Package ‘ensemblQueryR’

May 23, 2023

Type Package

Title Simple, Fast and Efficient Querying of the 'Ensembl' API

Version 0.1.0

Author Aine Fairbrother-Browne

Maintainer Aine Fairbrother-Browne <ucbtas8@ucl.ac.uk>

Description Suite of tools to enable fast, efficient, high-throughput, R workflow integrable querying of 'Ensembl' REST API endpoints (https://rest.ensembl.org/).

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.2.1

Imports dplyr, httr, jsonlite, magrittr, parallel, purrr, rlang, tibble, tidyr, utils, vroom, xml2

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

NeedsCompilation no

Repository CRAN

Date/Publication 2023-05-23 09:10:05 UTC

R topics documented:

ensemblQueryGetPops ........................................ 2
ensemblQueryLDwithSNPpair ............................. 2
ensemblQueryLDwithSNPpairDataframe ................. 3
ensemblQueryLDwithSNPregion ......................... 4
ensemblQueryLDwithSNPregionDataframe ............... 4
ensemblQueryLDwithSNPwindow ......................... 6
ensemblQueryLDwithSNPwindowDataframe ............... 7
pingEnsembl ............................................. 8

Index 9
**Description**

Function to get list of populations that Ensembl has available to query LD.

**Usage**

```r
ensemblQueryGetPops()
```

**Value**

data.frame of populations.

**Examples**

```r
ensemblQueryGetPops()
```

---

**Description**

Function to query Ensembl LD data with a pair of rsIDs. This function will return r-squared and D’ values for the rsID pair.

**Usage**

```r
ensemblQueryLDwithSNPpair(rsid1, rsid2, pop = "1000GENOMES:phase_3:EUR")
```

**Arguments**

- `pop`: String. Population for which to compute LD. Use `ensemblQueryGetPops()` to retrieve a list of all populations with LD data. Default is 1000GENOMES:phase_3:EUR.

**Value**

A dataframe.
**Examples**

```r
ensemblQueryLDwithSNPpair(
    rsid1="rs6792369",
    rsid2="rs1042779",
    pop="1000GENOMES:phase_3:EUR"
)
```

---

**Description**

‘ensemblQueryLDwithSNPpairDataframe’ applies ‘ensemblQueryLDwithSNPpair’ to a data.frame of rsID pairs.

**Usage**

```r
ensemblQueryLDwithSNPpairDataframe(
    in.table,
    pop = "1000GENOMES:phase_3:EUR",
    cores = 1
)
```

**Arguments**

- **in.table**: data.frame containing SNP pairs. Columns must include ‘rsid1’ for the first member of the pair and ‘rsid2’ for the second member of the pair.
- **pop**: String. Population for which to compute LD. Use ‘ensemblQueryGetPops()’ to retrieve a list of all populations with LD data. Default is 1000GENOMES:phase_3:EUR.
- **cores**: Integer. A value between 1 and 10 is accepted, as this prevents the server returning overload-related errors.

**Value**

A data.frame.

**Examples**

```r
ensemblQueryLDwithSNPpairDataframe(
    in.table=data.frame(rsid1=rep("rs6792369", 10),
                        rsid2=rep("rs1042779", 10)),
    pop="1000GENOMES:phase_3:EUR"
)
```
Function to query Ensembl LD data within a genomic window to get all variant pairs in the specified region and associated LD metrics.

**Usage**

```r
ensemblQueryLDwithSNPregion(chr, start, end, pop = "1000GENOMES:phase_3:EUR")
```

**Arguments**

- `chr` String. Chromosome that the query region is located on.
- `start` String. Base pair that the query region starts at.
- `end` String. Base pair that the query region ends at.
- `pop` String. Population for which to compute LD. Use `ensemblQueryGetPops()` to retrieve a list of all populations with LD data. Default is `1000GENOMES:phase_3:EUR`.

**Value**

A dataframe.

**Examples**

```r
ensemblQueryLDwithSNPregion(
  chr="6",
  start="25837556",
  end="25843455",
  pop="1000GENOMES:phase_3:EUR"
)
```

**Value**

A dataframe.
Description

'ensemblQueryLDwithSNPregionDataframe' applies 'ensemblQueryLDwithSNPregion' to a data.frame of genomic coordinates and returns all variant pairs present in each specified genomic region and their associated LD metrics.

