Package: bugphyzz (via r-universe)

November 19, 2026

Title A harmonized data resource and software for enrichment analysis of microbial physiologies

Version 1.7.20

Description bugphyzz is an electronic database of standardized microbial annotations. It facilitates the creation of microbial signatures based on shared attributes, which are utilized for bug set enrichment analysis. The data also includes annotations imputed with ancestra state reconstruction methods.

biocViews MicrobiomeData, ReproducibleResearch

License Artistic-2.0

Config/testthat/edition 3

Encoding UTF-8

LazyData false

Roxygen list(markdown = TRUE)

VignetteBuilder knitr

Depends R (>= 4.4)

Imports rlang, dplyr, utils, purrr, stringr, tidyr, tidyselect, BiocFileCache, httr2, tools, S4Vectors

Suggests DT, forcats, ggplot2, htmltools, knitr, rmarkdown, sessioninfo, testthat, withr, EnrichmentBrowser, MicrobiomeBenchmarkData, mia, stats, limma, mgsub, methods, readr, crayon, tibble, magrittr, tidytext, BiocStyle

URL https://github.com/waldronlab/bugphyzz

BugReports https://github.com/waldronlab/bugphyzz/issues

Config/roxygen2/version 8.0.0

Config/pak/sysreqs libicu-dev libssl-dev

Repository https://waldronlab.r-universe.dev

Date/Publication 2026-11-19 18:07:42 UTC

RemoteUrl https://github.com/waldronlab/bugphyzz

RemoteRef HEAD

RemoteSha 0d4e04eac88aeb4d7d297c060d3d11423f50a541

Contents

getTaxonSignatures 2
importBugphyzz 3
makeSignatures 4
physiologies 6
showPhys 7

Index **8**

`getTaxonSignatures` *Get Taxon Signatures*

Description

`getTaxonSignatures` returns the names of all of the signatures associated with a particular taxon. More details can be found in the main bugphyzz vignette; please run `browseVignettes("bugphyzz")`.

Usage

```
getTaxonSignatures(tax, bp, ...)
```

Arguments

`tax`	A valid NCBI ID or taxon name. If taxon name is used, the argument taxIdType = "Taxon_name" must also be used.
`bp`	List of data.frames imported with `importBugphyzz`.
`...`	Arguments passed to `makeSignatures`.

Value

A character vector with the names of the signatures for a taxon.

Examples

```
taxid <- "562"
taxonName <- "Escherichia coli"
bp <- importBugphyzz()
sig_names_1 <- getTaxonSignatures(taxid, bp)
sig_names_2 <- getTaxonSignatures(taxonName, bp, taxIdType = "Taxon_name")
```

`importBugphyzz` *Import bugphyzz*

Description

`importBugphyzz` imports bugphyzz annotations as a list of tidy data.frames. To learn more about the structure of the data.frames please check the bugphyzz vignette with `browseVignettes("bugphyzz")` or ‘vignette("bugphyzz", "bugphyzz").

Usage

```
importBugphyzz(
  version = "10.5281/zenodo.24166279",
  forceDownload = FALSE,
  kappaThreshold = 0.7,
  r2Threshold = 0.5,
  asrScoreThreshold = 0.5,
  excludeRarely,
  v = NULL
)
```

Arguments

`version`	Character string indicating the version. Default is the Zenodo release the package was built against (v1.2.0, November 2026). Options: a Zenodo DOI, a GitHub commit hash of bugphyzzExports, or "devel" for the weekly export.
`forceDownload`	Logical value. Force a fresh download of the data or use the one stored in the cache (if available). Default is FALSE.
`kappaThreshold`	Minimum cross-validated Cohen’s kappa (lower 95% bound) for keeping the imputed annotations of a discrete attribute. Default 0.7 (see details).
`r2Threshold`	Minimum cross-validated R-squared (lower 95% bound) for keeping the imputed annotations of a numeric attribute. Default 0.5 (see details).
`asrScoreThreshold`	Minimum `Score` of an imputed annotation. Default 0.5 (see details).
`excludeRarely`	Deprecated. `excludeRarely = FALSE` is equivalent to `asrScoreThreshold = 0`; curated annotations with Frequency "rarely" or "never" are always excluded.
`v`	Deprecated. A single threshold applied to both `kappaThreshold` and `r2Threshold`.

Details

Data structure:

The data structure of the data.frames imported with `importBugphyzz` are detailed in the main vignette. Please run `browseVignettes("bugphyzz")`.

Imputed annotations (`kappaThreshold`, `r2Threshold`):

Data imported with `importBugphyzz` includes annotations imputed through ancestral state reconstruction (ASR), with Evidence "asr". An attribute's imputed annotations are kept all together or not at all, based on a cross-validation of the reconstruction (https://github.com/waldronlab/taxPProValidation): Cohen's kappa for discrete attributes (linearly weighted for attributes with ordered states, such as aerophilicity) and R-squared for numeric attributes, with held-out taxa weighted to match the phylogenetic distance of the taxa to be imputed. The threshold applies to the lower bound of the 95% confidence interval of the score. The defaults are the thresholds applied by the export (bugphyzzExports ADR 0006). Curated annotations are never filtered by these thresholds.

