Package ‘restez’

May 9, 2026

Type Package

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

Version 2.1.5

Maintainer Joel H. Nitta <joelnitta@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, https://docs.ropensci.org/restez/

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

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Depends R (>= 3.3.0)

Imports utils, rentrez, DBI (>= 1.0.0), curl, cli, crayon, stringi, duckdb, fs, assertthat, ape

Suggests sessioninfo, testthat, knitr, R.utils, rmarkdown, mockery

RoxygenNote 7.3.2

NeedsCompilation no

Author Joel H. Nitta [aut, cre] (ORCID: <https://orcid.org/0000-0003-4719-7472>), Dom Bennett [aut] (ORCID: <https://orcid.org/0000-0003-2722-1359>)

Repository CRAN

Date/Publication 2025-03-07 00:00:02 UTC

Contents

`add_rcrd_log` *Log files added to the SQL database in the restez path*

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(fl)
```

Arguments

`fl`	filename, character

See Also

Other private: `cat_line()`, `char()`, `check_connection()`, `cleanup()`, `connected()`, `connection_get()`, `db_download_intern()`, `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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `testdatadir_get()`

`cat_line` *Cat lines*

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()`, `db_download_intern()`, `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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `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()`, `db_download_intern()`, `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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `testdatadir_get()`

`check_connection` *Helper function to test if a stable internet connection can be established.*

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()`, `db_download_intern()`, `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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `testdatadir_get()`

`cleanup` *Clean up test data*

Description

Removes all temporary test data created.

Usage

```
cleanup()
```

See Also

Other private: `add_rcrd_log()`, `cat_line()`, `char()`, `check_connection()`, `connected()`, `connection_get()`, `db_download_intern()`, `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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `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()`, `db_download_intern()`, `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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `testdatadir_get()`

`connection_get` *Retrieve restez connection*

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()`, `db_download_intern()`, `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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `testdatadir_get()`

`count_db_ids` *Return the number of ids*

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)
(count_db_ids())

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

`db_create` *Create new NCBI database*

Description

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

Usage

```
db_create(
  db_type = "nucleotide",
  min_length = 0,
  max_length = NULL,
  acc_filter = NULL,
  invert = FALSE,
  alt_restez_path = NULL,
  scan = FALSE
)
```

Arguments

`db_type`	character, database type
`min_length`	Minimum sequence length, default 0.
`max_length`	Maximum sequence length, default NULL.
`acc_filter`	Character vector; accessions to include or exclude from the database as specified by `invert`.
`invert`	Logical vector of length 1; if TRUE, accessions in `acc_filter` will be excluded from the database; if FALSE, only accessions in `acc_filter` will be included in the database. Default FALSE.
`alt_restez_path`	Alternative restez path if you would like to use the downloads from a different restez path.
`scan`	Logical vector of length 1; should the sequence file be scanned for accessions in `acc_filter` prior to processing? Requires zgrep to be installed (so does not work on Windows). Only used if `acc_filter` is not NULL and `invert` is FALSE. Default FALSE.

Details

All .seq.gz files are added to the database by default. A user can specify minimum/maximum sequence lengths or accession numbers to limit the sequences to be added to the database – smaller databases are faster to search. The final selection of sequences is the result of applying all filters (`acc_filter`, `min_length`, `max_length`) in combination.

The `scan` option can decrease the time needed to build a database if only a small number of sequences should be written to the database compared to the number of the sequences downloaded from GenBank; i.e., if many of the files downloaded from GenBank do not contain any sequences that should be written to the database. When set to TRUE, if a file does not contain any of the accessions in `acc_filter`, further processing of that file will be skipped and none of the sequences it contains will be added to the database.

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:
# Example of general usage
library(restez)
restez_path_set(filepath = 'path/for/downloads/and/database')
db_download()
db_create()

# Example of using `acc_filter`
#
# Download files to temporary directory
temp_dir <- paste0(tempdir(), "/restez", collapse = "")
dir.create(temp_dir)
restez_path_set(filepath = temp_dir)
# Choose GenBank domain 20 ('unannotated'), the smallest
db_download(preselection = 20)
# Only include three accessions in database
db_create(
  acc_filter = c("AF000122", "AF000123", "AF000124")
)
list_db_ids()
db_delete()
unlink(temp_dir)

## 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.

Usage

