(restez)
restez_path_set(filepath = tempdir())
demo_db_create(n = 5)
restez_connect()
(count_db_ids())

# delete demo after example
db_delete(everything = TRUE)
```

custom_download

Helper function to perform customized downloads

Description

To achieve the most stable download experience, ftp file downloads are customized for each operating system.

Usage

```
custom_download(...)
```

Arguments

... additional arguments that shall be passed to [download](#)

Author(s)

Hajk-Georg Drost

See Also

Other private: [add_rcrd_log](#), [cat_line](#), [char](#), [check_connection](#), [cleanup](#), [connected](#), [connection_get](#), [custom_download2](#), [db_sqlngths_get](#), [db_sqlngths_log](#), [dir_size](#), [dwnld_path_get](#), [dwnld_rcrd_log](#), [entrez_fasta_get](#), [entrez_gb_get](#), [extract_accession](#), [extract_by_patterns](#), [extract_clean_sequence](#), [extract_definition](#), [extract_features](#), [extract_infocpart](#), [extract_keywords](#), [extract_locus](#), [extract_organism](#), [extract_seqrecpart](#), [extract_sequence](#), [extract_version](#), [file_download](#), [filename_log](#), [flatfile_read](#), [gb_build2](#), [gb_build](#), [gb_df_create](#), [gb_df_generate](#), [gb_sql_add](#), [gb_sql_query](#), [gbrelease_check](#), [gbrelease_get](#), [gbrelease_log](#), [has_data](#), [identify_downloadable_files](#), [last_add_get](#), [last_dwnld_get](#), [last_entry_get](#), [latest_genbank_release_notes](#), [latest_genbank_release](#),

[message_missing](#), [mock_def](#), [mock_gb_df_generate](#), [mock_org](#), [mock_rec](#), [mock_seq](#), [predict_datasizes](#), [print.status](#), [quiet_connect](#), [readme_log](#), [restez_path_check](#), [restez_rl](#), [seshinfo_log](#), [setup](#), [slctn_get](#), [slctn_log](#), [sql_path_get](#), [status_class](#), [stat](#), [testdatadir_get](#)

custom_download2 *Callr version of custom_download()*

Description

Runs [custom_download](#) as an independent R process. This allows the user to kill the process. Additionally, the process will print spinning dots to indicate it is still active.

Usage

```
custom_download2(url, destfile)
```

Arguments

url URL of source file, character.
destfile filepath to where the file should be saved.

See Also

Other private: [add_rcrd_log](#), [cat_line](#), [char](#), [check_connection](#), [cleanup](#), [connected](#), [connection_get](#), [custom_download](#), [db_sqlngths_get](#), [db_sqlngths_log](#), [dir_size](#), [dwnld_path_get](#), [dwnld_rcrd_log](#), [entrez_fasta_get](#), [entrez_gb_get](#), [extract_accession](#), [extract_by_patterns](#), [extract_clean_sequence](#), [extract_definition](#), [extract_features](#), [extract_infocpart](#), [extract_keywords](#), [extract_locus](#), [extract_organism](#), [extract_seqrepart](#), [extract_sequence](#), [extract_version](#), [file_download](#), [filename_log](#), [flatfile_read](#), [gb_build2](#), [gb_build](#), [gb_df_create](#), [gb_df_generate](#), [gb_sql_add](#), [gb_sql_query](#), [gbrelease_check](#), [gbrelease_get](#), [gbrelease_log](#), [has_data](#), [identify_downloadable_files](#), [last_add_get](#), [last_dwnld_get](#), [last_entry_get](#), [latest_genbank_release_notes](#), [latest_genbank_release](#), [message_missing](#), [mock_def](#), [mock_gb_df_generate](#), [mock_org](#), [mock_rec](#), [mock_seq](#), [predict_datasizes](#), [print.status](#), [quiet_connect](#), [readme_log](#), [restez_path_check](#), [restez_rl](#), [seshinfo_log](#), [setup](#), [slctn_get](#), [slctn_log](#), [sql_path_get](#), [status_class](#), [stat](#), [testdatadir_get](#)

db_create *Create new NCBI database*

Description

Create a new local SQL database from downloaded files. Currently only GenBank/nucleotide/nucleotide database is supported.

Usage

```
db_create(db_type = "nucleotide", min_length = 0, max_length = NULL,
          alt_restez_path = NULL)
```

Arguments

db_type	character, database type
min_length	Minimum sequence length, default 0.
max_length	Maximum sequence length, default NULL.
alt_restez_path	Alternative restez path if you would like to use the downloads from a different restez path.

Details

All .seq.gz files are added to the database. A user can specify sequence limit sizes for those sequences to be added to the database – smaller databases are faster to search.

Alternatively, a user can use the `alt_restez_path` to add the files from an alternative restez file path. For example, you may wish to have a database of all environmental sequences but then an additional smaller one of just the sequences with lengths below 100 bp. Instead of having to download all environmental sequences twice, you can generate multiple restez databases using the same downloaded files from a single restez path.

This function will not overwrite a pre-existing database. Old databases must be deleted before a new one can be created. Use `db_delete` with `everything=FALSE` to delete an SQL database.

Connections/disconnections to the database are made automatically.

See Also

Other database: [count_db_ids](#), [db_delete](#), [db_download](#), [demo_db_create](#), [is_in_db](#), [list_db_ids](#)

Examples

```
## Not run:
library(restez)
restez_path_set(filepath = 'path/for/downloads/and/database')
db_download()
db_create()

## End(Not run)
```

db_delete

Delete database

Description

Delete the local SQL database and/or restez folder.

Usage

```
db_delete(everything = FALSE)
```

Arguments

everything T/F, delete the whole restez folder as well?

Details

Any connected database will be automatically disconnected.

See Also

Other database: [count_db_ids](#), [db_create](#), [db_download](#), [demo_db_create](#), [is_in_db](#), [list_db_ids](#)

Examples

```
library(restez)
fp <- tempdir()
restez_path_set(filepath = fp)
demo_db_create(n = 10)
db_delete(everything = FALSE)
# Will not run: gb_sequence_get(id = 'demo_1')
# only the SQL database is deleted
db_delete(everything = TRUE)
# Now returns NULL
(restez_path_get())
```

db_download

Download database

Description

Download .seq.tar files from the latest GenBank release. The user interactively selects the parts of GenBank to download (e.g. primates, plants, bacteria ...)

Usage

```
db_download(db = "nucleotide", overwrite = FALSE,
  preselection = NULL)
```

Arguments

db Database type, only 'nucleotide' currently available.
 overwrite T/F, overwrite pre-existing downloaded files?
 preselection character of user input

Details

The downloaded files will appear in the restez filepath under downloads.

Value

T/F, if all files download correctly, TRUE else FALSE.

See Also

Other database: [count_db_ids](#), [db_create](#), [db_delete](#), [demo_db_create](#), [is_in_db](#), [list_db_ids](#)

Examples

```
## Not run:
library(restez)
restez_path_set(filepath = 'path/for/downloads')
db_download()

## End(Not run)
```

 db_sqlngths_get

Return the minimum and maximum sequence lengths in db

Description

Returns the maximum and minimum sequence lengths as set by the user upon db creation.

Usage

```
db_sqlngths_get()
```

Details

If no file found, returns empty character vector.

Value

vector of integers

See Also

Other private: [add_rcrd_log](#), [cat_line](#), [char](#), [check_connection](#), [cleanup](#), [connected](#), [connection_get](#), [custom_download2](#), [custom_download](#), [db_sqlngths_log](#), [dir_size](#), [dwnld_path_get](#), [dwnld_rcrd_log](#), [entrez_fasta_get](#), [entrez_gb_get](#), [extract_accession](#), [extract_by_patterns](#), [extract_clean_sequence](#), [extract_definition](#), [extract_features](#), [extract_infopart](#), [extract_keywords](#), [extract_locus](#), [extract_organism](#), [extract_seqpart](#), [extract_sequence](#), [extract_version](#), [file_download](#), [filename_log](#), [flatfile_read](#), [gb_build2](#), [gb_build](#), [gb_df_create](#), [gb_df_generate](#), [gb_sql_add](#), [gb_sql_query](#), [gbrelease_check](#), [gbrelease_get](#), [gbrelease_log](#), [has_data](#), [identify_downloadable_files](#), [last_add_get](#), [last_dwnld_get](#), [last_entry_get](#), [latest_genbank_release_notes](#), [latest_genbank_release](#), [message_missing](#), [mock_def](#), [mock_gb_df_generate](#), [mock_org](#), [mock_rec](#), [mock_seq](#), [predict_datasizes](#), [print.status](#), [quiet_connect](#), [readme_log](#), [restez_path_check](#), [restez_rl](#), [seshinfo_log](#), [setup](#), [slctn_get](#), [slctn_log](#), [sql_path_get](#), [status_class](#), [stat](#), [testdatadir_get](#)

db_sqlngths_log *Log the min and max sequence lengths*

Description

Log the min and maximum sequence length used in the created db.

Usage

```
db_sqlngths_log(min_lngth, max_lngth)
```

Arguments

min_lngth	Minimum length
max_lngth	Maximum length

See Also

Other private: [add_rcrd_log](#), [cat_line](#), [char](#), [check_connection](#), [cleanup](#), [connected](#), [connection_get](#), [custom_download2](#), [custom_download](#), [db_sqlngths_get](#), [dir_size](#), [dwnld_path_get](#), [dwnld_rcrd_log](#), [entrez_fasta_get](#), [entrez_gb_get](#), [extract_accession](#), [extract_by_patterns](#), [extract_clean_sequence](#), [extract_definition](#), [extract_features](#), [extract_inforecpart](#), [extract_keywords](#), [extract_locus](#), [extract_organism](#), [extract_seqrecpart](#), [extract_sequence](#), [extract_version](#), [file_download](#), [filename_log](#), [flatfile_read](#), [gb_build2](#), [gb_build](#), [gb_df_create](#), [gb_df_generate](#), [gb_sql_add](#), [gb_sql_query](#), [gbrelease_check](#), [gbrelease_get](#), [gbrelease_log](#), [has_data](#), [identify_downloadable_files](#), [last_add_get](#), [last_dwnld_get](#), [last_entry_get](#), [latest_genbank_release_notes](#), [latest_genbank_release](#), [message_missing](#), [mock_def](#), [mock_gb_df_generate](#), [mock_org](#), [mock_rec](#), [mock_seq](#), [predict_datasizes](#), [print.status](#), [quiet_connect](#), [readme_log](#), [restez_path_check](#), [restez_rl](#), [seshinfo_log](#), [setup](#), [slctn_get](#), [slctn_log](#), [sql_path_get](#), [status_class](#), [stat](#), [testdatadir_get](#)

demo_db_create *Create demo database*

Description

Creates a local mock SQL database from package test data for demonstration purposes. No internet connection required.

Usage

```
demo_db_create(db_type = "nucleotide", n = 100)
```

Arguments

db_type	character, database type
n	integer, number of mock sequences

See Also

Other database: [count_db_ids](#), [db_create](#), [db_delete](#), [db_download](#), [is_in_db](#), [list_db_ids](#)

Examples

```
library(restez)
# set the restez path to a temporary dir
restez_path_set(filepath = tempdir())
# create demo database
demo_db_create(n = 5)
restez_connect()
# in the demo, IDs are 'demo_1', 'demo_2' ...
(gb_sequence_get(id = 'demo_1'))

# Delete a demo database after an example
db_delete(everything = TRUE)
```

dir_size

Calculate the size of a directory

Description

Returns the size of directory in GB

Usage

```
dir_size(fp)
```

Arguments

fp File path, character

Value

numeric

See Also

Other private: [add_rcrd_log](#), [cat_line](#), [char](#), [check_connection](#), [cleanup](#), [connected](#), [connection_get](#), [custom_download2](#), [custom_download](#), [db_sqlngths_get](#), [db_sqlngths_log](#), [dwnld_path_get](#), [dwnld_rcrd_log](#), [entrez_fasta_get](#), [entrez_gb_get](#), [extract_accession](#), [extract_by_patterns](#), [extract_clean_sequence](#), [extract_definition](#), [extract_features](#), [extract_inforecpart](#), [extract_keywords](#), [extract_locus](#), [extract_organism](#), [extract_seqrecpart](#), [extract_sequence](#), [extract_version](#), [file_download](#), [filename_log](#), [flatfile_read](#), [gb_build2](#), [gb_build](#), [gb_df_create](#), [gb_df_generate](#), [gb_sql_add](#), [gb_sql_query](#), [gbrelease_check](#), [gbrelease_get](#), [gbrelease_log](#), [has_data](#), [identify_downloadable_files](#), [last_add_get](#), [last_dwnld_get](#), [last_entry_get](#), [latest_genbank_release_notes](#), [latest_genbank_release](#), [message_missing](#), [mock_def](#), [mock_gb_df_generate](#), [mock_org](#), [mock_rec](#), [mock_seq](#), [predict_datasizes](#), [print.status](#), [quiet_connect](#), [readme_log](#),

restez_path_check, restez_rl, seshinfo_log, setup, slctn_get, slctn_log, sql_path_get, status_class, stat, testdatadir_get

dwnld_path_get *Get dwnld path*

Description

Return path to folder where raw .seq files are stored.

Usage

dwnld_path_get()

Value

character

See Also

Other private: add_rcrd_log, cat_line, char, check_connection, cleanup, connected, connection_get, custom_download2, custom_download, db_sqlngths_get, db_sqlngths_log, dir_size, dwnld_rcrd_log, entrez_fasta_get, entrez_gb_get, extract_accession, extract_by_patterns, extract_clean_sequence, extract_definition, extract_features, extract_infocpart, extract_keywords, extract_locus, extract_organism, extract_seqrepart, extract_sequence, extract_version, file_download, filename_log, flatfile_read, gb_build2, gb_build, gb_df_create, gb_df_generate, gb_sql_add, gb_sql_query, gbrelease_check, gbrelease_get, gbrelease_log, has_data, identify_downloadable_files, last_add_get, last_dwnld_get, last_entry_get, latest_genbank_release_notes, latest_genbank_release, message_missing, mock_def, mock_gb_df_generate, mock_org, mock_rec, mock_seq, predict_datasizes, print.status, quiet_connect, readme_log, restez_path_check, restez_rl, seshinfo_log, setup, slctn_get, slctn_log, sql_path_get, status_class, stat, testdatadir_get

dwnld_rcrd_log *Log a downloaded file in the restez path*

Description

This function is called whenever a file is successfully downloaded. A row entry is added to the 'download_log.tsv' in the user's restez path containing the file name, the GB release number and the time of successfully download. The log is to help users keep track of when they downloaded files and to determine if the downloaded files are out of date.

Usage

dwnld_rcrd_log(f1)

Arguments

f1 file name, character

See Also

Other private: [add_rcrd_log](#), [cat_line](#), [char](#), [check_connection](#), [cleanup](#), [connected](#), [connection_get](#), [custom_download2](#), [custom_download](#), [db_sqlngths_get](#), [db_sqlngths_log](#), [dir_size](#), [dwnld_path_get](#), [entrez_fasta_get](#), [entrez_gb_get](#), [extract_accession](#), [extract_by_patterns](#), [extract_clean_sequence](#), [extract_definition](#), [extract_features](#), [extract_inforecpart](#), [extract_keywords](#), [extract_locus](#), [extract_organism](#), [extract_seqrecpart](#), [extract_sequence](#), [extract_version](#), [file_download](#), [filename_log](#), [flatfile_read](#), [gb_build2](#), [gb_build](#), [gb_df_create](#), [gb_df_generate](#), [gb_sql_add](#), [gb_sql_query](#), [gbrelease_check](#), [gbrelease_get](#), [gbrelease_log](#), [has_data](#), [identify_downloadable_files](#), [last_add_get](#), [last_dwnld_get](#), [last_entry_get](#), [latest_genbank_release_notes](#), [latest_genbank_release](#), [message_missing](#), [mock_def](#), [mock_gb_df_generate](#), [mock_org](#), [mock_rec](#), [mock_seq](#), [predict_datasizes](#), [print.status](#), [quiet_connect](#), [readme_log](#), [restez_path_check](#), [restez_rl](#), [seshinfo_log](#), [setup](#), [slctn_get](#), [slctn_log](#), [sql_path_get](#), [status_class](#), [stat](#), [testdatadir_get](#)

entrez_fasta_get *Get Entrez fasta*

Description

Return fasta format as expected from an Entrez call. If not all IDs are returned, will run `rentrez::entrez_fetch`.

Usage

