

Package ‘bedrockbio’

May 28, 2026

Title Open-Access Computational Biology Datasets

Version 1.4.0

Description Efficiently access the 'Bedrock Bio' library of open-access computational biology datasets. Lazily query datasets backed by 'DuckDB' and 'Apache Iceberg', with support for predicate pushdown and column projection to the cloud storage backend. This enables quick, iterative access to otherwise massive, unwieldy datasets without downloading them in full. See <<https://bedrock.bio>> for available datasets and documentation.

Language en-US

License GPL (>= 3)

URL <https://bedrock.bio>, <https://github.com/bedrock-bio/bedrock-bio>

BugReports <https://github.com/bedrock-bio/bedrock-bio/issues>

Depends R (>= 4.1)

Imports curl, DBI, dbplyr, dplyr, duckdb, jsonlite

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

OS_type unix

Encoding UTF-8

RoxygenNote 7.3.3

NeedsCompilation no

Author Liam Abbott [aut, cre, cph]

Maintainer Liam Abbott <liam@bedrock.bio>

Repository CRAN

Date/Publication 2026-05-28 13:50:08 UTC

Contents

describe_namespace	2
describe_table	3
list_namespaces	3
list_tables	4
load_table	4
Index	6

describe_namespace	<i>Describe a namespace's metadata, citation, license, and tables</i>
--------------------	---

Description

Describe a namespace's metadata, citation, license, and tables

Usage

```
describe_namespace(name)
```

Arguments

name	Namespace identifier (e.g., "ukb_ppp")
------	--

Value

A named list with id, name, description, source_url, license, instructions, citation, and tables (character vector of fully-qualified table identifiers). Use describe_table() for per-table details.

Examples

```
## Not run:
library(bedrockbio)
info <- describe_namespace("ukb_ppp")
info$tables

## End(Not run)
```

describe_table	<i>Describe a table's metadata, citation, and columns</i>
----------------	---

Description

Describe a table's metadata, citation, and columns

Usage

```
describe_table(name)
```

Arguments

name	Table identifier (e.g., "ukb_ppp.pqtls")
------	--

Value

A named list with name, description, citation, source_url, license, partition_by, sort_by, and columns.

Examples

```
## Not run:  
library(bedrockbio)  
info <- describe_table("ukb_ppp.pqtls")  
info$name  
  
## End(Not run)
```

list_namespaces	<i>List available namespaces (data sources) in the Bedrock Bio library</i>
-----------------	--

Description

List available namespaces (data sources) in the Bedrock Bio library

Usage

```
list_namespaces()
```

Value

A character vector of namespace identifiers

Examples

```
## Not run:  
library(bedrockbio)  
list_namespaces()  
  
## End(Not run)
```

list_tables	<i>List available tables in the Bedrock Bio library</i>
-------------	---

Description

List available tables in the Bedrock Bio library

Usage

```
list_tables()
```

Value

A character vector of table identifiers

Examples

```
## Not run:  
library(bedrockbio)  
list_tables()  
  
## End(Not run)
```

load_table	<i>Lazily query a table</i>
------------	-----------------------------

Description

Lazily query a table

Usage

```
load_table(name)
```

Arguments

name Table identifier (e.g., "ukb_ppp.pqtls")

Value

A lazy tbl backed by DuckDB, compatible with dplyr verbs. Use `describe_table()` to see partition columns and per-column allowed values; filter on partition columns for fastest reads.

Examples

```
## Not run:  
library(bedrockbio)  
library(dplyr)  
  
df <- load_table("dbsnp.vcf") |>  
  filter(assembly == "GRCh38", chromosome == "22") |>  
  select(rsid, position, ref_allele, alt_allele) |>  
  head(5) |>  
  collect()  
  
## End(Not run)
```

Index

`describe_namespace`, [2](#)

`describe_table`, [3](#)

`list_namespaces`, [3](#)

`list_tables`, [4](#)

`load_table`, [4](#)