```
db_download(
  db = "nucleotide",
  overwrite = FALSE,
  preselection = NULL,
  max_tries = 1
)
```

Arguments

`db`	Database type, only 'nucleotide' currently available.
`overwrite`	T/F, overwrite pre-existing downloaded files?
`preselection`	Character vector of length 1; GenBank domains to download. If not specified (default), a menu will be provided for selection. To specify, provide either a single number or a single character string of numbers separated by spaces, e.g. "19 20" for 'Phage' (19) and 'Unannotated' (20).
`max_tries`	Numeric vector of length 1; maximum number of times to attempt downloading database (default 1).

Details

In default mode, the user interactively selects the parts (i.e., "domains") of GenBank to download (e.g. primates, plants, bacteria ...). Alternatively, the selected domains can be provided as a character string to `preselection`.

The `max_tries` argument is useful for large databases that may otherwise fail due to periodic lapses in internet connectivity. This value can be set to `Inf` to continuously try until the database download succeeds (not recommended if you do not have an internet connection!).

Value

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

See Also

`ncbi_acc_get()`

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_download_intern` *Download database (internal version)*

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 ...). This is an internal function so the download can be wrapped in `while()` to enable persistent downloading.

Usage

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

Arguments

`db`	Database type, only 'nucleotide' currently available.
`overwrite`	T/F, overwrite pre-existing downloaded files?
`preselection`	Character vector of length 1; GenBank domains to download. If not specified (default), a menu will be provided for selection. To specify, provide either a single number or a single character string of numbers separated by spaces, e.g. "19 20" for 'Phage' (19) and 'Unannotated' (20).

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 private: `add_rcrd_log()`, `cat_line()`, `char()`, `check_connection()`, `cleanup()`, `connected()`, `connection_get()`, `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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `testdatadir_get()`

`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()`, `db_download_intern()`, `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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `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()`, `db_download_intern()`, `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_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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `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)
# 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()`, `db_download_intern()`, `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_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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `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()`, `db_download_intern()`, `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_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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `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(fl)
```

Arguments

`fl`	file name, character

See Also

Other private: `add_rcrd_log()`, `cat_line()`, `char()`, `check_connection()`, `cleanup()`, `connected()`, `connection_get()`, `db_download_intern()`, `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_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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `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()`, `db_download_intern()`, `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_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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `testdatadir_get()`

`entrez_fetch` *Entrez fetch*

Description

Wrapper for rentrez::entrez_fetch.

Usage

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

Arguments

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

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 'uilist' 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

`rentrez::entrez_fetch()`

Examples

```
library(restez)
restez_path_set(tempdir())
demo_db_create(n = 5)
# 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

`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()`, `db_download_intern()`, `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_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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `testdatadir_get()`

`extract_accession` *Extract accession*

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()`, `db_download_intern()`, `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_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`(), `latest_genbank_release_notes`(), `message_missing`(), `mock_def`(), `mock_gb_df_generate`(), `mock_org`(), `mock_rec`(), `mock_seq`(), `predict_datasizes`(), `readme_log`(), `restez_connect`(), `restez_disconnect`(), `restez_path_check`(), `restez_rl`(), `search_gz`(), `seshinfo_log`(), `setup`(), `slctn_get`(), `slctn_log`(), `sql_path_get`(), `stat`(), `status_class`(), `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`(), `db_download_intern`(), `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_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`(), `latest_genbank_release_notes`(), `message_missing`(), `mock_def`(), `mock_gb_df_generate`(), `mock_org`(), `mock_rec`(), `mock_seq`(), `predict_datasizes`(), `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `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, max_len = 1e+08)
```

Arguments

`seqrecpart`	Sequence part of a GenBank record, character
`max_len`	Number: maximum number of characters allowed in a single record before splitting the record into parts. Does not affect output, but only internal calculations, so generally should not be changed. Default = 1e8.