```
entrez_fasta_get(id, ...)
```

Arguments

id vector, unique ID(s) for record(s)
 ... arguments passed on to `rentrez`

Value

character string containing the file created

See Also

Other private: [add_rcrd_log](#), [cat_line](#), [char](#), [check_connection](#), [cleanup](#), [connected](#), [connection_get](#), [custom_download2](#), [custom_download](#), [db_sqlngths_get](#), [db_sqlngths_log](#), [dir_size](#), [dwnld_path_get](#), [dwnld_rcrd_log](#), [entrez_gb_get](#), [extract_accession](#), [extract_by_patterns](#), [extract_clean_sequence](#), [extract_definition](#), [extract_features](#), [extract_inforecpart](#), [extract_keywords](#), [extract_locus](#), [extract_organism](#), [extract_seqrecpart](#), [extract_sequence](#), [extract_version](#), [file_download](#), [filename_log](#), [flatfile_read](#), [gb_build2](#), [gb_build](#), [gb_df_create](#), [gb_df_generate](#), [gb_sql_add](#), [gb_sql_query](#), [gbrelease_check](#), [gbrelease_get](#), [gbrelease_log](#), [has_data](#), [identify_downloadable_files](#),

[last_add_get](#), [last_dwnld_get](#), [last_entry_get](#), [latest_genbank_release_notes](#), [latest_genbank_release](#), [message_missing](#), [mock_def](#), [mock_gb_df_generate](#), [mock_org](#), [mock_rec](#), [mock_seq](#), [predict_datasizes](#), [print.status](#), [quiet_connect](#), [readme_log](#), [restez_path_check](#), [restez_rl](#), [seshinfo_log](#), [setup](#), [slctn_get](#), [slctn_log](#), [sql_path_get](#), [status_class](#), [stat](#), [testdatadir_get](#)

entrez_fetch	<i>Entrez_fetch</i>
--------------	---------------------

Description

Wrapper for `rentrez::entrez_fetch`.

Usage

```
entrez_fetch(db, id = NULL, rettype, retmode = "", ...)
```

Arguments

<code>db</code>	character, name of the database
<code>id</code>	vector, unique ID(s) for record(s)
<code>rettype</code>	character, data format
<code>retmode</code>	character, data mode
<code>...</code>	Arguments to be passed on to <code>rentrez</code>

Details

Attempts to first search local database with user-specified parameters, if the record is missing in the database, the function then calls `rentrez::entrez_fetch` to search GenBank remotely.

`rettype='fasta'` and `rettype='gb'` are respectively equivalent to [gb_fasta_get](#) and [gb_record_get](#).

Value

character string containing the file created

Supported return types and modes

XML `retmode` is not supported. Rettypes `'seqid'`, `'ft'`, `'acc'` and `'uilst'` are also not supported.

Note

It is advisable to call `restez` and `rentrez` functions with `'::'` notation rather than `library()` calls to avoid namespace issues. e.g. `restez::entrez_fetch()`.

See Also

[entrez_fetch](#)

Examples

```

library(restez)
rester_path_set(tempdir())
demo_db_create(n = 5)
rester_connect()
# return fasta record
fasta_res <- entrez_fetch(db = 'nucleotide',
                        id = c('demo_1', 'demo_2'),
                        rettype = 'fasta')

cat(fasta_res)
# return whole GB record in text format
gb_res <- entrez_fetch(db = 'nucleotide',
                    id = c('demo_1', 'demo_2'),
                    rettype = 'gb')

cat(gb_res)
# NOT RUN
# whereas these request would go through rentrez
# fasta_res <- entrez_fetch(db = 'nucleotide',
#                          id = c('S71333', 'S71334'),
#                          rettype = 'fasta')
# gb_res <- entrez_fetch(db = 'nucleotide',
#                       id = c('S71333', 'S71334'),
#                       rettype = 'gb')

# delete demo after example
db_delete(everything = TRUE)

```

entrez_gb_get

Get Entrez GenBank record

Description

Return gb and gbwithparts format as expected from an Entrez call. If not all IDs are returned, will run `rentrez::entrez_fetch`.

Usage

```
entrez_gb_get(id, ...)
```

Arguments

<code>id</code>	vector, unique ID(s) for record(s)
<code>...</code>	arguments passed on to <code>rentrez</code>

Value

character string containing the file created

See Also

Other private: add_rcrd_log, cat_line, char, check_connection, cleanup, connected, connection_get, custom_download2, custom_download, db_sqlngths_get, db_sqlngths_log, dir_size, dwnld_path_get, dwnld_rcrd_log, entrez_fasta_get, extract_accession, extract_by_patterns, extract_clean_sequence, extract_definition, extract_features, extract_infocpart, extract_keywords, extract_locus, extract_organism, extract_seqrecpart, extract_sequence, extract_version, file_download, filename_log, flatfile_read, gb_build2, gb_build, gb_df_create, gb_df_generate, gb_sql_add, gb_sql_query, gbrelease_check, gbrelease_get, gbrelease_log, has_data, identify_downloadable_files, last_add_get, last_dwnld_get, last_entry_get, latest_genbank_release_notes, latest_genbank_release, message_missing, mock_def, mock_gb_df_generate, mock_org, mock_rec, mock_seq, predict_datasizes, print.status, quiet_connect, readme_log, restez_path_check, restez_rl, seshinfo_log, setup, slctn_get, slctn_log, sql_path_get, status_class, stat, testdatadir_get

extract_accession	<i>Extract accession</i>
-------------------	--------------------------

Description

Return accession ID from GenBank record

Usage

```
extract_accession(record)
```

Arguments

record	GenBank record in text format, character
--------	--

Details

If element is not found, "" returned.

Value

character

See Also

Other private: add_rcrd_log, cat_line, char, check_connection, cleanup, connected, connection_get, custom_download2, custom_download, db_sqlngths_get, db_sqlngths_log, dir_size, dwnld_path_get, dwnld_rcrd_log, entrez_fasta_get, entrez_gb_get, extract_by_patterns, extract_clean_sequence, extract_definition, extract_features, extract_infocpart, extract_keywords, extract_locus, extract_organism, extract_seqrecpart, extract_sequence, extract_version, file_download, filename_log, flatfile_read, gb_build2, gb_build, gb_df_create, gb_df_generate, gb_sql_add, gb_sql_query, gbrelease_check, gbrelease_get, gbrelease_log, has_data, identify_downloadable_files, last_add_get, last_dwnld_get, last_entry_get, latest_genbank_release_notes, latest_genbank_release, message_missing, mock_def, mock_gb_df_generate, mock_org, mock_rec, mock_seq, predict_datasizes, print.status, quiet_connect, readme_log, restez_path_check, restez_rl, seshinfo_log, setup, slctn_get, slctn_log, sql_path_get, status_class, stat, testdatadir_get

extract_by_patterns *Extract by keyword*

Description

Search through GenBank record for a keyword and return text up to the end_pattern.

Usage

```
extract_by_patterns(record, start_pattern, end_pattern = "\n")
```

Arguments

record	GenBank record in text format, character
start_pattern	REGEX pattern indicating the point to start extraction, character
end_pattern	REGEX pattern indicating the point to stop extraction, character

Details

The start_pattern should be any of the capitalized elements in a GenBank record (e.g. LOCUS, DESCRIPTION, ACCESSION). The end_pattern depends on how much of the selected element a user wants returned. By default, the extraction will stop at the next newline. If keyword or end pattern not found, returns NULL.

Value

character or NULL

See Also

Other private: [add_rcrd_log](#), [cat_line](#), [char](#), [check_connection](#), [cleanup](#), [connected](#), [connection_get](#), [custom_download2](#), [custom_download](#), [db_sqlngths_get](#), [db_sqlngths_log](#), [dir_size](#), [dwnld_path_get](#), [dwnld_rcrd_log](#), [entrez_fasta_get](#), [entrez_gb_get](#), [extract_accession](#), [extract_clean_sequence](#), [extract_definition](#), [extract_features](#), [extract_infocpart](#), [extract_keywords](#), [extract_locus](#), [extract_organism](#), [extract_seqrecpart](#), [extract_sequence](#), [extract_version](#), [file_download](#), [filename_log](#), [flatfile_read](#), [gb_build2](#), [gb_build](#), [gb_df_create](#), [gb_df_generate](#), [gb_sql_add](#), [gb_sql_query](#), [gbrelease_check](#), [gbrelease_get](#), [gbrelease_log](#), [has_data](#), [identify_downloadable_files](#), [last_add_get](#), [last_dwnld_get](#), [last_entry_get](#), [latest_genbank_release_notes](#), [latest_genbank_release](#), [message_missing](#), [mock_def](#), [mock_gb_df_generate](#), [mock_org](#), [mock_rec](#), [mock_seq](#), [predict_datasizes](#), [print.status](#), [quiet_connect](#), [readme_log](#), [restez_path_check](#), [restez_rl](#), [seshinfo_log](#), [setup](#), [slctn_get](#), [slctn_log](#), [sql_path_get](#), [status_class](#), [stat](#), [testdatadir_get](#)

extract_clean_sequence

Extract clean sequence from sequence part

Description

Return clean sequence from seqrecpart of a GenBank record

Usage

extract_clean_sequence(seqrecpart)

Arguments

seqrecpart Sequence part of a GenBank record, character

Details

If element is not found, "" returned.

Value

character

See Also

Other private: add_rcrd_log, cat_line, char, check_connection, cleanup, connected, connection_get, custom_download2, custom_download, db_sqlngths_get, db_sqlngths_log, dir_size, dwnld_path_get, dwnld_rcrd_log, entrez_fasta_get, entrez_gb_get, extract_accession, extract_by_patterns, extract_definition, extract_features, extract_infocpart, extract_keywords, extract_locus, extract_organism, extract_seqrecpart, extract_sequence, extract_version, file_download, filename_log, flatfile_read, gb_build2, gb_build, gb_df_create, gb_df_generate, gb_sql_add, gb_sql_query, gbrelease_check, gbrelease_get, gbrelease_log, has_data, identify_downloadable_files, last_add_get, last_dwnld_get, last_entry_get, latest_genbank_release_notes, latest_genbank_release, message_missing, mock_def, mock_gb_df_generate, mock_org, mock_rec, mock_seq, predict_datasizes, print.status, quiet_connect, readme_log, restez_path_check, restez_rl, seshinfo_log, setup, slctn_get, slctn_log, sql_path_get, status_class, stat, testdatadir_get

extract_definition	<i>Extract definition</i>
--------------------	---------------------------

Description

Return definition from GenBank record.

Usage

```
extract_definition(record)
```

Arguments

record	GenBank record in text format, character
--------	--

Details

If element is not found, "" returned.

Value

character

See Also

Other private: [add_rcrd_log](#), [cat_line](#), [char](#), [check_connection](#), [cleanup](#), [connected](#), [connection_get](#), [custom_download2](#), [custom_download](#), [db_sqlngths_get](#), [db_sqlngths_log](#), [dir_size](#), [dwnld_path_get](#), [dwnld_rcrd_log](#), [entrez_fasta_get](#), [entrez_gb_get](#), [extract_accession](#), [extract_by_patterns](#), [extract_clean_sequence](#), [extract_features](#), [extract_inforecpart](#), [extract_keywords](#), [extract_locus](#), [extract_organism](#), [extract_seqrecpart](#), [extract_sequence](#), [extract_version](#), [file_download](#), [filename_log](#), [flatfile_read](#), [gb_build2](#), [gb_build](#), [gb_df_create](#), [gb_df_generate](#), [gb_sql_add](#), [gb_sql_query](#), [gbrelease_check](#), [gbrelease_get](#), [gbrelease_log](#), [has_data](#), [identify_downloadable_files](#), [last_add_get](#), [last_dwnld_get](#), [last_entry_get](#), [latest_genbank_release_notes](#), [latest_genbank_release](#), [message_missing](#), [mock_def](#), [mock_gb_df_generate](#), [mock_org](#), [mock_rec](#), [mock_seq](#), [predict_datasizes](#), [print.status](#), [quiet_connect](#), [readme_log](#), [restez_path_check](#), [restez_rl](#), [seshinfo_log](#), [setup](#), [slctn_get](#), [slctn_log](#), [sql_path_get](#), [status_class](#), [stat](#), [testdatadir_get](#)

extract_features	<i>Extract features</i>
------------------	-------------------------

Description

Return feature table as list from GenBank record

Usage

```
extract_features(record)
```

Arguments

record GenBank record in text format, character

Details

If element is not found, empty list returned.

Value

list of lists

See Also

Other private: [add_rcrd_log](#), [cat_line](#), [char](#), [check_connection](#), [cleanup](#), [connected](#), [connection_get](#), [custom_download2](#), [custom_download](#), [db_sqlngths_get](#), [db_sqlngths_log](#), [dir_size](#), [dwnld_path_get](#), [dwnld_rcrd_log](#), [entrez_fasta_get](#), [entrez_gb_get](#), [extract_accession](#), [extract_by_patterns](#), [extract_clean_sequence](#), [extract_definition](#), [extract_inforecpart](#), [extract_keywords](#), [extract_locus](#), [extract_organism](#), [extract_seqrecpart](#), [extract_sequence](#), [extract_version](#), [file_download](#), [filename_log](#), [flatfile_read](#), [gb_build2](#), [gb_build](#), [gb_df_create](#), [gb_df_generate](#), [gb_sql_add](#), [gb_sql_query](#), [gbrelease_check](#), [gbrelease_get](#), [gbrelease_log](#), [has_data](#), [identify_downloadable_f](#), [last_add_get](#), [last_dwnld_get](#), [last_entry_get](#), [latest_genbank_release_notes](#), [latest_genbank_release](#), [message_missing](#), [mock_def](#), [mock_gb_df_generate](#), [mock_org](#), [mock_rec](#), [mock_seq](#), [predict_datasizes](#), [print.status](#), [quiet_connect](#), [readme_log](#), [restez_path_check](#), [restez_rl](#), [seshinfo_log](#), [setup](#), [slctn_get](#), [slctn_log](#), [sql_path_get](#), [status_class](#), [stat](#), [testdatadir_get](#)

extract_inforecpart *Extract the information record part*

Description

Return information part from GenBank record

Usage

```
extract_inforecpart(record)
```

Arguments

record GenBank record in text format, character

Details

If element is not found, "" returned.

Value

character

See Also

Other private: `add_rcrd_log`, `cat_line`, `char`, `check_connection`, `cleanup`, `connected`, `connection_get`, `custom_download2`, `custom_download`, `db_sqlngths_get`, `db_sqlngths_log`, `dir_size`, `dwlnld_path_get`, `dwlnld_rcrd_log`, `entrez_fasta_get`, `entrez_gb_get`, `extract_accession`, `extract_by_patterns`, `extract_clean_sequence`, `extract_definition`, `extract_features`, `extract_keywords`, `extract_locus`, `extract_organism`, `extract_seqrepart`, `extract_sequence`, `extract_version`, `file_download`, `filename_log`, `flatfile_read`, `gb_build2`, `gb_build`, `gb_df_create`, `gb_df_generate`, `gb_sql_add`, `gb_sql_query`, `gbrelease_check`, `gbrelease_get`, `gbrelease_log`, `has_data`, `identify_downloadable_files`, `last_add_get`, `last_dwlnld_get`, `last_entry_get`, `latest_genbank_release_notes`, `latest_genbank_release`, `message_missing`, `mock_def`, `mock_gb_df_generate`, `mock_org`, `mock_rec`, `mock_seq`, `predict_datasizes`, `print.status`, `quiet_connect`, `readme_log`, `restez_path_check`, `restez_rl`, `seshinfo_log`, `setup`, `slctn_get`, `slctn_log`, `sql_path_get`, `status_class`, `stat`, `testdatadir_get`

extract_keywords	<i>Extract keywords</i>
------------------	-------------------------

Description

Return keywords as list from GenBank record

Usage