Score of imputed annotations (`asrScoreThreshold`):

For discrete attributes, `Score` is the posterior probability of the imputed state. For numeric attributes it is a step function of the distance to the nearest annotated taxon (NSTI): 0.9 up to 0.05, 0.5 up to 0.15, 0.1 up to 0.2 and 0 beyond. Raise `asrScoreThreshold` to keep fewer, more reliable imputed annotations.

Thresholds below the defaults:

The export removes imputed annotations below the default thresholds and keeps them in a separate file. Setting any threshold below its default makes `importBugphyzz` read that file as well, for versions that have it.

Frequency:

Curated annotations with Frequency "rarely" or "never" are excluded. Use `physiologies()` for all curated data.

Sources:

By default, the datasets imported with the `importBugphuzz` function will always return a shortened version of the source. Please use vigette("sources", "bugphyz") to see the full sources.

Value

A list of tidy data frames.

Examples

```
bp <- importBugphyzz()
names(bp)
```

`makeSignatures` *Make signatures*

Description

`makeSignatures` Creates signatures for a list of bug signatures from a tidy data.frame imported through the `importBugphyzz` function. Please run `browseVignettes("bugphyz")` for detailed examples.

Usage

```
makeSignatures(
  dat,
  taxIdType = c("NCBI_ID", "Taxon_name"),
 taxLevel = c("mixed", "superkingdom", "phylum", "class", "order", "family", "genus",
    "species", "strain"),
  evidence = c("exp", "igc", "tas", "nas", "tax", "asr"),
  frequency = c("always", "usually", "sometimes", "unknown"),
  minSize = 10,
  min = NULL,
  max = NULL
)
```

Arguments

dat	A data.frame.
taxIdType	A character string. Valid options: NCBI_ID, Taxon_name.
taxLevel	A character vector. Taxonomic rank. Valid options: superkingdom, kingdom, phylum, class, order, family, genus, species, strain. They can be combined. "mixed" is equivalent to select all valid ranks.
evidence	A character vector. Valid options: exp, igc, nas, tas, tax, asr. They can be combined. Default is all.
frequency	A character vector. Valid options: always, usually, sometimes, rarely, unknown. They can be combined. By default, "rarely" is excluded.
minSize	Minimum number of bugs in a signature. Default is 10.
min	Minimum value (inclusive). Only for numeric attributes. Default is NULL.
max	Maximum value (inclusive). Only for numeric attributes. Default is NULL.

Value

A list of character vectors with scientific names or taxids.

Examples

```
bp <- importBugphyzz()
sigs <- purrr::map(bp, makeSignatures)
sigs <- purrr::list_flatten(sigs, name_spec = "{inner}")
```

physiologies *Import physiologies (for devs)*

Description

physiologies imports a list of data.frames. This data is in "raw" state before cleaning and going through the data imputation steps. It should be used by developers/curators of the package.

Usage

```
physiologies(
  keyword = "all",
  fullSource = FALSE,
  sources = TRUE,
  forceDownload = FALSE
)
```

Arguments

keyword	Character vector with one or more valid keywords. Valid keyboards can be checked with showPhys. If 'all', all physiologies are imported.
fullSource	Logical. If TRUE, the Attribute_source column will contain full source information. If FALSE, the Attribute_source column will contain shortened versions of the sources. Default is FALSE.
sources	Logical. Also read the programmatic sources (Bergey's Manual extraction, Madin et al. 2020, ...) kept under sources/ in https://github.com/waldronlab/bugphyzzExports. Default is TRUE.
forceDownload	Logical. Re-download the programmatic sources instead of using the cached copy. Default is FALSE. The location can be changed with options(bugphyzz.sourcesBaseUrl = ...), e.g. to a local checkout.

Details

Rows come from three places: the curation Google Sheets listed in inst/extdata/spreadsheet_links.tsv, the BacDive strain table (built from the BacDive API, sources/bacdive in bugphyzzExports), and the other programmatic sources above. The last are produced by scripts, never edited by hand, and are cached with BiocFileCache like the exports.

Attribute values from all three are mapped to the controlled vocabulary in inst/extdata/attributes.tsv with the synonym table inst/extdata/attribute_synonyms.tsv (columns attribute_group, source_term, attribute), e.g. BacDive's "rod-shaped" to "bacillus". A source term can map to several attributes; an empty attribute drops the row on purpose.

Value

A list of data.frames in tidy format.

Examples

```
## Not run:
## Reads every curation sheet and source over the network (several minutes)
l <- physiologies('all')
df <- physiologies('aerophilicity')[[1]]

## End(Not run)
```

---

showPhys *Show list of available physiologies (for devs)*

Description

showPhys prints the names of the available physiologies that can be imported with the physiologies function. This function should be used by developers/curators.

Usage

```
showPhys(whichNames = "all")
```

Arguments

whichNames	A character string. Options: 'all' (default), 'spreadsheets', 'bacdive'.

Value

A character vector with the names of the physiologies.

Examples

```
showPhys()
showPhys('bacdive')
showPhys('spreadsheets')
```

Index

getTaxonSignatures, 2

importBugphyzz, 3

makeSignatures, 4

physiologies, 6, *7*

showPhys, 7