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()`, `db_download_intern()`, `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_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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `testdatadir_get()`

`extract_definition` *Extract definition*

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()`, `db_download_intern()`, `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_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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `testdatadir_get()`

`extract_features` *Extract features*

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()`, `db_download_intern()`, `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_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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `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()`, `db_download_intern()`, `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_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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `testdatadir_get()`

`extract_keywords` *Extract keywords*

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()`, `db_download_intern()`, `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_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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `testdatadir_get()`

`extract_locus` *Extract locus*

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()`, `db_download_intern()`, `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_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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `testdatadir_get()`

`extract_organism` *Extract organism*

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()`, `db_download_intern()`, `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_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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `testdatadir_get()`

`extract_seqrecpart` *Extract the sequence record part*

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()`, `db_download_intern()`, `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_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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `testdatadir_get()`

`extract_sequence` *Extract sequence*

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()`, `db_download_intern()`, `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_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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `testdatadir_get()`

`extract_version` *Extract version*

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()`, `db_download_intern()`, `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_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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `testdatadir_get()`

`filename_log` *Write filenames to log files*

Description

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

Usage

```
filename_log(fl, fp)
```

Arguments

`fl`	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()`, `db_download_intern()`, `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_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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `testdatadir_get()`

`file_download` *Download a file*

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(fl, overwrite = FALSE)
```

Arguments

`fl`	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()`, `db_download_intern()`, `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()`, `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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `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()`, `db_download_intern()`, `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()`, `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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `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()`, `db_download_intern()`, `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_get()`, `gbrelease_log()`, `has_data()`, `identify_downloadable_files()`, `last_add_get()`, `last_dwnld_get()`, `last_entry_get()`, `latest_genbank_release()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `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()`, `db_download_intern()`, `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_log()`, `has_data()`, `identify_downloadable_files()`, `last_add_get()`, `last_dwnld_get()`, `last_entry_get()`, `latest_genbank_release()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `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 if 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()`, `db_download_intern()`, `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()`, `has_data()`, `identify_downloadable_files()`, `last_add_get()`, `last_dwnld_get()`, `last_entry_get()`, `latest_genbank_release()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `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,
  acc_filter = NULL,
  invert = FALSE,
  scan = FALSE
)
```

Arguments

`dpth`	Download path (where seq_files are stored)
`seq_files`	.seq.tar seq file names
`max_length`	Maximum sequence length, default NULL.
`min_length`	Minimum sequence length, default 0.

`acc_filter`	Character vector; accessions to include or exclude from the database as specified by `invert`.
`invert`	Logical vector of length 1; if TRUE, accessions in `acc_filter` will be excluded from the database; if FALSE, only accessions in `acc_filter` will be included in the database. Default FALSE.
`scan`	Logical vector of length 1; should the sequence file be scanned for accessions in `acc_filter` prior to processing? Requires zgrep to be installed (so does not work on Windows). Only used if `acc_filter` is not NULL and `invert` is FALSE. Default FALSE.

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()`, `db_download_intern()`, `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_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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `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

`ncbi_acc_get()`

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)
(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()`, `db_download_intern()`, `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_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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `testdatadir_get()`

`gb_df_generate` *Generate GenBank records data.frame*

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,
  acc_filter = NULL,
  invert = FALSE
)
```

Arguments

`records`	character, vector of GenBank records in text format
`min_length`	Minimum sequence length, default 0.
`max_length`	Maximum sequence length, default NULL.
`acc_filter`	Character vector; accessions to include or exclude from the database as specified by `invert`.
`invert`	Logical vector of length 1; if TRUE, accessions in `acc_filter` will be excluded from the database; if FALSE, only accessions in `acc_filter` will be included in the database. Default FALSE.

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 ?charToRaw, in order to save on RAM. The raw_record contains the entire GenBank record in text format.

Use `acc_filter` and max and min sequence lengths to minimize 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. The final selection of sequences is the result of applying all filters (`acc_filter`, `min_length`, `max_length`) in combination.

Value

data.frame, or NULL if no records pass filters

See Also

Other private: `add_rcrd_log()`, `cat_line()`, `char()`, `check_connection()`, `cleanup()`, `connected()`, `connection_get()`, `db_download_intern()`, `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_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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `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 `rentrez::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

`ncbi_acc_get()`

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)
(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` *Get organism from GenBank*

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

`ncbi_acc_get()`

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)
(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` *Get record from GenBank*

Description

Return the entire GenBank record for an accession ID. Equivalent to `rettype='gb'` in `rentrez::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

`ncbi_acc_get()`

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)
(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` *Get sequence from GenBank*

Description

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

Usage