```
extract_keywords(record)
```

Arguments

record	GenBank record in text format, character
--------	--

Details

If element is not found, "" returned.

Value

character vector

See Also

Other private: `add_rcrd_log`, `cat_line`, `char`, `check_connection`, `cleanup`, `connected`, `connection_get`, `custom_download2`, `custom_download`, `db_sqlngths_get`, `db_sqlngths_log`, `dir_size`, `dwlnld_path_get`, `dwlnld_rcrd_log`, `entrez_fasta_get`, `entrez_gb_get`, `extract_accession`, `extract_by_patterns`, `extract_clean_sequence`, `extract_definition`, `extract_features`, `extract_inforepart`, `extract_locus`, `extract_organism`, `extract_seqrepart`, `extract_sequence`, `extract_version`, `file_download`, `filename_log`, `flatfile_read`, `gb_build2`, `gb_build`, `gb_df_create`, `gb_df_generate`, `gb_sql_add`, `gb_sql_query`, `gbrelease_check`, `gbrelease_get`, `gbrelease_log`, `has_data`, `identify_downloadable_files`, `last_add_get`, `last_dwlnld_get`, `last_entry_get`, `latest_genbank_release_notes`, `latest_genbank_release`, `message_missing`, `mock_def`, `mock_gb_df_generate`, `mock_org`, `mock_rec`, `mock_seq`, `predict_datasizes`,

print.status, quiet_connect, readme_log, restez_path_check, restez_rl, seshinfo_log, setup, slctn_get, slctn_log, sql_path_get, status_class, stat, testdatadir_get

extract_locus	<i>Extract locus</i>
---------------	----------------------

Description

Return locus information from GenBank record

Usage

```
extract_locus(record)
```

Arguments

record GenBank record in text format, character

Details

If element is not found, "" returned.

Value

named character vector

See Also

Other private: add_rcrd_log, cat_line, char, check_connection, cleanup, connected, connection_get, custom_download2, custom_download, db_sqlngths_get, db_sqlngths_log, dir_size, dwnld_path_get, dwnld_rcrd_log, entrez_fasta_get, entrez_gb_get, extract_accession, extract_by_patterns, extract_clean_sequence, extract_definition, extract_features, extract_inforecpart, extract_keywords, extract_organism, extract_seqrecpart, extract_sequence, extract_version, file_download, filename_log, flatfile_read, gb_build2, gb_build, gb_df_create, gb_df_generate, gb_sql_add, gb_sql_query, gbrelease_check, gbrelease_get, gbrelease_log, has_data, identify_downloadable_f, last_add_get, last_dwnld_get, last_entry_get, latest_genbank_release_notes, latest_genbank_release, message_missing, mock_def, mock_gb_df_generate, mock_org, mock_rec, mock_seq, predict_datasizes, print.status, quiet_connect, readme_log, restez_path_check, restez_rl, seshinfo_log, setup, slctn_get, slctn_log, sql_path_get, status_class, stat, testdatadir_get

extract_organism	<i>Extract organism</i>
------------------	-------------------------

Description

Return organism name from GenBank record

Usage

```
extract_organism(record)
```

Arguments

record	GenBank record in text format, character
--------	--

Details

If element is not found, "" returned.

Value

character

See Also

Other private: [add_rcrd_log](#), [cat_line](#), [char](#), [check_connection](#), [cleanup](#), [connected](#), [connection_get](#), [custom_download2](#), [custom_download](#), [db_sqlngths_get](#), [db_sqlngths_log](#), [dir_size](#), [dwnld_path_get](#), [dwnld_rcrd_log](#), [entrez_fasta_get](#), [entrez_gb_get](#), [extract_accession](#), [extract_by_patterns](#), [extract_clean_sequence](#), [extract_definition](#), [extract_features](#), [extract_inforecpart](#), [extract_keywords](#), [extract_locus](#), [extract_seqrecpart](#), [extract_sequence](#), [extract_version](#), [file_download](#), [filename_log](#), [flatfile_read](#), [gb_build2](#), [gb_build](#), [gb_df_create](#), [gb_df_generate](#), [gb_sql_add](#), [gb_sql_query](#), [gbrelease_check](#), [gbrelease_get](#), [gbrelease_log](#), [has_data](#), [identify_downloadable_f](#), [last_add_get](#), [last_dwnld_get](#), [last_entry_get](#), [latest_genbank_release_notes](#), [latest_genbank_release](#), [message_missing](#), [mock_def](#), [mock_gb_df_generate](#), [mock_org](#), [mock_rec](#), [mock_seq](#), [predict_datasizes](#), [print.status](#), [quiet_connect](#), [readme_log](#), [restez_path_check](#), [restez_rl](#), [seshinfo_log](#), [setup](#), [slctn_get](#), [slctn_log](#), [sql_path_get](#), [status_class](#), [stat](#), [testdatadir_get](#)

extract_seqrecpart	<i>Extract the sequence record part</i>
--------------------	---

Description

Return sequence part from GenBank record

Usage

```
extract_seqrecpart(record)
```

Arguments

record GenBank record in text format, character

Details

If element is not found, "" returned.

Value

character

See Also

Other private: [add_rcrd_log](#), [cat_line](#), [char](#), [check_connection](#), [cleanup](#), [connected](#), [connection_get](#), [custom_download2](#), [custom_download](#), [db_sqlngths_get](#), [db_sqlngths_log](#), [dir_size](#), [dwnld_path_get](#), [dwnld_rcrd_log](#), [entrez_fasta_get](#), [entrez_gb_get](#), [extract_accession](#), [extract_by_patterns](#), [extract_clean_sequence](#), [extract_definition](#), [extract_features](#), [extract_inforecpart](#), [extract_keywords](#), [extract_locus](#), [extract_organism](#), [extract_sequence](#), [extract_version](#), [file_download](#), [filename_log](#), [flatfile_read](#), [gb_build2](#), [gb_build](#), [gb_df_create](#), [gb_df_generate](#), [gb_sql_add](#), [gb_sql_query](#), [gbrelease_check](#), [gbrelease_get](#), [gbrelease_log](#), [has_data](#), [identify_downloadable_f](#), [last_add_get](#), [last_dwnld_get](#), [last_entry_get](#), [latest_genbank_release_notes](#), [latest_genbank_release](#), [message_missing](#), [mock_def](#), [mock_gb_df_generate](#), [mock_org](#), [mock_rec](#), [mock_seq](#), [predict_datasizes](#), [print.status](#), [quiet_connect](#), [readme_log](#), [restez_path_check](#), [restez_rl](#), [seshinfo_log](#), [setup](#), [slctn_get](#), [slctn_log](#), [sql_path_get](#), [status_class](#), [stat](#), [testdatadir_get](#)

extract_sequence	<i>Extract sequence</i>
------------------	-------------------------

Description

Return sequence from GenBank record

Usage

```
extract_sequence(record)
```

Arguments

record GenBank record in text format, character

Details

If element is not found, "" returned.

Value

character

See Also

Other private: add_rcrd_log, cat_line, char, check_connection, cleanup, connected, connection_get, custom_download2, custom_download, db_sqlngths_get, db_sqlngths_log, dir_size, dwnld_path_get, dwnld_rcrd_log, entrez_fasta_get, entrez_gb_get, extract_accession, extract_by_patterns, extract_clean_sequence, extract_definition, extract_features, extract_inforecpart, extract_keywords, extract_locus, extract_organism, extract_seqrecpart, extract_version, file_download, filename_log, flatfile_read, gb_build2, gb_build, gb_df_create, gb_df_generate, gb_sql_add, gb_sql_query, gbrelease_check, gbrelease_get, gbrelease_log, has_data, identify_downloadable_f, last_add_get, last_dwnld_get, last_entry_get, latest_genbank_release_notes, latest_genbank_release, message_missing, mock_def, mock_gb_df_generate, mock_org, mock_rec, mock_seq, predict_datasizes, print.status, quiet_connect, readme_log, restez_path_check, restez_rl, seshinfo_log, setup, slctn_get, slctn_log, sql_path_get, status_class, stat, testdatadir_get

extract_version	<i>Extract version</i>
-----------------	------------------------

Description

Return accession + version ID from GenBank record

Usage

```
extract_version(record)
```

Arguments

record	GenBank record in text format, character
--------	--

Details

If element is not found, "" returned.

Value

character

See Also

Other private: add_rcrd_log, cat_line, char, check_connection, cleanup, connected, connection_get, custom_download2, custom_download, db_sqlngths_get, db_sqlngths_log, dir_size, dwnld_path_get, dwnld_rcrd_log, entrez_fasta_get, entrez_gb_get, extract_accession, extract_by_patterns, extract_clean_sequence, extract_definition, extract_features, extract_inforecpart, extract_keywords, extract_locus, extract_organism, extract_seqrecpart, extract_sequence, file_download, filename_log, flatfile_read, gb_build2, gb_build, gb_df_create, gb_df_generate, gb_sql_add, gb_sql_query, gbrelease_check, gbrelease_get, gbrelease_log, has_data, identify_downloadable_f, last_add_get, last_dwnld_get, last_entry_get, latest_genbank_release_notes, latest_genbank_release, message_missing, mock_def, mock_gb_df_generate, mock_org, mock_rec, mock_seq, predict_datasizes, print.status, quiet_connect, readme_log, restez_path_check, restez_rl, seshinfo_log, setup, slctn_get, slctn_log, sql_path_get, status_class, stat, testdatadir_get

filename_log	<i>Write filenames to log files</i>
--------------	-------------------------------------

Description

Record a filename in a log file along with GB release and time.

Usage

```
filename_log(f1, fp)
```

Arguments

f1	file name, character
fp	filepath to log file, character

See Also

Other private: [add_rcrd_log](#), [cat_line](#), [char](#), [check_connection](#), [cleanup](#), [connected](#), [connection_get](#), [custom_download2](#), [custom_download](#), [db_sqlngths_get](#), [db_sqlngths_log](#), [dir_size](#), [dwnld_path_get](#), [dwnld_rcrd_log](#), [entrez_fasta_get](#), [entrez_gb_get](#), [extract_accession](#), [extract_by_patterns](#), [extract_clean_sequence](#), [extract_definition](#), [extract_features](#), [extract_inforecpart](#), [extract_keywords](#), [extract_locus](#), [extract_organism](#), [extract_seqrecpart](#), [extract_sequence](#), [extract_version](#), [file_download](#), [flatfile_read](#), [gb_build2](#), [gb_build](#), [gb_df_create](#), [gb_df_generate](#), [gb_sql_add](#), [gb_sql_query](#), [gbrelease_check](#), [gbrelease_get](#), [gbrelease_log](#), [has_data](#), [identify_downloadable_f](#), [last_add_get](#), [last_dwnld_get](#), [last_entry_get](#), [latest_genbank_release_notes](#), [latest_genbank_release](#), [message_missing](#), [mock_def](#), [mock_gb_df_generate](#), [mock_org](#), [mock_rec](#), [mock_seq](#), [predict_datasizes](#), [print.status](#), [quiet_connect](#), [readme_log](#), [restez_path_check](#), [restez_rl](#), [seshinfo_log](#), [setup](#), [slctn_get](#), [slctn_log](#), [sql_path_get](#), [status_class](#), [stat](#), [testdatadir_get](#)

file_download	<i>Download a file</i>
---------------	------------------------

Description

Download a GenBank .seq.tar file. Check the file has downloaded properly. If not, returns FALSE. If overwrite is true, any previous file will be overwritten.

Usage

```
file_download(f1, overwrite = FALSE)
```

Arguments

f1	character, base filename (e.g. gbpri9.seq) to be downloaded
overwrite	T/F

Value

T/F

See Also

Other private: `add_rcrd_log`, `cat_line`, `char`, `check_connection`, `cleanup`, `connected`, `connection_get`, `custom_download2`, `custom_download`, `db_sqlngths_get`, `db_sqlngths_log`, `dir_size`, `dnwld_path_get`, `dnwld_rcrd_log`, `entrez_fasta_get`, `entrez_gb_get`, `extract_accession`, `extract_by_patterns`, `extract_clean_sequence`, `extract_definition`, `extract_features`, `extract_inforecpart`, `extract_keywords`, `extract_locus`, `extract_organism`, `extract_seqrecpart`, `extract_sequence`, `extract_version`, `filename_log`, `flatfile_read`, `gb_build2`, `gb_build`, `gb_df_create`, `gb_df_generate`, `gb_sql_add`, `gb_sql_query`, `gbrelease_check`, `gbrelease_get`, `gbrelease_log`, `has_data`, `identify_downloadable_f`, `last_add_get`, `last_dnwld_get`, `last_entry_get`, `latest_genbank_release_notes`, `latest_genbank_release`, `message_missing`, `mock_def`, `mock_gb_df_generate`, `mock_org`, `mock_rec`, `mock_seq`, `predict_datasizes`, `print.status`, `quiet_connect`, `readme_log`, `restez_path_check`, `restez_rl`, `sessinfo_log`, `setup`, `slctn_get`, `slctn_log`, `sql_path_get`, `status_class`, `stat`, `testdatadir_get`

flatfile_read

*Read flatfile sequence records***Description**

Read records from a .seq file.

Usage

```
flatfile_read(flpth)
```

Arguments

flpth Path to .seq file

Value

list of GenBank records in text format

See Also

Other private: `add_rcrd_log`, `cat_line`, `char`, `check_connection`, `cleanup`, `connected`, `connection_get`, `custom_download2`, `custom_download`, `db_sqlngths_get`, `db_sqlngths_log`, `dir_size`, `dnwld_path_get`, `dnwld_rcrd_log`, `entrez_fasta_get`, `entrez_gb_get`, `extract_accession`, `extract_by_patterns`, `extract_clean_sequence`, `extract_definition`, `extract_features`, `extract_inforecpart`, `extract_keywords`, `extract_locus`, `extract_organism`, `extract_seqrecpart`, `extract_sequence`, `extract_version`, `file_download`, `filename_log`, `gb_build2`, `gb_build`, `gb_df_create`, `gb_df_generate`, `gb_sql_add`, `gb_sql_query`, `gbrelease_check`, `gbrelease_get`, `gbrelease_log`, `has_data`, `identify_downloadable_f`, `last_add_get`, `last_dnwld_get`, `last_entry_get`, `latest_genbank_release_notes`, `latest_genbank_release`, `message_missing`, `mock_def`, `mock_gb_df_generate`, `mock_org`, `mock_rec`, `mock_seq`, `predict_datasizes`,

[print.status](#), [quiet_connect](#), [readme_log](#), [restez_path_check](#), [restez_rl](#), [seshinfo_log](#), [setup](#), [slctn_get](#), [slctn_log](#), [sql_path_get](#), [status_class](#), [stat](#), [testdatadir_get](#)

gbrelease_check *Check if the last GenBank release number is the latest*

Description

Returns TRUE if the GenBank release number is the most recent GenBank release available.

Usage

```
gbrelease_check()
```

Value

logical

See Also

Other private: [add_rcrd_log](#), [cat_line](#), [char](#), [check_connection](#), [cleanup](#), [connected](#), [connection_get](#), [custom_download2](#), [custom_download](#), [db_sqlngths_get](#), [db_sqlngths_log](#), [dir_size](#), [dwnld_path_get](#), [dwnld_rcrd_log](#), [entrez_fasta_get](#), [entrez_gb_get](#), [extract_accession](#), [extract_by_patterns](#), [extract_clean_sequence](#), [extract_definition](#), [extract_features](#), [extract_inforecpart](#), [extract_keywords](#), [extract_locus](#), [extract_organism](#), [extract_seqrecpart](#), [extract_sequence](#), [extract_version](#), [file_download](#), [filename_log](#), [flatfile_read](#), [gb_build2](#), [gb_build](#), [gb_df_create](#), [gb_df_generate](#), [gb_sql_add](#), [gb_sql_query](#), [gbrelease_get](#), [gbrelease_log](#), [has_data](#), [identify_downloadable_fi](#), [last_add_get](#), [last_dwnld_get](#), [last_entry_get](#), [latest_genbank_release_notes](#), [latest_genbank_release](#), [message_missing](#), [mock_def](#), [mock_gb_df_generate](#), [mock_org](#), [mock_rec](#), [mock_seq](#), [predict_datasizes](#), [print.status](#), [quiet_connect](#), [readme_log](#), [restez_path_check](#), [restez_rl](#), [seshinfo_log](#), [setup](#), [slctn_get](#), [slctn_log](#), [sql_path_get](#), [status_class](#), [stat](#), [testdatadir_get](#)

gbrelease_get *Get the GenBank release number in the restez path*

Description

Returns the GenBank release number. Returns empty character if none found.

Usage