Usage

ensemblQueryLDwithSNPregionDataframe(
  in.table,
  pop = "1000GENOMES:phase_3:EUR",
  cores = 1
)

Arguments

in.table Dataframe containing genomic coordinates. Columns must include 'chr' (the chromosome), 'start' (the starting genomic coordinate) and 'end' (the ending genomic coordinate).

pop String. Population for which to compute LD. Use 'ensemblQueryGetPops()' to retrieve a list of all populations with LD data. Default is 1000GENOMES:phase_3:EUR.

cores Integer. A value between 1 and 10 is accepted, as this prevents the server returning overload-related errors.

Value

A dataframe.

Examples

## Not run:
library(magrittr)

data.frame(
  chr=rep(c("6"), 5),
  start=rep(c("25837556"), 5),
  end=rep(c("25943455"), 5)
) %>%
  ensemblQueryLDwithSNPregionDataframe(
    in.table=.,
    pop="1000GENOMES:phase_3:EUR",
    cores = 1
  )

## End(Not run)
**ensemblQueryLDwithSNPwindow**

*Function to query Ensembl LD data with a single rsID*

**Description**

Function to query Ensembl LD data with a single rsID

**Usage**

```r
ensemblQueryLDwithSNPwindow(
  rsid,
  pop = "1000GENOMES:phase_3:EUR",
  r2 = NA,
  d.prime = NA,
  window.size = NA
)
```

**Arguments**

- **rsid**: String. Variant ID.
- **pop**: String. Population for which to compute LD. Use `ensemblQueryGetPops()` to retrieve a list of all populations with LD data. Default is 1000GENOMES:phase_3:EUR.
- **r2**: Float. Measure of LD. If r-squared is provided only return pairs of variants whose r-squared value is equal to or greater than the value provided.
- **d.prime**: Float. Measure of LD. If D' is provided only return pairs of variants whose D' value is equal to or greater than the value provided.
- **window.size**: Integer. Window size in kb. The maximum allowed value for the window size is 500 kb. LD is computed for the given variant and all variants that are located within the specified window.

**Value**

A dataframe.

**Examples**

```r
ensemblQueryLDwithSNPwindow(rsid="rs3851179", r2=0.8, d.prime=0.8,
                             window.size=500, pop="1000GENOMES:phase_3:EUR")
```
"ensemblQueryLDwithSNPwindowDataframe" applies "ensemblQueryLDwithSNPwindow" to a data.frame of rsIDs.

Description

"ensemblQueryLDwithSNPwindowDataframe" applies "ensemblQueryLDwithSNPwindow" to a data.frame of rsIDs.

Usage

```r
ensemblQueryLDwithSNPwindowDataframe(
  in.table,
  pop = "1000GENOMES:phase_3:EUR",
  r2 = NA,
  d.prime = NA,
  window.size = NA,
  cores = 1
)
```

Arguments

- `in.table`: data.frame containing SNP pairs. Columns must include ‘rsid1’ for the first member of the pair and ‘rsid2’ for the second member of the pair.
- `pop`: String. Population for which to compute LD. Use ‘ensemblQueryGetPops()’ to retrieve a list of all populations with LD data. Default is 1000GENOMES:phase_3:EUR.
- `r2`: Float. Measure of LD. If r-squared is provided only return pairs of variants whose r-squared value is equal to or greater than the value provided.
- `d.prime`: Float. Measure of LD. If D’ is provided only return pairs of variants whose D’ value is equal to or greater than the value provided.
- `window.size`: Integer. Window size in kb. The maximum allowed value for the window size is 500 kb. LD is computed for the given variant and all variants that are located within the specified window.
- `cores`: Integer. A value between 1 and 10 is accepted, as this prevents the server returning overload-related errors.

Value

A dataframe.

Examples

```r
## Not run:
in.table = data.frame(rsid = rep(c("rs7153434","rs1963154","rs12672022",
                                  "rs3852802","rs12324408","rs56346870"), 5))
```
Function to check whether the Ensembl server is up by pinging it.

Usage

pingEnsembl()

Value

Integer. Where 1 is indicative of a successful ping.

Examples

pingEnsembl()
Index

ensemblQueryGetPops, 2
ensemblQueryLDwithSNPpair, 2
ensemblQueryLDwithSNPpairDataframe, 3
ensemblQueryLDwithSNPregion, 4
ensemblQueryLDwithSNPregionDataframe, 4
ensemblQueryLDwithSNPwindow, 6
ensemblQueryLDwithSNPwindowDataframe, 7

pingEnsembl, 8