```
gb_sequence_get(id, dnabin = FALSE)
```

Arguments

`id`	character, sequence accession ID(s)
`dnabin`	Logical vector of length 1; should the sequences be returned using the bit-level coding scheme of the ape package? Default FALSE.

Details

For more information about the `dnabin` format, see `ape::DNAbin()`.

Value

named vector of sequences, if no results found NULL

See Also

`ncbi_acc_get()`

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)
(seq <- gb_sequence_get(id = 'demo_1'))
(seqs <- gb_sequence_get(id = c('demo_1', 'demo_2')))
(fasta_dnabin <- gb_sequence_get(id = 'demo_1', dnabin = TRUE))

# 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()`, `db_download_intern()`, `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_query()`, `gbrelease_check()`, `gbrelease_get()`, `gbrelease_log()`, `has_data()`, `identify_downloadable_files()`, `last_add_get()`, `last_dwnld_get()`, `last_entry_get()`, `latest_genbank_release()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `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()`, `db_download_intern()`, `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()`, `gbrelease_check()`, `gbrelease_get()`, `gbrelease_log()`, `has_data()`, `identify_downloadable_files()`, `last_add_get()`, `last_dwnld_get()`, `last_entry_get()`, `latest_genbank_release()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `testdatadir_get()`

`gb_version_get` *Get version from GenBank*

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

`ncbi_acc_get()`

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)
(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` *Does the connected database have data?*

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()`, `db_download_intern()`, `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()`, `identify_downloadable_files()`, `last_add_get()`, `last_dwnld_get()`, `last_entry_get()`, `latest_genbank_release()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `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()`, `db_download_intern()`, `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()`, `last_add_get()`, `last_dwnld_get()`, `last_entry_get()`, `latest_genbank_release()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `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)
# 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` *Return date and time of the last added sequence*

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()`, `db_download_intern()`, `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_dwnld_get()`, `last_entry_get()`, `latest_genbank_release()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `testdatadir_get()`

`last_dwnld_get` *Return date and time of the last download*

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()`, `db_download_intern()`, `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_entry_get()`, `latest_genbank_release()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `testdatadir_get()`

`last_entry_get` *Return the last entry*

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()`, `db_download_intern()`, `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()`, `latest_genbank_release()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `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()`, `db_download_intern()`, `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()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `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()`, `db_download_intern()`, `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()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `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
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()`, `db_download_intern()`, `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()`, `latest_genbank_release_notes()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `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()`, `db_download_intern()`, `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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `testdatadir_get()`

`mock_gb_df_generate` *Generate mock GenBank records data.frame*

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()`, `db_download_intern()`, `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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `testdatadir_get()`

`mock_org` *Mock org*

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()`, `db_download_intern()`, `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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `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()`, `db_download_intern()`, `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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `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()`, `db_download_intern()`, `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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `testdatadir_get()`

`ncbi_acc_get` *Get accession numbers by querying NCBI GenBank*

Description

The query string can be formatted using GenBank advanced query terms to obtain accession numbers corresponding to a specific set of criteria.

Usage

```
ncbi_acc_get(query, strict = TRUE, drop_ver = TRUE)
```

Arguments

`query`	Character vector of length 1; query string to search GenBank.
`strict`	Logical vector of length 1; should an error be issued if the number of unique accessions retrieved does not match the number of hits from GenBank? Default TRUE.
`drop_ver`	Logical vector of length 1; should the version part of the accession number (e.g., '.1' in 'AB001538.1') be dropped? Default TRUE.

Details

Note this queries NCBI GenBank, not the local database generated with restez.

It can be used either to restrict the accessions used to construct the local database (`acc_filter` argument of `db_create()`) or to specify accessions to read from the local database (`id` argument of `gb_fasta_get()` and other gb_*_get() functions).

Value

Character vector; accession numbers resulting from query.

See Also

`db_create()`, `gb_fasta_get()`

Examples

```
## Not run:
  # requires an internet connection
  cmin_accs <- ncbi_acc_get("Crepidomanes minutum")
  length(cmin_accs)
  head(cmin_accs)