```
gbrelease_get()
```

Details

If no file found, returns empty character vector.

Value

character

See Also

Other private: `add_rcrd_log`, `cat_line`, `char`, `check_connection`, `cleanup`, `connected`, `connection_get`, `custom_download2`, `custom_download`, `db_sqlngths_get`, `db_sqlngths_log`, `dir_size`, `dnwld_path_get`, `dnwld_rcrd_log`, `entrez_fasta_get`, `entrez_gb_get`, `extract_accession`, `extract_by_patterns`, `extract_clean_sequence`, `extract_definition`, `extract_features`, `extract_inforecpart`, `extract_keywords`, `extract_locus`, `extract_organism`, `extract_seqrecpart`, `extract_sequence`, `extract_version`, `file_download`, `filename_log`, `flatfile_read`, `gb_build2`, `gb_build`, `gb_df_create`, `gb_df_generate`, `gb_sql_add`, `gb_sql_query`, `gbrelease_check`, `gbrelease_log`, `has_data`, `identify_downloadable_files`, `last_add_get`, `last_dnwld_get`, `last_entry_get`, `latest_genbank_release_notes`, `latest_genbank_release`, `message_missing`, `mock_def`, `mock_gb_df_generate`, `mock_org`, `mock_rec`, `mock_seq`, `predict_datasizes`, `print.status`, `quiet_connect`, `readme_log`, `restez_path_check`, `restez_rl`, `seshinfo_log`, `setup`, `slctn_get`, `slctn_log`, `sql_path_get`, `status_class`, `stat`, `testdatadir_get`

gbrelease_log

Log the GenBank release number in the restez path

Description

This function is called whenever `db_download` is run. It logs the GB release number in the 'gb_release.txt' in the user's `restez` path. The log is to help users keep track of whether their database is out of date.

Usage

```
gbrelease_log(release)
```

Arguments

release GenBank release number, character

See Also

Other private: `add_rcrd_log`, `cat_line`, `char`, `check_connection`, `cleanup`, `connected`, `connection_get`, `custom_download2`, `custom_download`, `db_sqlngths_get`, `db_sqlngths_log`, `dir_size`, `dnwld_path_get`, `dnwld_rcrd_log`, `entrez_fasta_get`, `entrez_gb_get`, `extract_accession`, `extract_by_patterns`, `extract_clean_sequence`, `extract_definition`, `extract_features`, `extract_inforecpart`, `extract_keywords`, `extract_locus`, `extract_organism`, `extract_seqrecpart`, `extract_sequence`, `extract_version`, `file_download`, `filename_log`, `flatfile_read`, `gb_build2`, `gb_build`, `gb_df_create`, `gb_df_generate`, `gb_sql_add`, `gb_sql_query`, `gbrelease_check`, `gbrelease_get`, `has_data`, `identify_downloadable_files`, `last_add_get`, `last_dnwld_get`, `last_entry_get`, `latest_genbank_release_notes`, `latest_genbank_release`, `message_missing`, `mock_def`, `mock_gb_df_generate`, `mock_org`, `mock_rec`, `mock_seq`, `predict_datasizes`, `print.status`, `quiet_connect`, `readme_log`, `restez_path_check`, `restez_rl`, `seshinfo_log`, `setup`, `slctn_get`, `slctn_log`, `sql_path_get`, `status_class`, `stat`, `testdatadir_get`

 gb_build

Read and add .seq files to database

Description

Given a list of seq_files, read and add the contents of the files to a SQL-like database. If any errors during the process, FALSE is returned.

Usage

```
gb_build(dpth, seq_files, max_length, min_length)
```

Arguments

dpth	Download path (where seq_files are stored)
seq_files	.seq.tar seq file names
max_length	Maximum sequence length
min_length	Minimum sequence length

Details

This function will automatically connect to the restez database.

Value

Logical

See Also

Other private: `add_rcrd_log`, `cat_line`, `char`, `check_connection`, `cleanup`, `connected`, `connection_get`, `custom_download2`, `custom_download`, `db_sqlngths_get`, `db_sqlngths_log`, `dir_size`, `dwncld_path_get`, `dwncld_rcrd_log`, `entrez_fasta_get`, `entrez_gb_get`, `extract_accession`, `extract_by_patterns`, `extract_clean_sequence`, `extract_definition`, `extract_features`, `extract_inforecpart`, `extract_keywords`, `extract_locus`, `extract_organism`, `extract_seqrecpart`, `extract_sequence`, `extract_version`, `file_download`, `filename_log`, `flatfile_read`, `gb_build2`, `gb_df_create`, `gb_df_generate`, `gb_sql_add`, `gb_sql_query`, `gbrelease_check`, `gbrelease_get`, `gbrelease_log`, `has_data`, `identify_downloadable_files`, `last_add_get`, `last_dwncld_get`, `last_entry_get`, `latest_genbank_release_notes`, `latest_genbank_release`, `message_missing`, `mock_def`, `mock_gb_df_generate`, `mock_org`, `mock_rec`, `mock_seq`, `predict_datasizes`, `print.status`, `quiet_connect`, `readme_log`, `restez_path_check`, `restez_rl`, `seshinfo_log`, `setup`, `slctn_get`, `slctn_log`, `sql_path_get`, `status_class`, `stat`, `testdatadir_get`

gb_build2	<i>Callr version of gb_build()</i>
-----------	------------------------------------

Description

Runs [gb_build](#) in callr. This allows the user to kill the process. Additionally, the process will print spinning dots to indicate it is still active.

Usage

```
gb_build2(dpth, seq_files, max_length, min_length)
```

Arguments

dpth	Download path (where seq_files are stored)
seq_files	.seq.tar seq file names
max_length	Maximum sequence length.
min_length	Minimum sequence length.

Value

Logical

See Also

Other private: [add_rcrd_log](#), [cat_line](#), [char](#), [check_connection](#), [cleanup](#), [connected](#), [connection_get](#), [custom_download2](#), [custom_download](#), [db_sqlngths_get](#), [db_sqlngths_log](#), [dir_size](#), [dwnld_path_get](#), [dwnld_rcrd_log](#), [entrez_fasta_get](#), [entrez_gb_get](#), [extract_accession](#), [extract_by_patterns](#), [extract_clean_sequence](#), [extract_definition](#), [extract_features](#), [extract_inforecpart](#), [extract_keywords](#), [extract_locus](#), [extract_organism](#), [extract_seqrecpart](#), [extract_sequence](#), [extract_version](#), [file_download](#), [filename_log](#), [flatfile_read](#), [gb_build](#), [gb_df_create](#), [gb_df_generate](#), [gb_sql_add](#), [gb_sql_query](#), [gbrelease_check](#), [gbrelease_get](#), [gbrelease_log](#), [has_data](#), [identify_downloadable_files](#), [last_add_get](#), [last_dwnld_get](#), [last_entry_get](#), [latest_genbank_release_notes](#), [latest_genbank_release](#), [message_missing](#), [mock_def](#), [mock_gb_df_generate](#), [mock_org](#), [mock_rec](#), [mock_seq](#), [predict_datasizes](#), [print.status](#), [quiet_connect](#), [readme_log](#), [restez_path_check](#), [restez_rl](#), [seshinfo_log](#), [setup](#), [slctn_get](#), [slctn_log](#), [sql_path_get](#), [status_class](#), [stat](#), [testdatadir_get](#)

gb_definition_get *Get definition from GenBank*

Description

Return the definition line for an accession ID.

Usage

```
gb_definition_get(id)
```

Arguments

id character, sequence accession ID(s)

Value

named vector of definitions, if no results found NULL

See Also

Other get: [gb_fasta_get](#), [gb_organism_get](#), [gb_record_get](#), [gb_sequence_get](#), [gb_version_get](#)

Examples

```
library(restez)
restez_path_set(filepath = tempdir())
demo_db_create(n = 5)
restez_connect()
(def <- gb_definition_get(id = 'demo_1'))
(defs <- gb_definition_get(id = c('demo_1', 'demo_2'))))

# delete demo after example
db_delete(everything = TRUE)
```

gb_df_create *Create GenBank data.frame*

Description

Make data.frame from columns vectors for nucleotide entries. As part of `gb_df_generate()`.

Usage

```
gb_df_create(accessions, versions, organisms, definitions, sequences,
             records)
```

Arguments

accessions	character, vector of accessions
versions	character, vector of accessions + versions
organisms	character, vector of organism names
definitions	character, vector of sequence definitions
sequences	character, vector of sequences
records	character, vector of GenBank records in text format

Value

data.frame

See Also

Other private: [add_rcrd_log](#), [cat_line](#), [char](#), [check_connection](#), [cleanup](#), [connected](#), [connection_get](#), [custom_download2](#), [custom_download](#), [db_sqlngths_get](#), [db_sqlngths_log](#), [dir_size](#), [dwnld_path_get](#), [dwnld_rcrd_log](#), [entrez_fasta_get](#), [entrez_gb_get](#), [extract_accession](#), [extract_by_patterns](#), [extract_clean_sequence](#), [extract_definition](#), [extract_features](#), [extract_inforecpart](#), [extract_keywords](#), [extract_locus](#), [extract_organism](#), [extract_seqrecpart](#), [extract_sequence](#), [extract_version](#), [file_download](#), [filename_log](#), [flatfile_read](#), [gb_build2](#), [gb_build](#), [gb_df_generate](#), [gb_sql_add](#), [gb_sql_query](#), [gbrelease_check](#), [gbrelease_get](#), [gbrelease_log](#), [has_data](#), [identify_downloadable_f](#), [last_add_get](#), [last_dwnld_get](#), [last_entry_get](#), [latest_genbank_release_notes](#), [latest_genbank_release](#), [message_missing](#), [mock_def](#), [mock_gb_df_generate](#), [mock_org](#), [mock_rec](#), [mock_seq](#), [predict_datasizes](#), [print.status](#), [quiet_connect](#), [readme_log](#), [restez_path_check](#), [restez_rl](#), [seshinfo_log](#), [setup](#), [slctn_get](#), [slctn_log](#), [sql_path_get](#), [status_class](#), [stat](#), [testdatadir_get](#)

gb_df_generate	<i>Generate GenBank records data.frame</i>
----------------	--

Description

For a list of records, construct a data.frame for insertion into SQL database.

Usage

```
gb_df_generate(records, min_length = 0, max_length = NULL)
```

Arguments

records	character, vector of GenBank records in text format
min_length	Minimum sequence length, default 0.
max_length	Maximum sequence length, default NULL.

Details

The resulting data.frame has five columns: accession, organism, raw_definition, raw_sequence, raw_record. The prefix 'raw_' indicates the data has been converted to the raw format, see ?char-ToRaw, in order to save on RAM. The raw_record contains the entire GenBank record in text format.

Use max and min sequence lengths to minimise the size of the database. All sequences have to be at least as long as min and less than or equal in length to max, unless max is NULL in which there is no maximum length.

Value

data.frame

See Also

Other private: [add_rcrd_log](#), [cat_line](#), [char](#), [check_connection](#), [cleanup](#), [connected](#), [connection_get](#), [custom_download2](#), [custom_download](#), [db_sqlngths_get](#), [db_sqlngths_log](#), [dir_size](#), [dwnld_path_get](#), [dwnld_rcrd_log](#), [entrez_fasta_get](#), [entrez_gb_get](#), [extract_accession](#), [extract_by_patterns](#), [extract_clean_sequence](#), [extract_definition](#), [extract_features](#), [extract_inforecpart](#), [extract_keywords](#), [extract_locus](#), [extract_organism](#), [extract_seqrecpart](#), [extract_sequence](#), [extract_version](#), [file_download](#), [filename_log](#), [flatfile_read](#), [gb_build2](#), [gb_build](#), [gb_df_create](#), [gb_sql_add](#), [gb_sql_query](#), [gbrelease_check](#), [gbrelease_get](#), [gbrelease_log](#), [has_data](#), [identify_downloadable_f](#), [last_add_get](#), [last_dwnld_get](#), [last_entry_get](#), [latest_genbank_release_notes](#), [latest_genbank_release](#), [message_missing](#), [mock_def](#), [mock_gb_df_generate](#), [mock_org](#), [mock_rec](#), [mock_seq](#), [predict_datasizes](#), [print.status](#), [quiet_connect](#), [readme_log](#), [restez_path_check](#), [restez_rl](#), [seshinfo_log](#), [setup](#), [slctn_get](#), [slctn_log](#), [sql_path_get](#), [status_class](#), [stat](#), [testdatadir_get](#)

gb_extract

Extract elements of a GenBank record

Description

Return elements of GenBank record e.g. sequence, definition ...

Usage

```
gb_extract(record, what = c("accession", "version", "organism",
  "sequence", "definition", "locus", "features", "keywords"))
```

Arguments

record	GenBank record in text format, character
what	Which element to extract

Details

This function uses a REGEX to extract particular elements of a GenBank record. All of the what options return a single character with the exception of 'locus' or 'keywords' that return character vectors and 'features' that returns a list of lists for all features.

The accuracy of these functions cannot be guaranteed due to the enormity of the GenBank database. But the function is regularly tested on a range of GenBank records.

Note: all non-latin1 characters are converted to '-'.

Value

character or list of lists (what='features') or named character vector (what='locus')

Examples

```
library(restez)
data('record')
(gb_extract(record = record, what = 'locus'))
```

gb_fasta_get

Get fasta from GenBank

Description

Get sequence and definition data in FASTA format. Equivalent to rettype='fasta' in [entrez_fetch](#).

Usage

```
gb_fasta_get(id, width = 70)
```

Arguments

id	character, sequence accession ID(s)
width	integer, maximum number of characters in a line

Value

named vector of fasta sequences, if no results found NULL

See Also

Other get: [gb_definition_get](#), [gb_organism_get](#), [gb_record_get](#), [gb_sequence_get](#), [gb_version_get](#)

Examples

```
library(restez)
restez_path_set(filepath = tempdir())
demo_db_create(n = 5)
restez_connect()
(fasta <- gb_fasta_get(id = 'demo_1'))
(fastas <- gb_fasta_get(id = c('demo_1', 'demo_2'))))

# delete demo after example
db_delete(everything = TRUE)
```

gb_organism_get	<i>Get organism from GenBank</i>
-----------------	----------------------------------

Description

Return the organism name for an accession ID.

Usage

```
gb_organism_get(id)
```

Arguments

id character, sequence accession ID(s)

Value

named vector of definitions, if no results found NULL

See Also

Other get: [gb_definition_get](#), [gb_fasta_get](#), [gb_record_get](#), [gb_sequence_get](#), [gb_version_get](#)

Examples

```
library(restez)
restez_path_set(filepath = tempdir())
demo_db_create(n = 5)
restez_connect()
(org <- gb_organism_get(id = 'demo_1'))
(orgs <- gb_organism_get(id = c('demo_1', 'demo_2'))))

# delete demo after example
db_delete(everything = TRUE)
```

gb_record_get	<i>Get record from GenBank</i>
---------------	--------------------------------

Description

Return the entire GenBank record for an accession ID. Equivalent to `ret type='gb'` in [entrez_fetch](#).

Usage

```
gb_record_get(id)
```

Arguments

id character, sequence accession ID(s)

Value

named vector of records, if no results found NULL

See Also

Other get: [gb_definition_get](#), [gb_fasta_get](#), [gb_organism_get](#), [gb_sequence_get](#), [gb_version_get](#)

Examples

```
library(restez)
restez_path_set(filepath = tempdir())
demo_db_create(n = 5)
restez_connect()
(rec <- gb_record_get(id = 'demo_1'))
(recs <- gb_record_get(id = c('demo_1', 'demo_2')))
```



```
# delete demo after example
db_delete(everything = TRUE)
```

gb_sequence_get	<i>Get sequence from GenBank</i>
-----------------	----------------------------------

Description

Return the sequence(s) for a record(s) from the accession ID(s).

Usage

```
gb_sequence_get(id)
```

Arguments

id character, sequence accession ID(s)

Value

named vector of sequences, if no results found NULL

See Also

Other get: [gb_definition_get](#), [gb_fasta_get](#), [gb_organism_get](#), [gb_record_get](#), [gb_version_get](#)

Examples

```
library(restez)
restez_path_set(filepath = tempdir())
demo_db_create(n = 5)
restez_connect()
(seq <- gb_sequence_get(id = 'demo_1'))
(seqs <- gb_sequence_get(id = c('demo_1', 'demo_2'))))

# delete demo after example
db_delete(everything = TRUE)
```

gb_sql_add

Add to GenBank SQL database

Description

Add records data.frame to SQL-like database.

Usage

