

# Package ‘bugphyzz’

October 31, 2024

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

**Version** 1.0.0

**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)

**RoxygenNote** 7.3.1

**VignetteBuilder** knitr

**Depends** R (>= 4.4)

**Imports** dplyr, utils, purrr, stringr, tidyr, tidyselct,  
BiocFileCache, httr2, tools, S4Vectors

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

**URL** <https://github.com/waldronlab/bugphyzz>

**BugReports** <https://github.com/waldronlab/bugphyzz/issues>

**git\_url** <https://git.bioconductor.org/packages/bugphyzz>

**git\_branch** RELEASE\_3\_20

**git\_last\_commit** 50358b0

**git\_last\_commit\_date** 2024-10-29

**Repository** Bioconductor 3.20

**Date/Publication** 2024-10-31

**Author** Samuel Gamboa [aut, cre] (<<https://orcid.org/0000-0002-6863-7943>>),  
 Levi Waldron [aut] (<<https://orcid.org/0000-0003-2725-0694>>),  
 Kelly Eckenrode [aut],  
 Jonathan Ye [aut],  
 Jennifer Wokaty [aut]

**Maintainer** Samuel Gamboa <Samuel.Gamboa.Tuz@gmail.com>

## Contents

getTaxonSignatures . . . . .	2
importBugphyzz . . . . .	3
makeSignatures . . . . .	4
physiologies . . . . .	5
showPhys . . . . .	5
<b>Index</b>	<b>7</b>

---

getTaxonSignatures	<i>Get Taxon Signatures</i>
--------------------	-----------------------------

---

## 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 <code>taxIdType = "Taxon_name"</code> must also be used.
bp	List of data.frames imported with <code>importBugphyzz</code> .
...	Arguments passed to <code>makeSignatures</code> .

## 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	<i>Import bugphyzz</i>
----------------	------------------------

---

## 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.12574596",
  forceDownload = FALSE,
  v = 0.8,
  excludeRarely = TRUE
)
```

## Arguments

version	Character string indicating the version. Default is the latest release on Zenodo. Options: Zenodo DOI, GitHub commit hash, or devel.
forceDownload	Logical value. Force a fresh download of the data or use the one stored in the cache (if available). Default is FALSE.
v	Validation value. Default 0.8 (see details).
excludeRarely	Default is TRUE. Exclude values with Frequency == FALSE (see details).

## Details

### Data structure:

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

### Validation (v argument):

Data imported with importBugphyzz includes annotations imputed through ancestral state reconstruction (ASR) methods. A 10-fold cross-validation approach was implemented to assess the reliability of the data imputed. Mathew's correlation coefficient (MCC) and R-squared (R2) were used for the validation of discrete and numeric attributes. Details can be found at: <https://github.com/waldronlab/taxPPPr>. By default, imputed annotations with a MCC or R2 value greater than 0.5 are imported. The minimum value can be adjusted with the v argument (only values between 0 and 1).

### Frequency (excludeRarely argument):

One of the variables in the bugphyzz data.frames is "Frequency", which can adopt values of "always", "usually", "sometimes", "rarely", or "never". By default "never" and "rarely" are excluded. "rarely" could be included with `excludeRarely = FALSE`. To learn more about these frequency keywords please check the bugphyzz vignette with `browseVignettes("bugphyzz")`.

### Sources:

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

**Value**

A list of tidy data frames.

**Examples**

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

---

makeSignatures	<i>Make signatures</i>
----------------	------------------------

---

**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	<i>Import physiologies (for devs)</i>
--------------	---------------------------------------

---

**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)
```

**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.

**Value**

A list of data.frames in tidy format.

**Examples**

```
l <- physiologies('all')
df <- physiologies('aerophilicity')[[1]]
```

---

showPhys	<i>Show list of available physiologies (for devs)</i>
----------	-------------------------------------------------------

---

**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', 'bacdiv'.
------------	-------------------------------------------------------------------------

**Value**

A character vector with the names of the physiologies.

**Examples**

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

# Index

`getTaxonSignatures`, [2](#)

`importBugphyzz`, [3](#)

`makeSignatures`, [4](#)

`physiologies`, [5](#), [5](#)

`showPhys`, [5](#)