## End(Not run)
```

`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()`, `db_download_intern()`, `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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `testdatadir_get()`

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

Description

Prints to screen the three sections of the status class. Not meant to be used interactively.

Usage

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

Arguments

`x`	Status object
`...`	Other arguments (not used by this function)

`readme_log` *Create README in restez_path*

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()`, `db_download_intern()`, `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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `testdatadir_get()`

`record` *Example GenBank record*

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_connect` *Connect to the restez database*

Description

Sets a connection to the local database.

Usage

```
restez_connect(read_only = FALSE)
```

Arguments

`read_only`	Logical; should the connection be made in read-only mode? Read-only mode is required for multiple R processes to access the database simultaneously. Default FALSE.

See Also

Other private: `add_rcrd_log()`, `cat_line()`, `char()`, `check_connection()`, `cleanup()`, `connected()`, `connection_get()`, `db_download_intern()`, `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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `testdatadir_get()`

`restez_disconnect` *Disconnect from restez database*

Description

Safely disconnect from the restez connection

Usage

```
restez_disconnect()
```

See Also

Other private: `add_rcrd_log()`, `cat_line()`, `char()`, `check_connection()`, `cleanup()`, `connected()`, `connection_get()`, `db_download_intern()`, `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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `testdatadir_get()`

`restez_path_check` *Check restez filepath*

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()`, `db_download_intern()`, `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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `testdatadir_get()`

`restez_path_get` *Get restez path*

Description

Return filepath to where the restez database is stored.

Usage

```
restez_path_get()
```

Value

character

See Also

Other setup: `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` *Set restez path*

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_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` *Unset restez path*

Description

Set the restez path to NULL

Usage

```
restez_path_unset()
```

See Also

Other setup: `restez_path_get()`, `restez_path_set()`, `restez_ready()`, `restez_status()`

`restez_ready` *Is restez ready?*

Description

Returns TRUE if a restez SQL database is available. Use restez_status() for more information.

Usage

```
restez_ready()
```

Value

Logical

See Also

Other setup: `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_ready())
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()`, `db_download_intern()`, `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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `testdatadir_get()`

`restez_status` *Check restez status*

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

Set gb_check=TRUE to see if your downloads are up-to-date.

Value

Status class

See Also

Other setup: `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_status()
db_delete(everything = TRUE)
# Errors:
# restez_status()
```

`search_gz` *Scan a gzipped file for text*

Description

Scans a zipped file for text strings and returns TRUE if any are present.

Usage

```
search_gz(terms, path)
```

Arguments

`terms`	Character vector; search terms (most likely GenBank accession numbers)
`path`	Path to the gzipped file to scan

Value

Logical

See Also

Other private: `add_rcrd_log()`, `cat_line()`, `char()`, `check_connection()`, `cleanup()`, `connected()`, `connection_get()`, `db_download_intern()`, `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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `testdatadir_get()`

`seshinfo_log` *Log the system session information in restez path*

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()`, `db_download_intern()`, `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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `testdatadir_get()`

`setup` *Set up test common test data*

Description

Creates temporary test folders.

Usage

```
setup()
```

See Also

Other private: `add_rcrd_log()`, `cat_line()`, `char()`, `check_connection()`, `cleanup()`, `connected()`, `connection_get()`, `db_download_intern()`, `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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `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()`, `db_download_intern()`, `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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`, `testdatadir_get()`

`slctn_log` *Log the GenBank selection made by a user*

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()`, `db_download_intern()`, `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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `sql_path_get()`, `stat()`, `status_class()`, `testdatadir_get()`

`sql_path_get` *Get SQL path*

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()`, `db_download_intern()`, `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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `stat()`, `status_class()`, `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()`, `db_download_intern()`, `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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `status_class()`, `testdatadir_get()`

`status_class` *Generate a list class for storing status information*

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()`, `db_download_intern()`, `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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `testdatadir_get()`

`testdatadir_get` *Get test data directory*

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()`, `db_download_intern()`, `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()`, `latest_genbank_release_notes()`, `message_missing()`, `mock_def()`, `mock_gb_df_generate()`, `mock_org()`, `mock_rec()`, `mock_seq()`, `predict_datasizes()`, `readme_log()`, `restez_connect()`, `restez_disconnect()`, `restez_path_check()`, `restez_rl()`, `search_gz()`, `seshinfo_log()`, `setup()`, `slctn_get()`, `slctn_log()`, `sql_path_get()`, `stat()`, `status_class()`

Index