```
gb_sql_add(df)
```

Arguments

df Records data.frame

See Also

Other private: [add_rcrd_log](#), [cat_line](#), [char](#), [check_connection](#), [cleanup](#), [connected](#), [connection_get](#), [custom_download2](#), [custom_download](#), [db_sqlngths_get](#), [db_sqlngths_log](#), [dir_size](#), [dwnld_path_get](#), [dwnld_rcrd_log](#), [entrez_fasta_get](#), [entrez_gb_get](#), [extract_accession](#), [extract_by_patterns](#), [extract_clean_sequence](#), [extract_definition](#), [extract_features](#), [extract_inforecpart](#), [extract_keywords](#), [extract_locus](#), [extract_organism](#), [extract_seqrecpart](#), [extract_sequence](#), [extract_version](#), [file_download](#), [filename_log](#), [flatfile_read](#), [gb_build2](#), [gb_build](#), [gb_df_create](#), [gb_df_generate](#), [gb_sql_query](#), [gbrelease_check](#), [gbrelease_get](#), [gbrelease_log](#), [has_data](#),

identify_downloadable_files, last_add_get, last_dwnld_get, last_entry_get, latest_genbank_release_notes, latest_genbank_release, message_missing, mock_def, mock_gb_df_generate, mock_org, mock_rec, mock_seq, predict_datasizes, print.status, quiet_connect, readme_log, restez_path_check, restez_rl, seshinfo_log, setup, slctn_get, slctn_log, sql_path_get, status_class, stat, testdatadir_get

 gb_sql_query

Query the GenBank SQL

Description

Generic query function for retrieving data from the SQL database for the get functions.

Usage

```
gb_sql_query(nm, id)
```

Arguments

nm	character, column name
id	character, sequence accession ID(s)

Value

data.frame

See Also

Other private: add_rcrd_log, cat_line, char, check_connection, cleanup, connected, connection_get, custom_download2, custom_download, db_sqlngths_get, db_sqlngths_log, dir_size, dwnld_path_get, dwnld_rcrd_log, entrez_fasta_get, entrez_gb_get, extract_accession, extract_by_patterns, extract_clean_sequence, extract_definition, extract_features, extract_inforecpart, extract_keywords, extract_locus, extract_organism, extract_seqrecpart, extract_sequence, extract_version, file_download, filename_log, flatfile_read, gb_build2, gb_build, gb_df_create, gb_df_generate, gb_sql_add, gbrelease_check, gbrelease_get, gbrelease_log, has_data, identify_downloadable_files, last_add_get, last_dwnld_get, last_entry_get, latest_genbank_release_notes, latest_genbank_release, message_missing, mock_def, mock_gb_df_generate, mock_org, mock_rec, mock_seq, predict_datasizes, print.status, quiet_connect, readme_log, restez_path_check, restez_rl, seshinfo_log, setup, slctn_get, slctn_log, sql_path_get, status_class, stat, testdatadir_get

gb_version_get	<i>Get version from GenBank</i>
----------------	---------------------------------

Description

Return the accession version for an accession ID.

Usage

```
gb_version_get(id)
```

Arguments

id character, sequence accession ID(s)

Value

named vector of versions, if no results found NULL

See Also

Other get: [gb_definition_get](#), [gb_fasta_get](#), [gb_organism_get](#), [gb_record_get](#), [gb_sequence_get](#)

Examples

```
library(restez)
restez_path_set(filepath = tempdir())
demo_db_create(n = 5)
restez_connect()
(ver <- gb_version_get(id = 'demo_1'))
(vers <- gb_version_get(id = c('demo_1', 'demo_2')))
```



```
# delete demo after example
db_delete(everything = TRUE)
```

has_data	<i>Does the connected database have data?</i>
----------	---

Description

Returns TRUE if a restez SQL database has data.

Usage

```
has_data()
```

Value

Logical

See Also

Other private: `add_rcrd_log`, `cat_line`, `char`, `check_connection`, `cleanup`, `connected`, `connection_get`, `custom_download2`, `custom_download`, `db_sqlngths_get`, `db_sqlngths_log`, `dir_size`, `dwncld_path_get`, `dwncld_rcrd_log`, `entrez_fasta_get`, `entrez_gb_get`, `extract_accession`, `extract_by_patterns`, `extract_clean_sequence`, `extract_definition`, `extract_features`, `extract_inforecpart`, `extract_keywords`, `extract_locus`, `extract_organism`, `extract_seqrecpart`, `extract_sequence`, `extract_version`, `file_download`, `filename_log`, `flatfile_read`, `gb_build2`, `gb_build`, `gb_df_create`, `gb_df_generate`, `gb_sql_add`, `gb_sql_query`, `gbrelease_check`, `gbrelease_get`, `gbrelease_log`, `identify_downloadable_files`, `last_add_get`, `last_dwncld_get`, `last_entry_get`, `latest_genbank_release_notes`, `latest_genbank_release`, `message_missing`, `mock_def`, `mock_gb_df_generate`, `mock_org`, `mock_rec`, `mock_seq`, `predict_datasizes`, `print.status`, `quiet_connect`, `readme_log`, `restez_path_check`, `restez_rl`, `sessinfo_log`, `setup`, `slctn_get`, `slctn_log`, `sql_path_get`, `status_class`, `stat`, `testdatadir_get`

identify_downloadable_files

Identify downloadable files

Description

Searches through the release notes for a GenBank release to find all listed .seq files. Returns a data.frame for all .seq files and their description.

Usage

```
identify_downloadable_files()
```

Value

data.frame

See Also

Other private: `add_rcrd_log`, `cat_line`, `char`, `check_connection`, `cleanup`, `connected`, `connection_get`, `custom_download2`, `custom_download`, `db_sqlngths_get`, `db_sqlngths_log`, `dir_size`, `dwncld_path_get`, `dwncld_rcrd_log`, `entrez_fasta_get`, `entrez_gb_get`, `extract_accession`, `extract_by_patterns`, `extract_clean_sequence`, `extract_definition`, `extract_features`, `extract_inforecpart`, `extract_keywords`, `extract_locus`, `extract_organism`, `extract_seqrecpart`, `extract_sequence`, `extract_version`, `file_download`, `filename_log`, `flatfile_read`, `gb_build2`, `gb_build`, `gb_df_create`, `gb_df_generate`, `gb_sql_add`, `gb_sql_query`, `gbrelease_check`, `gbrelease_get`, `gbrelease_log`, `has_data`, `last_add_get`, `last_dwncld_get`, `last_entry_get`, `latest_genbank_release_notes`, `latest_genbank_release`, `message_missing`, `mock_def`, `mock_gb_df_generate`, `mock_org`, `mock_rec`, `mock_seq`, `predict_datasizes`, `print.status`, `quiet_connect`, `readme_log`, `restez_path_check`,

[restez_rl](#), [seshinfo_log](#), [setup](#), [slctn_get](#), [slctn_log](#), [sql_path_get](#), [status_class](#), [stat](#), [testdatadir_get](#)

is_in_db

Is in db

Description

Determine whether an id(s) is/are present in a database.

Usage

```
is_in_db(id, db = "nucleotide")
```

Arguments

id	character, sequence accession ID(s)
db	character, database name

Value

named vector of booleans

See Also

Other database: [count_db_ids](#), [db_create](#), [db_delete](#), [db_download](#), [demo_db_create](#), [list_db_ids](#)

Examples

```
library(restez)
# set the restez path to a temporary dir
restez_path_set(filepath = tempdir())
# create demo database
demo_db_create(n = 5)
restez_connect()
# in the demo, IDs are 'demo_1', 'demo_2' ...
ids <- c('thisisnotanid', 'demo_1', 'demo_2')
(is_in_db(id = ids))

# delete demo after example
db_delete(everything = TRUE)
```

last_add_get	<i>Return date and time of the last added sequence</i>
--------------	--

Description

Return the date and time of the last added sequence as determined using the 'add_log.tsv'.

Usage

```
last_add_get()
```

Details

If no file found, returns empty character vector.

Value

character

See Also

Other private: [add_rcrd_log](#), [cat_line](#), [char](#), [check_connection](#), [cleanup](#), [connected](#), [connection_get](#), [custom_download2](#), [custom_download](#), [db_sqlngths_get](#), [db_sqlngths_log](#), [dir_size](#), [dwnld_path_get](#), [dwnld_rcrd_log](#), [entrez_fasta_get](#), [entrez_gb_get](#), [extract_accession](#), [extract_by_patterns](#), [extract_clean_sequence](#), [extract_definition](#), [extract_features](#), [extract_inforecpart](#), [extract_keywords](#), [extract_locus](#), [extract_organism](#), [extract_seqrecpart](#), [extract_sequence](#), [extract_version](#), [file_download](#), [filename_log](#), [flatfile_read](#), [gb_build2](#), [gb_build](#), [gb_df_create](#), [gb_df_generate](#), [gb_sql_add](#), [gb_sql_query](#), [gbrelease_check](#), [gbrelease_get](#), [gbrelease_log](#), [has_data](#), [identify_downloadable_files](#), [last_dwnld_get](#), [last_entry_get](#), [latest_genbank_release_notes](#), [latest_genbank_release](#), [message_missing](#), [mock_def](#), [mock_gb_df_generate](#), [mock_org](#), [mock_rec](#), [mock_seq](#), [predict_datasizes](#), [print.status](#), [quiet_connect](#), [readme_log](#), [restez_path_check](#), [restez_rl](#), [seshinfo_log](#), [setup](#), [slctn_get](#), [slctn_log](#), [sql_path_get](#), [status_class](#), [stat](#), [testdatadir_get](#)

last_dwnld_get	<i>Return date and time of the last download</i>
----------------	--

Description

Return the date and time of the last download as determined using the 'download_log.tsv'.

Usage

```
last_dwnld_get()
```

Details

If no file found, returns empty character vector.

Value

character

See Also

Other private: [add_rcrd_log](#), [cat_line](#), [char](#), [check_connection](#), [cleanup](#), [connected](#), [connection_get](#), [custom_download2](#), [custom_download](#), [db_sqlngths_get](#), [db_sqlngths_log](#), [dir_size](#), [dwnld_path_get](#), [dwnld_rcrd_log](#), [entrez_fasta_get](#), [entrez_gb_get](#), [extract_accession](#), [extract_by_patterns](#), [extract_clean_sequence](#), [extract_definition](#), [extract_features](#), [extract_inforecpart](#), [extract_keywords](#), [extract_locus](#), [extract_organism](#), [extract_seqrecpart](#), [extract_sequence](#), [extract_version](#), [file_download](#), [filename_log](#), [flatfile_read](#), [gb_build2](#), [gb_build](#), [gb_df_create](#), [gb_df_generate](#), [gb_sql_add](#), [gb_sql_query](#), [gbrelease_check](#), [gbrelease_get](#), [gbrelease_log](#), [has_data](#), [identify_downloadable_files](#), [last_add_get](#), [last_entry_get](#), [latest_genbank_release_notes](#), [latest_genbank_release](#), [message_missing](#), [mock_def](#), [mock_gb_df_generate](#), [mock_org](#), [mock_rec](#), [mock_seq](#), [predict_datasizes](#), [print.status](#), [quiet_connect](#), [readme_log](#), [restez_path_check](#), [restez_rl](#), [seshinfo_log](#), [setup](#), [slctn_get](#), [slctn_log](#), [sql_path_get](#), [status_class](#), [stat](#), [testdatadir_get](#)

last_entry_get	<i>Return the last entry</i>
----------------	------------------------------

Description

Return the last entry from a tab-delimited log file.

Usage

```
last_entry_get(fp)
```

Arguments

fp	Filepath, character
----	---------------------

Value

vector

See Also

Other private: `add_rcrd_log`, `cat_line`, `char`, `check_connection`, `cleanup`, `connected`, `connection_get`, `custom_download2`, `custom_download`, `db_sqlngths_get`, `db_sqlngths_log`, `dir_size`, `dnwld_path_get`, `dnwld_rcrd_log`, `entrez_fasta_get`, `entrez_gb_get`, `extract_accession`, `extract_by_patterns`, `extract_clean_sequence`, `extract_definition`, `extract_features`, `extract_inforecpart`, `extract_keywords`, `extract_locus`, `extract_organism`, `extract_seqrecpart`, `extract_sequence`, `extract_version`, `file_download`, `filename_log`, `flatfile_read`, `gb_build2`, `gb_build`, `gb_df_create`, `gb_df_generate`, `gb_sql_add`, `gb_sql_query`, `gbrelease_check`, `gbrelease_get`, `gbrelease_log`, `has_data`, `identify_downloadable_files`, `last_add_get`, `last_dnwld_get`, `latest_genbank_release_notes`, `latest_genbank_release`, `message_missing`, `mock_def`, `mock_gb_df_generate`, `mock_org`, `mock_rec`, `mock_seq`, `predict_datasizes`, `print.status`, `quiet_connect`, `readme_log`, `restez_path_check`, `restez_rl`, `seshinfo_log`, `setup`, `slctn_get`, `slctn_log`, `sql_path_get`, `status_class`, `stat`, `testdatadir_get`

latest_genbank_release

Retrieve latest GenBank release number

Description

Downloads the latest GenBank release number and returns it.

Usage

```
latest_genbank_release()
```

Value

character

See Also

Other private: `add_rcrd_log`, `cat_line`, `char`, `check_connection`, `cleanup`, `connected`, `connection_get`, `custom_download2`, `custom_download`, `db_sqlngths_get`, `db_sqlngths_log`, `dir_size`, `dnwld_path_get`, `dnwld_rcrd_log`, `entrez_fasta_get`, `entrez_gb_get`, `extract_accession`, `extract_by_patterns`, `extract_clean_sequence`, `extract_definition`, `extract_features`, `extract_inforecpart`, `extract_keywords`, `extract_locus`, `extract_organism`, `extract_seqrecpart`, `extract_sequence`, `extract_version`, `file_download`, `filename_log`, `flatfile_read`, `gb_build2`, `gb_build`, `gb_df_create`, `gb_df_generate`, `gb_sql_add`, `gb_sql_query`, `gbrelease_check`, `gbrelease_get`, `gbrelease_log`, `has_data`, `identify_downloadable_files`, `last_add_get`, `last_dnwld_get`, `last_entry_get`, `latest_genbank_release_notes`, `message_missing`, `mock_def`, `mock_gb_df_generate`, `mock_org`, `mock_rec`, `mock_seq`, `predict_datasizes`, `print.status`, `quiet_connect`, `readme_log`, `restez_path_check`, `restez_rl`, `seshinfo_log`, `setup`, `slctn_get`, `slctn_log`, `sql_path_get`, `status_class`, `stat`, `testdatadir_get`

latest_genbank_release_notes

Download the latest GenBank Release Notes

Description

Downloads the latest GenBank release notes to a user's restez download path.

Usage

```
latest_genbank_release_notes()
```

See Also

Other private: [add_rcrd_log](#), [cat_line](#), [char](#), [check_connection](#), [cleanup](#), [connected](#), [connection_get](#), [custom_download2](#), [custom_download](#), [db_sqlngths_get](#), [db_sqlngths_log](#), [dir_size](#), [dwnld_path_get](#), [dwnld_rcrd_log](#), [entrez_fasta_get](#), [entrez_gb_get](#), [extract_accession](#), [extract_by_patterns](#), [extract_clean_sequence](#), [extract_definition](#), [extract_features](#), [extract_inforecpart](#), [extract_keywords](#), [extract_locus](#), [extract_organism](#), [extract_seqrecpart](#), [extract_sequence](#), [extract_version](#), [file_download](#), [filename_log](#), [flatfile_read](#), [gb_build2](#), [gb_build](#), [gb_df_create](#), [gb_df_generate](#), [gb_sql_add](#), [gb_sql_query](#), [gbrelease_check](#), [gbrelease_get](#), [gbrelease_log](#), [has_data](#), [identify_downloadable_files](#), [last_add_get](#), [last_dwnld_get](#), [last_entry_get](#), [latest_genbank_release](#), [message_missing](#), [mock_def](#), [mock_gb_df_generate](#), [mock_org](#), [mock_rec](#), [mock_seq](#), [predict_datasizes](#), [print.status](#), [quiet_connect](#), [readme_log](#), [restez_path_check](#), [restez_rl](#), [seshinfo_log](#), [setup](#), [slctn_get](#), [slctn_log](#), [sql_path_get](#), [status_class](#), [stat](#), [testdatadir_get](#)

list_db_ids

List database IDs

Description

Return a vector of all IDs in a database.

Usage

```
list_db_ids(db = "nucleotide", n = 100)
```

Arguments

db	character, database name
n	Maximum number of IDs to return, if NULL returns all

Details

Warning: can return very large vectors for large databases.

Value

vector of characters

See Also

Other database: [count_db_ids](#), [db_create](#), [db_delete](#), [db_download](#), [demo_db_create](#), [is_in_db](#)

Examples

```
library(restez)
restez_path_set(filepath = tempdir())
demo_db_create(n = 5)
# Warning: not recommended for real databases
# with potentially millions of IDs
restez_connect()
all_ids <- list_db_ids()

# What shall we do with these IDs?
# ... how about make a mock fasta file
seqs <- gb_sequence_get(id = all_ids)
defs <- gb_definition_get(id = all_ids)
# paste together
fasta_seqs <- paste0('>', defs, '\n', seqs)
fasta_file <- paste0(fasta_seqs, collapse = '\n')
cat(fasta_file)

# delete after example
db_delete(everything = TRUE)
```

message_missing *Produce message of missing IDs*

Description

Sends message to console stating number of missing IDs.

Usage

```
message_missing(n)
```

Arguments

n Number of missing IDs

See Also

Other private: add_rcrd_log, cat_line, char, check_connection, cleanup, connected, connection_get, custom_download2, custom_download, db_sqlngths_get, db_sqlngths_log, dir_size, dwnld_path_get, dwnld_rcrd_log, entrez_fasta_get, entrez_gb_get, extract_accession, extract_by_patterns, extract_clean_sequence, extract_definition, extract_features, extract_inforecpart, extract_keywords, extract_locus, extract_organism, extract_seqrecpart, extract_sequence, extract_version, file_download, filename_log, flatfile_read, gb_build2, gb_build, gb_df_create, gb_df_generate, gb_sql_add, gb_sql_query, gbrelease_check, gbrelease_get, gbrelease_log, has_data, identify_downloadable_files, last_add_get, last_dwnld_get, last_entry_get, latest_genbank_release_notes, latest_genbank_release, mock_def, mock_gb_df_generate, mock_org, mock_rec, mock_seq, predict_datasizes, print.status, quiet_connect, readme_log, restez_path_check, restez_rl, seshinfo_log, setup, slctn_get, slctn_log, sql_path_get, status_class, stat, testdatadir_get

 mock_def

Mock def

Description

Make a mock sequence definition. Designed to be part of a loop.

Usage

```
mock_def(i)
```

Arguments

i integer, iterator

Value

character

See Also

Other private: add_rcrd_log, cat_line, char, check_connection, cleanup, connected, connection_get, custom_download2, custom_download, db_sqlngths_get, db_sqlngths_log, dir_size, dwnld_path_get, dwnld_rcrd_log, entrez_fasta_get, entrez_gb_get, extract_accession, extract_by_patterns, extract_clean_sequence, extract_definition, extract_features, extract_inforecpart, extract_keywords, extract_locus, extract_organism, extract_seqrecpart, extract_sequence, extract_version, file_download, filename_log, flatfile_read, gb_build2, gb_build, gb_df_create, gb_df_generate, gb_sql_add, gb_sql_query, gbrelease_check, gbrelease_get, gbrelease_log, has_data, identify_downloadable_files, last_add_get, last_dwnld_get, last_entry_get, latest_genbank_release_notes, latest_genbank_release, message_missing, mock_gb_df_generate, mock_org, mock_rec, mock_seq, predict_datasizes, print.status, quiet_connect, readme_log, restez_path_check, restez_rl, seshinfo_log, setup, slctn_get, slctn_log, sql_path_get, status_class, stat, testdatadir_get

mock_gb_df_generate	<i>Generate mock GenBank records data.frame</i>
---------------------	---

Description

Make a mock nucleotide data.frame for entry into a demonstration SQL database.

Usage

```
mock_gb_df_generate(n)
```

Arguments

n integer, number of entries

Value

data.frame

See Also

Other private: [add_rcrd_log](#), [cat_line](#), [char](#), [check_connection](#), [cleanup](#), [connected](#), [connection_get](#), [custom_download2](#), [custom_download](#), [db_sqlngths_get](#), [db_sqlngths_log](#), [dir_size](#), [dwnld_path_get](#), [dwnld_rcrd_log](#), [entrez_fasta_get](#), [entrez_gb_get](#), [extract_accession](#), [extract_by_patterns](#), [extract_clean_sequence](#), [extract_definition](#), [extract_features](#), [extract_inforecpart](#), [extract_keywords](#), [extract_locus](#), [extract_organism](#), [extract_seqrecpart](#), [extract_sequence](#), [extract_version](#), [file_download](#), [filename_log](#), [flatfile_read](#), [gb_build2](#), [gb_build](#), [gb_df_create](#), [gb_df_generate](#), [gb_sql_add](#), [gb_sql_query](#), [gbrelease_check](#), [gbrelease_get](#), [gbrelease_log](#), [has_data](#), [identify_downloadable_files](#), [last_add_get](#), [last_dwnld_get](#), [last_entry_get](#), [latest_genbank_release_notes](#), [latest_genbank_release](#), [message_missing](#), [mock_def](#), [mock_org](#), [mock_rec](#), [mock_seq](#), [predict_datasizes](#), [print.status](#), [quiet_connect](#), [readme_log](#), [restez_path_check](#), [restez_rl](#), [seshinfo_log](#), [setup](#), [slctn_get](#), [slctn_log](#), [sql_path_get](#), [status_class](#), [stat](#), [testdatadir_get](#)

mock_org	<i>Mock org</i>
----------	-----------------

Description

Make a mock sequence organism. Designed to be part of a loop.

Usage

```
mock_org(i)
```

Arguments

i integer, iterator

Value

character

See Also

Other private: add_rcrd_log, cat_line, char, check_connection, cleanup, connected, connection_get, custom_download2, custom_download, db_sqlngths_get, db_sqlngths_log, dir_size, dwnld_path_get, dwnld_rcrd_log, entrez_fasta_get, entrez_gb_get, extract_accession, extract_by_patterns, extract_clean_sequence, extract_definition, extract_features, extract_inforecpart, extract_keywords, extract_locus, extract_organism, extract_seqrecpart, extract_sequence, extract_version, file_download, filename_log, flatfile_read, gb_build2, gb_build, gb_df_create, gb_df_generate, gb_sql_add, gb_sql_query, gbrelease_check, gbrelease_get, gbrelease_log, has_data, identify_downloadable_files, last_add_get, last_dwnld_get, last_entry_get, latest_genbank_release_notes, latest_genbank_release, message_missing, mock_def, mock_gb_df_generate, mock_rec, mock_seq, predict_datasizes, print.status, quiet_connect, readme_log, restez_path_check, restez_rl, seshinfo_log, setup, slctn_get, slctn_log, sql_path_get, status_class, stat, testdatadir_get

mock_rec

Mock rec

Description

Create a mock GenBank record for demo-ing and testing purposes. Designed to be part of a loop. Accession, organism... etc. are optional arguments.

Usage

```
mock_rec(i, definition = NULL, accession = NULL, version = NULL,
         organism = NULL, sequence = NULL)
```

Arguments

i integer, iterator
 definition character
 accession character
 version character
 organism character
 sequence character

Value

character

See Also

Other private: add_rcrd_log, cat_line, char, check_connection, cleanup, connected, connection_get, custom_download2, custom_download, db_sqlngths_get, db_sqlngths_log, dir_size, dnwld_path_get, dnwld_rcrd_log, entrez_fasta_get, entrez_gb_get, extract_accession, extract_by_patterns, extract_clean_sequence, extract_definition, extract_features, extract_inforecpart, extract_keywords, extract_locus, extract_organism, extract_seqrecpart, extract_sequence, extract_version, file_download, filename_log, flatfile_read, gb_build2, gb_build, gb_df_create, gb_df_generate, gb_sql_add, gb_sql_query, gbrelease_check, gbrelease_get, gbrelease_log, has_data, identify_downloadable_files, last_add_get, last_dnwld_get, last_entry_get, latest_genbank_release_notes, latest_genbank_release, message_missing, mock_def, mock_gb_df_generate, mock_org, mock_seq, predict_datasizes, print.status, quiet_connect, readme_log, restez_path_check, restez_rl, seshinfo_log, setup, slctn_get, slctn_log, sql_path_get, status_class, stat, testdatadir_get

mock_seq

Mock seq

Description

Make a mock sequence. Designed to be part of a loop.

Usage

```
mock_seq(i, sqlngth = 10)
```

Arguments

i	integer, iterator
sqlngth	integer, sequence length

Value

character

See Also

Other private: add_rcrd_log, cat_line, char, check_connection, cleanup, connected, connection_get, custom_download2, custom_download, db_sqlngths_get, db_sqlngths_log, dir_size, dnwld_path_get, dnwld_rcrd_log, entrez_fasta_get, entrez_gb_get, extract_accession, extract_by_patterns, extract_clean_sequence, extract_definition, extract_features, extract_inforecpart, extract_keywords, extract_locus, extract_organism, extract_seqrecpart, extract_sequence, extract_version, file_download, filename_log, flatfile_read, gb_build2, gb_build, gb_df_create, gb_df_generate, gb_sql_add, gb_sql_query, gbrelease_check, gbrelease_get, gbrelease_log, has_data, identify_downloadable_files, last_add_get, last_dnwld_get, last_entry_get, latest_genbank_release_notes, latest_genbank_release, message_missing, mock_def, mock_gb_df_generate, mock_org, mock_rec, predict_datasizes, print.status, quiet_connect, readme_log, restez_path_check, restez_rl, seshinfo_log, setup, slctn_get, slctn_log, sql_path_get, status_class, stat, testdatadir_get

`predict_datasizes` *Print file size predictions to screen*

Description

Predicts the file sizes of the downloads and the database from the GenBank filesize information. Conversion factors are based on previous restez downloads.

Usage

```
predict_datasizes(uncompressed_filesize)
```

Arguments

```
uncompressed_filesize
    GBs of the stated filesize, numeric
```

See Also

Other private: [add_rcrd_log](#), [cat_line](#), [char](#), [check_connection](#), [cleanup](#), [connected](#), [connection_get](#), [custom_download2](#), [custom_download](#), [db_sqlngths_get](#), [db_sqlngths_log](#), [dir_size](#), [dwnld_path_get](#), [dwnld_rcrd_log](#), [entrez_fasta_get](#), [entrez_gb_get](#), [extract_accession](#), [extract_by_patterns](#), [extract_clean_sequence](#), [extract_definition](#), [extract_features](#), [extract_inforecpart](#), [extract_keywords](#), [extract_locus](#), [extract_organism](#), [extract_seqrecpart](#), [extract_sequence](#), [extract_version](#), [file_download](#), [filename_log](#), [flatfile_read](#), [gb_build2](#), [gb_build](#), [gb_df_create](#), [gb_df_generate](#), [gb_sql_add](#), [gb_sql_query](#), [gbrelease_check](#), [gbrelease_get](#), [gbrelease_log](#), [has_data](#), [identify_downloadable_files](#), [last_add_get](#), [last_dwnld_get](#), [last_entry_get](#), [latest_genbank_release_notes](#), [latest_genbank_release](#), [message_missing](#), [mock_def](#), [mock_gb_df_generate](#), [mock_org](#), [mock_rec](#), [mock_seq](#), [print.status](#), [quiet_connect](#), [readme_log](#), [restez_path_check](#), [restez_rl](#), [seshinfo_log](#), [setup](#), [slctn_get](#), [slctn_log](#), [sql_path_get](#), [status_class](#), [stat](#), [testdatadir_get](#)

`print.status` *Print method for status class*

Description

Prints to screen the three sections of the status class.

Usage

```
## S3 method for class 'status'
print(x)
```


Arguments

x Status object

See Also

Other private: `add_rcrd_log`, `cat_line`, `char`, `check_connection`, `cleanup`, `connected`, `connection_get`, `custom_download2`, `custom_download`, `db_sqlngths_get`, `db_sqlngths_log`, `dir_size`, `dwncld_path_get`, `dwncld_rcrd_log`, `entrez_fasta_get`, `entrez_gb_get`, `extract_accession`, `extract_by_patterns`, `extract_clean_sequence`, `extract_definition`, `extract_features`, `extract_inforecpart`, `extract_keywords`, `extract_locus`, `extract_organism`, `extract_seqrecpart`, `extract_sequence`, `extract_version`, `file_download`, `filename_log`, `flatfile_read`, `gb_build2`, `gb_build`, `gb_df_create`, `gb_df_generate`, `gb_sql_add`, `gb_sql_query`, `gbrelease_check`, `gbrelease_get`, `gbrelease_log`, `has_data`, `identify_downloadable_files`, `last_add_get`, `last_dwncld_get`, `last_entry_get`, `latest_genbank_release_notes`, `latest_genbank_release`, `message_missing`, `mock_def`, `mock_gb_df_generate`, `mock_org`, `mock_rec`, `mock_seq`, `predict_datasizes`, `quiet_connect`, `readme_log`, `restez_path_check`, `restez_rl`, `seshinfo_log`, `setup`, `slctn_get`, `slctn_log`, `sql_path_get`, `status_class`, `stat`, `testdatadir_get`

quiet_connect

Quiely connect to the restez database

Description

Quiet version of `restez_connect` for automatic connections.

Usage

`quiet_connect()`

See Also

Other private: `add_rcrd_log`, `cat_line`, `char`, `check_connection`, `cleanup`, `connected`, `connection_get`, `custom_download2`, `custom_download`, `db_sqlngths_get`, `db_sqlngths_log`, `dir_size`, `dwncld_path_get`, `dwncld_rcrd_log`, `entrez_fasta_get`, `entrez_gb_get`, `extract_accession`, `extract_by_patterns`, `extract_clean_sequence`, `extract_definition`, `extract_features`, `extract_inforecpart`, `extract_keywords`, `extract_locus`, `extract_organism`, `extract_seqrecpart`, `extract_sequence`, `extract_version`, `file_download`, `filename_log`, `flatfile_read`, `gb_build2`, `gb_build`, `gb_df_create`, `gb_df_generate`, `gb_sql_add`, `gb_sql_query`, `gbrelease_check`, `gbrelease_get`, `gbrelease_log`, `has_data`, `identify_downloadable_files`, `last_add_get`, `last_dwncld_get`, `last_entry_get`, `latest_genbank_release_notes`, `latest_genbank_release`, `message_missing`, `mock_def`, `mock_gb_df_generate`, `mock_org`, `mock_rec`, `mock_seq`, `predict_datasizes`, `print.status`, `readme_log`, `restez_path_check`, `restez_rl`, `seshinfo_log`, `setup`, `slctn_get`, `slctn_log`, `sql_path_get`, `status_class`, `stat`, `testdatadir_get`

readme_log	<i>Create README in restez_path</i>
------------	-------------------------------------

Description

Write notes for the curious sorts who peruse the restez_path.

Usage

```
readme_log()
```

See Also

Other private: [add_rcrd_log](#), [cat_line](#), [char](#), [check_connection](#), [cleanup](#), [connected](#), [connection_get](#), [custom_download2](#), [custom_download](#), [db_sqlngths_get](#), [db_sqlngths_log](#), [dir_size](#), [dwnld_path_get](#), [dwnld_rcrd_log](#), [entrez_fasta_get](#), [entrez_gb_get](#), [extract_accession](#), [extract_by_patterns](#), [extract_clean_sequence](#), [extract_definition](#), [extract_features](#), [extract_inforecpart](#), [extract_keywords](#), [extract_locus](#), [extract_organism](#), [extract_seqrecpart](#), [extract_sequence](#), [extract_version](#), [file_download](#), [filename_log](#), [flatfile_read](#), [gb_build2](#), [gb_build](#), [gb_df_create](#), [gb_df_generate](#), [gb_sql_add](#), [gb_sql_query](#), [gbrelease_check](#), [gbrelease_get](#), [gbrelease_log](#), [has_data](#), [identify_downloadable_files](#), [last_add_get](#), [last_dwnld_get](#), [last_entry_get](#), [latest_genbank_release_notes](#), [latest_genbank_release](#), [message_missing](#), [mock_def](#), [mock_gb_df_generate](#), [mock_org](#), [mock_rec](#), [mock_seq](#), [predict_datasizes](#), [print.status](#), [quiet_connect](#), [restez_path_check](#), [restez_rl](#), [seshinfo_log](#), [setup](#), [slctn_get](#), [slctn_log](#), [sql_path_get](#), [status_class](#), [stat](#), [testdatadir_get](#)

record	<i>Example GenBank record</i>
--------	-------------------------------

Description

Example GenBank record in text format for demonstration purposes.

Usage

```
data("record")
```

Format

A large character object containing record information and DNA sequence.

Source

<https://www.ncbi.nlm.nih.gov/nuccore/AY952423.1>

References

GenBank

Examples

```
data(record)
cat(record)
```

restez

restez: Create and Query a Local Copy of GenBank in R

Description

The restez package comes with five families of functions: setup, database, get, entrez and internal/private.

Setup functions

These functions allow a user to set the filepath for where the GenBank files should be stored, create connections and verify these settings.

Database functions

These functions download specific parts of GenBank and create the local SQL-like database.

GenBank functions

These functions allow a user to query the local SQL-like database. A user can use an NCBI accession ID to retrieve sequences or whole GenBank records.

Entrez functions

The entrez functions are wrappers to the `entrez_*` functions in the `rentrez` package. e.g the `restez`'s `entrez_fetch` will first try to search the local database, if it fails it will then call `rentrez`'s [entrez_fetch](#) with the same arguments.

Private/internal functions

These functions work behind the scenes to make everything work. If you're curious you can read their documentation using the form `?restez:::functionname`.

restez_connect	<i>Connect to the restez database</i>
----------------	---------------------------------------

Description

Sets a connection to the local database. If database connection cannot be made, an error is returned.

Usage

```
restez_connect()
```

See Also

Other setup: [restez_disconnect](#), [restez_path_get](#), [restez_path_set](#), [restez_path_unset](#), [restez_ready](#), [restez_status](#)

Examples

```
library(restez)
fp <- tempdir()
restez_path_set(filepath = fp)
demo_db_create(n = 5)
restez_connect()
restez_status()
restez_disconnect()
db_delete(everything = TRUE)
# Errors
# restez_status()
```

restez_disconnect	<i>Disconnect from restez database</i>
-------------------	--

Description

Safely disconnect from the restez connection

Usage

```
restez_disconnect()
```

See Also

Other setup: [restez_connect](#), [restez_path_get](#), [restez_path_set](#), [restez_path_unset](#), [restez_ready](#), [restez_status](#)

Examples

```
library(restez)
fp <- tempdir()
restez_path_set(filepath = fp)
demo_db_create(n = 5)
restez_connect()
restez_status()
# always remember to disconnect from a database when you've finished
restez_disconnect()
db_delete(everything = TRUE)
# Errors
# restez_status()
```

restez_path_check	<i>Check restez filepath</i>
-------------------	------------------------------

Description

Raises error if restez path does not exist.

Usage

```
restez_path_check()
```

See Also

Other private: [add_rcrd_log](#), [cat_line](#), [char](#), [check_connection](#), [cleanup](#), [connected](#), [connection_get](#), [custom_download2](#), [custom_download](#), [db_sqlngths_get](#), [db_sqlngths_log](#), [dir_size](#), [dwnld_path_get](#), [dwnld_rcrd_log](#), [entrez_fasta_get](#), [entrez_gb_get](#), [extract_accession](#), [extract_by_patterns](#), [extract_clean_sequence](#), [extract_definition](#), [extract_features](#), [extract_inforecpart](#), [extract_keywords](#), [extract_locus](#), [extract_organism](#), [extract_seqrecpart](#), [extract_sequence](#), [extract_version](#), [file_download](#), [filename_log](#), [flatfile_read](#), [gb_build2](#), [gb_build](#), [gb_df_create](#), [gb_df_generate](#), [gb_sql_add](#), [gb_sql_query](#), [gbrelease_check](#), [gbrelease_get](#), [gbrelease_log](#), [has_data](#), [identify_downloadable_files](#), [last_add_get](#), [last_dwnld_get](#), [last_entry_get](#), [latest_genbank_release_notes](#), [latest_genbank_release](#), [message_missing](#), [mock_def](#), [mock_gb_df_generate](#), [mock_org](#), [mock_rec](#), [mock_seq](#), [predict_datasizes](#), [print.status](#), [quiet_connect](#), [readme_log](#), [restez_rl](#), [seshinfo_log](#), [setup](#), [slctn_get](#), [slctn_log](#), [sql_path_get](#), [status_class](#), [stat](#), [testdatadir_get](#)

restez_path_get	<i>Get restez path</i>
-----------------	------------------------

Description

Return filepath to where the restez database is stored.

Usage

```
restez_path_get()
```

Value

character

See Also

Other setup: [restez_connect](#), [restez_disconnect](#), [restez_path_set](#), [restez_path_unset](#), [restez_ready](#), [restez_status](#)

Examples

```
library(restez)
# set a restez path with a tempdir
restez_path_set(filepath = tempdir())
# check what the set path is
(restez_path_get())
```

restez_path_set	<i>Set restez path</i>
-----------------	------------------------

Description

Specify the filepath for the local GenBank database.

Usage

```
restez_path_set(filepath)
```

Arguments

filepath character, valid filepath to the folder where the database should be stored.

Details

Adds 'restez_path' to options(). In this path the folder 'restez' will be created and all downloaded and database files will be stored there.

See Also

Other setup: [restez_connect](#), [restez_disconnect](#), [restez_path_get](#), [restez_path_unset](#), [restez_ready](#), [restez_status](#)

Examples

```
## Not run:
library(restez)
restez_path_set(filepath = 'path/to/where/you/want/files/to/download')

## End(Not run)
```

restez_path_unset	<i>Unset restez path</i>
-------------------	--------------------------

Description

Set the restez path to NULL

Usage

```
restez_path_unset()
```

Details

Any connected database will be automatically disconnected.

See Also

Other setup: [restez_connect](#), [restez_disconnect](#), [restez_path_get](#), [restez_path_set](#), [restez_ready](#), [restez_status](#)

restez_ready	<i>Is restez ready?</i>
--------------	-------------------------

Description

Returns TRUE if a restez SQL database is available, connected and has data. Use [restez_status\(\)](#) for more information.

Usage

```
restez_ready()
```

Value

Logical

See Also

Other setup: [restez_connect](#), [restez_disconnect](#), [restez_path_get](#), [restez_path_set](#), [restez_path_unset](#), [restez_status](#)

Examples

```
library(restez)
fp <- tempdir()
restez_path_set(filepath = fp)
demo_db_create(n = 5)
restez_connect()
(restez_ready())
restez_disconnect()
db_delete(everything = TRUE)
(restez_ready())
```

restez_rl

Restez readline

Description

Wrapper for base readline.

Usage

```
restez_rl(prompt)
```

Arguments

prompt character, display text

Value

character

See Also

Other private: [add_rcrd_log](#), [cat_line](#), [char](#), [check_connection](#), [cleanup](#), [connected](#), [connection_get](#), [custom_download2](#), [custom_download](#), [db_sqlngths_get](#), [db_sqlngths_log](#), [dir_size](#), [dwnld_path_get](#), [dwnld_rcrd_log](#), [entrez_fasta_get](#), [entrez_gb_get](#), [extract_accession](#), [extract_by_patterns](#), [extract_clean_sequence](#), [extract_definition](#), [extract_features](#), [extract_inforecpart](#), [extract_keywords](#), [extract_locus](#), [extract_organism](#), [extract_seqrecpart](#), [extract_sequence](#), [extract_version](#), [file_download](#), [filename_log](#), [flatfile_read](#), [gb_build2](#), [gb_build](#), [gb_df_create](#), [gb_df_generate](#), [gb_sql_add](#), [gb_sql_query](#), [gbrelease_check](#), [gbrelease_get](#), [gbrelease_log](#), [has_data](#), [identify_downloadable_files](#), [last_add_get](#), [last_dwnld_get](#), [last_entry_get](#), [latest_genbank_release_notes](#), [latest_genbank_release](#), [message_missing](#), [mock_def](#), [mock_gb_df_generate](#), [mock_org](#), [mock_rec](#), [mock_seq](#), [predict_datasizes](#), [print.status](#), [quiet_connect](#), [readme_log](#), [restez_path_check](#), [seshinfo_log](#), [setup](#), [slctn_get](#), [slctn_log](#), [sql_path_get](#), [status_class](#), [stat](#), [testdatadir_get](#)

restez_status	<i>Check restez status</i>
---------------	----------------------------

Description

Report to console current setup status of restez.

Usage

```
restez_status.gb_check = FALSE)
```

Arguments

`gb_check` Check whether last download was from latest GenBank release? Default FALSE.

Details

Always remember to run [restez_connect](#) before running this function. Set `gb_check=TRUE` to see if your downloads are up-to-date.

Value

Status class

See Also

Other setup: [restez_connect](#), [restez_disconnect](#), [restez_path_get](#), [restez_path_set](#), [restez_path_unset](#), [restez_ready](#)

Examples

```
library(restez)
fp <- tempdir()
restez_path_set(filepath = fp)
demo_db_create(n = 5)
restez_connect()
restez_status()
restez_disconnect()
db_delete(everything = TRUE)
# Errors:
# restez_status()
```

seshinfo_log	<i>Log the system session information in restez path</i>
--------------	--

Description

Records the session and system information to file.

Usage

seshinfo_log()

See Also

Other private: [add_rcrd_log](#), [cat_line](#), [char](#), [check_connection](#), [cleanup](#), [connected](#), [connection_get](#), [custom_download2](#), [custom_download](#), [db_sqlngths_get](#), [db_sqlngths_log](#), [dir_size](#), [dwnld_path_get](#), [dwnld_rcrd_log](#), [entrez_fasta_get](#), [entrez_gb_get](#), [extract_accession](#), [extract_by_patterns](#), [extract_clean_sequence](#), [extract_definition](#), [extract_features](#), [extract_inforecpart](#), [extract_keywords](#), [extract_locus](#), [extract_organism](#), [extract_seqrecpart](#), [extract_sequence](#), [extract_version](#), [file_download](#), [filename_log](#), [flatfile_read](#), [gb_build2](#), [gb_build](#), [gb_df_create](#), [gb_df_generate](#), [gb_sql_add](#), [gb_sql_query](#), [gbrelease_check](#), [gbrelease_get](#), [gbrelease_log](#), [has_data](#), [identify_downloadable_files](#), [last_add_get](#), [last_dwnld_get](#), [last_entry_get](#), [latest_genbank_release_notes](#), [latest_genbank_release](#), [message_missing](#), [mock_def](#), [mock_gb_df_generate](#), [mock_org](#), [mock_rec](#), [mock_seq](#), [predict_datasizes](#), [print.status](#), [quiet_connect](#), [readme_log](#), [restez_path_check](#), [restez_rl](#), [setup](#), [slctn_get](#), [slctn_log](#), [sql_path_get](#), [status_class](#), [stat](#), [testdatadir_get](#)

setup	<i>Set up test common test data</i>
-------	-------------------------------------

Description

Creates temporary test folders.

Usage

setup()

See Also

Other private: [add_rcrd_log](#), [cat_line](#), [char](#), [check_connection](#), [cleanup](#), [connected](#), [connection_get](#), [custom_download2](#), [custom_download](#), [db_sqlngths_get](#), [db_sqlngths_log](#), [dir_size](#), [dwnld_path_get](#), [dwnld_rcrd_log](#), [entrez_fasta_get](#), [entrez_gb_get](#), [extract_accession](#), [extract_by_patterns](#), [extract_clean_sequence](#), [extract_definition](#), [extract_features](#), [extract_inforecpart](#), [extract_keywords](#), [extract_locus](#), [extract_organism](#), [extract_seqrecpart](#), [extract_sequence](#), [extract_version](#), [file_download](#), [filename_log](#), [flatfile_read](#), [gb_build2](#), [gb_build](#), [gb_df_create](#), [gb_df_generate](#), [gb_sql_add](#), [gb_sql_query](#), [gbrelease_check](#), [gbrelease_get](#), [gbrelease_log](#),

has_data, identify_downloadable_files, last_add_get, last_dwnld_get, last_entry_get, latest_genbank_release_notes, latest_genbank_release, message_missing, mock_def, mock_gb_df_generate, mock_org, mock_rec, mock_seq, predict_datasizes, print.status, quiet_connect, readme_log, restez_path_check, restez_rl, seshinfo_log, slctn_get, slctn_log, sql_path_get, status_class, stat, testdatadir_get

slctn_get

Retrieve GenBank selections made by user

Description

Returns the selections made by the user.

Usage

```
slctn_get()
```

Details

If no file found, returns empty character vector.

Value

character vector

See Also

Other private: add_rcrd_log, cat_line, char, check_connection, cleanup, connected, connection_get, custom_download2, custom_download, db_sqlngths_get, db_sqlngths_log, dir_size, dwnld_path_get, dwnld_rcrd_log, entrez_fasta_get, entrez_gb_get, extract_accession, extract_by_patterns, extract_clean_sequence, extract_definition, extract_features, extract_inforecpart, extract_keywords, extract_locus, extract_organism, extract_seqrecpart, extract_sequence, extract_version, file_download, filename_log, flatfile_read, gb_build2, gb_build, gb_df_create, gb_df_generate, gb_sql_add, gb_sql_query, gbrelease_check, gbrelease_get, gbrelease_log, has_data, identify_downloadable_files, last_add_get, last_dwnld_get, last_entry_get, latest_genbank_release_notes, latest_genbank_release, message_missing, mock_def, mock_gb_df_generate, mock_org, mock_rec, mock_seq, predict_datasizes, print.status, quiet_connect, readme_log, restez_path_check, restez_rl, seshinfo_log, setup, slctn_log, sql_path_get, status_class, stat, testdatadir_get

slctn_log	<i>Log the GenBank selection made by a user</i>
-----------	---

Description

This function is called whenever a user makes a selection with the `db_download`. It records GenBank numbers selections.

Usage

```
slctn_log(selection)
```

Arguments

selection selected GenBank sequences, named vector

See Also

Other private: `add_rcrd_log`, `cat_line`, `char`, `check_connection`, `cleanup`, `connected`, `connection_get`, `custom_download2`, `custom_download`, `db_sqlngths_get`, `db_sqlngths_log`, `dir_size`, `dwncld_path_get`, `dwncld_rcrd_log`, `entrez_fasta_get`, `entrez_gb_get`, `extract_accession`, `extract_by_patterns`, `extract_clean_sequence`, `extract_definition`, `extract_features`, `extract_inforecpart`, `extract_keywords`, `extract_locus`, `extract_organism`, `extract_seqrecpart`, `extract_sequence`, `extract_version`, `file_download`, `filename_log`, `flatfile_read`, `gb_build2`, `gb_build`, `gb_df_create`, `gb_df_generate`, `gb_sql_add`, `gb_sql_query`, `gbrelease_check`, `gbrelease_get`, `gbrelease_log`, `has_data`, `identify_downloadable_files`, `last_add_get`, `last_dwncld_get`, `last_entry_get`, `latest_genbank_release_notes`, `latest_genbank_release`, `message_missing`, `mock_def`, `mock_gb_df_generate`, `mock_org`, `mock_rec`, `mock_seq`, `predict_datasizes`, `print.status`, `quiet_connect`, `readme_log`, `restez_path_check`, `restez_rl`, `seshinfo_log`, `setup`, `slctn_get`, `sql_path_get`, `status_class`, `stat`, `testdatadir_get`

sql_path_get	<i>Get SQL path</i>
--------------	---------------------

Description

Return path to where SQL database is stored.

Usage

```
sql_path_get()
```

Value

character

See Also

Other private: `add_rcrd_log`, `cat_line`, `char`, `check_connection`, `cleanup`, `connected`, `connection_get`, `custom_download2`, `custom_download`, `db_sqlngths_get`, `db_sqlngths_log`, `dir_size`, `dwncld_path_get`, `dwncld_rcrd_log`, `entrez_fasta_get`, `entrez_gb_get`, `extract_accession`, `extract_by_patterns`, `extract_clean_sequence`, `extract_definition`, `extract_features`, `extract_inforecpart`, `extract_keywords`, `extract_locus`, `extract_organism`, `extract_seqrecpart`, `extract_sequence`, `extract_version`, `file_download`, `filename_log`, `flatfile_read`, `gb_build2`, `gb_build`, `gb_df_create`, `gb_df_generate`, `gb_sql_add`, `gb_sql_query`, `gbrelease_check`, `gbrelease_get`, `gbrelease_log`, `has_data`, `identify_downloadable_files`, `last_add_get`, `last_dwncld_get`, `last_entry_get`, `latest_genbank_release_notes`, `latest_genbank_release`, `message_missing`, `mock_def`, `mock_gb_df_generate`, `mock_org`, `mock_rec`, `mock_seq`, `predict_datasizes`, `print.status`, `quiet_connect`, `readme_log`, `restez_path_check`, `restez_rl`, `seshinfo_log`, `setup`, `slctn_get`, `slctn_log`, `status_class`, `stat`, `testdatadir_get`

stat

Print blue

Description

Print to console blue text to indicate a number/statistic.

Usage

```
stat(...)
```

Arguments

... Any number of text arguments to print, character

Value

coloured character encoding, character

See Also

Other private: `add_rcrd_log`, `cat_line`, `char`, `check_connection`, `cleanup`, `connected`, `connection_get`, `custom_download2`, `custom_download`, `db_sqlngths_get`, `db_sqlngths_log`, `dir_size`, `dwncld_path_get`, `dwncld_rcrd_log`, `entrez_fasta_get`, `entrez_gb_get`, `extract_accession`, `extract_by_patterns`, `extract_clean_sequence`, `extract_definition`, `extract_features`, `extract_inforecpart`, `extract_keywords`, `extract_locus`, `extract_organism`, `extract_seqrecpart`, `extract_sequence`, `extract_version`, `file_download`, `filename_log`, `flatfile_read`, `gb_build2`, `gb_build`, `gb_df_create`, `gb_df_generate`, `gb_sql_add`, `gb_sql_query`, `gbrelease_check`, `gbrelease_get`, `gbrelease_log`, `has_data`, `identify_downloadable_files`, `last_add_get`, `last_dwncld_get`, `last_entry_get`, `latest_genbank_release_notes`, `latest_genbank_release`, `message_missing`, `mock_def`, `mock_gb_df_generate`, `mock_org`, `mock_rec`, `mock_seq`, `predict_datasizes`, `print.status`, `quiet_connect`, `readme_log`, `restez_path_check`, `restez_rl`, `seshinfo_log`, `setup`, `slctn_get`, `slctn_log`, `sql_path_get`, `status_class`, `testdatadir_get`

status_class	<i>Generate a list class for storing status information</i>
--------------	---

Description

Creates a three-part list for holding information on the status of the restez file path.

Usage

```
status_class()
```

Value

Status class

See Also

Other private: [add_rcrd_log](#), [cat_line](#), [char](#), [check_connection](#), [cleanup](#), [connected](#), [connection_get](#), [custom_download2](#), [custom_download](#), [db_sqlngths_get](#), [db_sqlngths_log](#), [dir_size](#), [dwnld_path_get](#), [dwnld_rcrd_log](#), [entrez_fasta_get](#), [entrez_gb_get](#), [extract_accession](#), [extract_by_patterns](#), [extract_clean_sequence](#), [extract_definition](#), [extract_features](#), [extract_inforecpart](#), [extract_keywords](#), [extract_locus](#), [extract_organism](#), [extract_seqrecpart](#), [extract_sequence](#), [extract_version](#), [file_download](#), [filename_log](#), [flatfile_read](#), [gb_build2](#), [gb_build](#), [gb_df_create](#), [gb_df_generate](#), [gb_sql_add](#), [gb_sql_query](#), [gbrelease_check](#), [gbrelease_get](#), [gbrelease_log](#), [has_data](#), [identify_downloadable_files](#), [last_add_get](#), [last_dwnld_get](#), [last_entry_get](#), [latest_genbank_release_notes](#), [latest_genbank_release](#), [message_missing](#), [mock_def](#), [mock_gb_df_generate](#), [mock_org](#), [mock_rec](#), [mock_seq](#), [predict_datasizes](#), [print.status](#), [quiet_connect](#), [readme_log](#), [restez_path_check](#), [restez_rl](#), [seshinfo_log](#), [setup](#), [slctn_get](#), [slctn_log](#), [sql_path_get](#), [stat](#), [testdatadir_get](#)

testdatadir_get	<i>Get test data directory</i>
-----------------	--------------------------------

Description

Get the folder containing test data.

Usage

```
testdatadir_get()
```

See Also

Other private: add_rcrd_log, cat_line, char, check_connection, cleanup, connected, connection_get, custom_download2, custom_download, db_sqlngths_get, db_sqlngths_log, dir_size, dwnld_path_get, dwnld_rcrd_log, entrez_fasta_get, entrez_gb_get, extract_accession, extract_by_patterns, extract_clean_sequence, extract_definition, extract_features, extract_inforecpart, extract_keywords, extract_locus, extract_organism, extract_seqrecpart, extract_sequence, extract_version, file_download, filename_log, flatfile_read, gb_build2, gb_build, gb_df_create, gb_df_generate, gb_sql_add, gb_sql_query, gbrelease_check, gbrelease_get, gbrelease_log, has_data, identify_downloadable_files, last_add_get, last_dwnld_get, last_entry_get, latest_genbank_release_notes, latest_genbank_release, message_missing, mock_def, mock_gb_df_generate, mock_org, mock_rec, mock_seq, predict_datasizes, print.status, quiet_connect, readme_log, restez_path_check, restez_rl, seshinfo_log, setup, slctn_get, slctn_log, sql_path_get, status_class, stat

Index

*Topic **datasets**

- record, 58
- add_rcrd_log, 4, 5–10, 13–17, 20–35, 37, 38, 42, 43, 45, 47–50, 52–58, 61, 64, 66–71
- cat_line, 4, 4, 5–10, 13–17, 20–35, 37, 38, 42, 43, 45, 47–50, 52–58, 61, 64, 66–71
- char, 4, 5, 5, 6–10, 13–17, 20–35, 37, 38, 42, 43, 45, 47–50, 52–58, 61, 64, 66–71
- check_connection, 4, 5, 6, 7–10, 13–17, 20–35, 37, 38, 42, 43, 45, 47–50, 52–58, 61, 64, 66–71
- cleanup, 4–6, 6, 7–10, 13–17, 20–35, 37, 38, 42, 43, 45, 47–50, 52–58, 61, 64, 66–71
- connected, 4–7, 7, 8–10, 13–17, 20–35, 37, 38, 42, 43, 45, 47–50, 52–58, 61, 64, 66–71
- connection_get, 4–7, 8, 9, 10, 13–17, 20–35, 37, 38, 42, 43, 45, 47–50, 52–58, 61, 64, 66–71
- count_db_ids, 8, 11–13, 15, 46, 51
- custom_download, 4–8, 9, 10, 13–17, 20–35, 37, 38, 42, 43, 45, 47–50, 52–58, 61, 64, 66–71
- custom_download2, 4–9, 10, 13–17, 20–35, 37, 38, 42, 43, 45, 47–50, 52–58, 61, 64, 66–71
- db_create, 9, 10, 12, 13, 15, 46, 51
- db_delete, 9, 11, 11, 13, 15, 46, 51
- db_download, 9, 11, 12, 12, 15, 46, 51, 68
- db_sqlngths_get, 4–10, 13, 14–17, 20–35, 37, 38, 42, 43, 45, 47–50, 52–58, 61, 64, 66–71
- db_sqlngths_log, 4–10, 13, 14, 15–17, 20–35, 37, 38, 42, 43, 45, 47–50, 52–58, 61, 64, 66–71
- demo_db_create, 9, 11–13, 14, 46, 51
- dir_size, 4–10, 13, 14, 15, 16, 17, 20–35, 37, 38, 42, 43, 45, 47–50, 52–58, 61, 64, 66–71
- download, 9
- dwnld_path_get, 4–10, 13–15, 16, 17, 20–35, 37, 38, 42, 43, 45, 47–50, 52–58, 61, 64, 66–71
- dwnld_rcrd_log, 4–10, 13–16, 16, 17, 20–35, 37, 38, 42, 43, 45, 47–50, 52–58, 61, 64, 66–71
- entrez_fasta_get, 4–10, 13–17, 17, 20–35, 37, 38, 42, 43, 45, 47–50, 52–58, 61, 64, 66–71
- entrez_fetch, 18, 18, 39, 41, 59
- entrez_gb_get, 4–10, 13–17, 19, 20–35, 37, 38, 42, 43, 45, 47–50, 52–58, 61, 64, 66–71
- extract_accession, 4–10, 13–17, 20, 20, 21–35, 37, 38, 42, 43, 45, 47–50, 52–58, 61, 64, 66–71
- extract_by_patterns, 4–10, 13–17, 20, 21, 22–35, 37, 38, 42, 43, 45, 47–50, 52–58, 61, 64, 66–71
- extract_clean_sequence, 4–10, 13–17, 20, 21, 22, 23–35, 37, 38, 42, 43, 45, 47–50, 52–58, 61, 64, 66–71
- extract_definition, 4–10, 13–17, 20–22, 23, 24–35, 37, 38, 42, 43, 45, 47–50, 52–58, 61, 64, 66–71
- extract_features, 4–10, 13–17, 20–23, 23, 25–35, 37, 38, 42, 43, 45, 47–50, 52–58, 61, 64, 66–71
- extract_inforecpart, 4–10, 13–17, 20–24, 24, 25–35, 37, 38, 42, 43, 45, 47–50, 52–58, 61, 64, 66–71
- extract_keywords, 4–10, 13–17, 20–25, 25, 26–35, 37, 38, 42, 43, 45, 47–50, 52–58, 61, 64, 66–71

- `extract_locus`, 4–10, 13–17, 20–25, 26, 27–35, 37, 38, 42, 43, 45, 47–50, 52–58, 61, 64, 66–71
- `extract_organism`, 4–10, 13–17, 20–26, 27, 28–35, 37, 38, 42, 43, 45, 47–50, 52–58, 61, 64, 66–71
- `extract_seqrepart`, 4–10, 13–17, 20–27, 27, 29–35, 37, 38, 42, 43, 45, 47–50, 52–58, 61, 64, 66–71
- `extract_sequence`, 4–10, 13–17, 20–28, 28, 29–35, 37, 38, 42, 43, 45, 47–50, 52–58, 61, 64, 66–71
- `extract_version`, 4–10, 13–17, 20–29, 29, 30–35, 37, 38, 42, 43, 45, 47–50, 52–58, 61, 64, 66–71
- `file_download`, 4–10, 13–17, 20–30, 30, 31–35, 37, 38, 42, 43, 45, 47–50, 52–58, 61, 64, 66–71
- `filename_log`, 4–10, 13–17, 20–29, 30, 31–35, 37, 38, 42, 43, 45, 47–50, 52–58, 61, 64, 66–71
- `flatfile_read`, 4–10, 13–17, 20–31, 31, 32–35, 37, 38, 42, 43, 45, 47–50, 52–58, 61, 64, 66–71
- `gb_build`, 4–10, 13–17, 20–33, 34, 35, 37, 38, 42, 43, 45, 47–50, 52–58, 61, 64, 66–71
- `gb_build2`, 4–10, 13–17, 20–34, 35, 37, 38, 42, 43, 45, 47–50, 52–58, 61, 64, 66–71
- `gb_definition_get`, 36, 39–42, 44
- `gb_df_create`, 4–10, 13–17, 20–35, 36, 38, 42, 43, 45, 47–50, 52–58, 61, 64, 66–71
- `gb_df_generate`, 4–10, 13–17, 20–35, 37, 37, 42, 43, 45, 47–50, 52–58, 61, 64, 66–71
- `gb_extract`, 38
- `gb_fasta_get`, 18, 36, 39, 40–42, 44
- `gb_organism_get`, 36, 39, 40, 41, 42, 44
- `gb_record_get`, 18, 36, 39, 40, 41, 42, 44
- `gb_sequence_get`, 36, 39–41, 41, 44
- `gb_sql_add`, 4–10, 13–17, 20–35, 37, 38, 42, 43, 45, 47–50, 52–58, 61, 64, 66–71
- `gb_sql_query`, 4–10, 13–17, 20–35, 37, 38, 42, 43, 45, 47–50, 52–58, 61, 64, 66–71
- `gb_version_get`, 36, 39–42, 44
- `gbrelease_check`, 4–10, 13–17, 20–31, 32, 33–35, 37, 38, 42, 43, 45, 47–50, 52–58, 61, 64, 66–71
- `gbrelease_get`, 4–10, 13–17, 20–32, 32, 33–35, 37, 38, 42, 43, 45, 47–50, 52–58, 61, 64, 66–71
- `gbrelease_log`, 4–10, 13–17, 20–33, 33, 34, 35, 37, 38, 42, 43, 45, 47–50, 52–58, 61, 64, 66–71
- `has_data`, 4–10, 13–17, 20–35, 37, 38, 42, 43, 44, 45, 47–50, 52–58, 61, 64, 66–71
- `identify_downloadable_files`, 4–10, 13–17, 20–35, 37, 38, 43, 45, 45, 47–50, 52–58, 61, 64, 66–71
- `is_in_db`, 9, 11–13, 15, 46, 51
- `last_add_get`, 4–10, 13–18, 20–35, 37, 38, 43, 45, 47, 48–50, 52–58, 61, 64, 66–71
- `last_dwnld_get`, 4–10, 13–18, 20–35, 37, 38, 43, 45, 47, 47, 49, 50, 52–58, 61, 64, 66–71
- `last_entry_get`, 4–10, 13–18, 20–35, 37, 38, 43, 45, 47, 48, 48, 49, 50, 52–58, 61, 64, 66–71
- `latest_genbank_release`, 4–10, 13–18, 20–35, 37, 38, 43, 45, 47–49, 49, 50, 52–58, 61, 64, 66–71
- `latest_genbank_release_notes`, 4–10, 13–18, 20–35, 37, 38, 43, 45, 47–49, 50, 52–58, 61, 64, 66–71
- `list_db_ids`, 9, 11–13, 15, 46, 50
- `message_missing`, 4–8, 10, 13–18, 20–35, 37, 38, 43, 45, 47–50, 51, 52–58, 61, 64, 66–71
- `mock_def`, 4–8, 10, 13–18, 20–35, 37, 38, 43, 45, 47–50, 52, 52, 53–58, 61, 64, 66–71
- `mock_gb_df_generate`, 4–8, 10, 13–18, 20–35, 37, 38, 43, 45, 47–50, 52, 53, 54–58, 61, 64, 66–71
- `mock_org`, 4–8, 10, 13–18, 20–35, 37, 38, 43, 45, 47–50, 52, 53, 53, 55–58, 61, 64, 66–71

- mock_rec, 4–8, 10, 13–18, 20–35, 37, 38, 43, 45, 47–50, 52–54, 54, 55–58, 61, 64, 66–71
- mock_seq, 4–8, 10, 13–18, 20–35, 37, 38, 43, 45, 47–50, 52–55, 55, 56–58, 61, 64, 66–71
- predict_datasizes, 4–8, 10, 13–18, 20–35, 37, 38, 43, 45, 47–50, 52–55, 56, 57, 58, 61, 64, 66–71
- print.status, 4–8, 10, 13–18, 20–35, 37, 38, 43, 45, 47–50, 52–56, 56, 57, 58, 61, 64, 66–71
- quiet_connect, 4–8, 10, 13–18, 20–35, 37, 38, 43, 45, 47–50, 52–57, 57, 58, 61, 64, 66–71
- readme_log, 4–8, 10, 13–18, 20–35, 37, 38, 43, 45, 47–50, 52–57, 58, 61, 64, 66–71
- record, 58
- restez, 59
- restez-package (restez), 59
- restez_connect, 60, 60, 62–65
- restez_disconnect, 60, 60, 62–65
- restez_path_check, 4–8, 10, 13, 14, 16–18, 20–35, 37, 38, 43, 45, 47–50, 52–58, 61, 64, 66–71
- restez_path_get, 60, 62, 63–65
- restez_path_set, 60, 62, 62, 63–65
- restez_path_unset, 60, 62, 63, 63, 64, 65
- restez_ready, 60, 62, 63, 63, 65
- restez_rl, 4–8, 10, 13, 14, 16–18, 20–35, 37, 38, 43, 45–50, 52–58, 61, 64, 66–71
- restez_status, 60, 62–64, 65
- seshinfo_log, 4–8, 10, 13, 14, 16–18, 20–35, 37, 38, 43, 45–50, 52–58, 61, 64, 66, 67–71
- setup, 4–8, 10, 13, 14, 16–18, 20–35, 37, 38, 43, 45–50, 52–58, 61, 64, 66, 66, 67–71
- slctn_get, 4–8, 10, 13, 14, 16–18, 20–35, 37, 38, 43, 45–50, 52–58, 61, 64, 66, 67, 67, 68–71
- slctn_log, 4–8, 10, 13, 14, 16–18, 20–35, 37, 38, 43, 45–50, 52–58, 61, 64, 66, 67, 68, 69–71
- sql_path_get, 4–8, 10, 13, 14, 16–18, 20–35, 37, 38, 43, 45–50, 52–58, 61, 64, 66–68, 68, 69–71
- stat, 4–8, 10, 13, 14, 16–18, 20–35, 37, 38, 43, 45–50, 52–58, 61, 64, 66–69, 69, 70, 71
- status_class, 4–8, 10, 13, 14, 16–18, 20–35, 37, 38, 43, 45–50, 52–58, 61, 64, 66–69, 70, 71
- testdatadir_get, 4–8, 10, 13, 14, 16–18, 20–35, 37, 38, 43, 45–50, 52–58, 61, 64, 66–70, 70