Package ‘TaxSEA’

January 16, 2026

Type Package

Title Taxon Set Enrichment Analysis

Version 1.3.1

Description TaxSEA is an R package for Taxon Set Enrichment Analysis, which utilises a Kolmogorov-Smirnov test analyses to investigate differential abundance analysis output for whether there are alternations in a-priori defined sets of taxa from public databases (BugSigDB, MiMeDB, GutMGene, mBodyMap, BacDive and GMRepoV2) and collated from the literature. TaxSEA takes as input a list of taxonomic identifiers (e.g. species names, NCBI IDs etc.) and a rank (E.g. fold change, correlation coefficient). TaxSEA be applied to any microbiota taxonomic profiling technology (array-based, 16S rRNA gene sequencing, shotgun metagenomics & metatranscriptomics etc.) and enables researchers to rapidly contextualize their findings within the broader literature to accelerate interpretation of results.

License GPL-3

Encoding UTF-8

LazyData false

VignetteBuilder knitr

RoxygenNote 7.3.3

biocViews Microbiome, Metagenomics, Sequencing, GeneSetEnrichment, RNASeq

URL `https://github.com/feargalr/taxsea`, `https://feargalr.github.io/TaxSEA/`

BugReports `https://github.com/feargalr/taxsea/issues`

Depends R (>= 4.5.0)

Suggests BiocStyle, bugsigdbr, fgsea, knitr, rmarkdown, testthat

Imports stats, utils

Config/testthat/edition 3

git_url https://git.bioconductor.org/packages/TaxSEA

git_branch devel

git_last_commit 4a2f425

git_last_commit_date 2026-01-14

Repository Bioconductor 3.23

Date/Publication 2026-01-15

Author Feargal Ryan [aut, cre, fnd] (ORCID: <`https://orcid.org/0000-0002-1565-4598`>, funding: Supported by NHMRC Investigator Grant)

Maintainer Feargal Ryan <`feargalr@gmail.com`>

Contents

`get_ncbi_taxon_ids` *Retrieve NCBI Taxonomy IDs for a list of taxon names*

Description

This function takes a vector of taxon names and returns a vector of NCBI taxonomy IDs by querying the NCBI Entrez API.

Usage

```
get_ncbi_taxon_ids(taxon_names)
```

Arguments

`taxon_names`	A character vector of taxon names

Value

A character vector of NCBI taxonomy IDs corresponding to the input taxon names

Examples

```
taxon_names <- c("Escherichia coli", "Staphylococcus aureus")
taxon_ids <- get_ncbi_taxon_ids(taxon_names)
```

`get_taxon_sets` *Retrieve Taxon Sets from TaxSEA Library*

Description

Retrieve from the TaxSEA database which taxon sets (metabolite producers and disease signatures) contain a taxon of interest.

Usage

```
get_taxon_sets(taxon_to_fetch = taxon)
```

Arguments

`taxon_to_fetch`	The taxon to search for in the TaxSEA database.

Value

A character vector containing the names of taxonomic sets where the specified taxon is present.

Examples

```
# Retrieve sets for Bifidobacterium longum
get_taxon_sets(taxon="Bifidobacterium_longum")
```

`NCBI_ids` *NCBI IDs Dataset*

Description

A dataset for mapping NCBI IDs to species/genus names. This named vector allows for lookup of NCBI IDs associated with species or genus names.

Usage

```
NCBI_ids
```

Format

A named vector where:

names NCBI IDs

values Species or genus names

A named vector mapping NCBI IDs to species or genus names.

Source

NCBI

Examples

```
data(NCBI_ids)
# Can look up either with or without spaces
NCBI_ids["Bifidobacterium_breve"]
NCBI_ids["Bifidobacterium breve"]
```

`TaxSEA` *TaxSEA: Taxon Set Enrichment Analysis*

Description

Modular TaxSEA implementation supporting enrichment (KS) and ORA (Fisher). Provide either `taxon_ranks` for enrichment or `input_taxa` for ORA.

Usage

```
TaxSEA(
  taxon_ranks = NULL,
  input_taxa = NULL,
  mode = NULL,
  lookup_missing = FALSE,
  min_set_size = 5,
  max_set_size = 300,
  custom_db = NULL
)
```

Arguments

`taxon_ranks`	Named numeric vector of statistics (e.g., log2 fold changes). Required for enrichment.
`input_taxa`	Character vector of taxa to treat as "hits"/selected taxa. Required for ORA.
`mode`	Character. One of `"enrichment"` or `"ora"`. If NULL, inferred from which input is provided.
`lookup_missing`	Logical indicating whether to fetch missing NCBI IDs. Default is FALSE.
`min_set_size`	Minimum size of taxon sets to include in the analysis. Default is 5.
`max_set_size`	Maximum size of taxon sets to include in the analysis. Default is 100.
`custom_db`	A user-provided list of taxon sets. If NULL (default), the built-in database is used.

Value

A list of data frames with taxon set results.

Examples

```
data("TaxSEA_test_data")
res <- TaxSEA(taxon_ranks = TaxSEA_test_data)
head(res$All_databases)

# ORA example (toy): treat taxa with positive values as "hits"
hits <- names(TaxSEA_test_data)
res_ora <- TaxSEA(input_taxa = hits, mode = "ora")
head(res_ora$All_databases)
```

`TaxSEA_db` *TaxSEA Database A dataset containing taxon sets. Each item in the list is a taxon set, and each member within a taxon set is a taxon.*

Description

TaxSEA Database A dataset containing taxon sets. Each item in the list is a taxon set, and each member within a taxon set is a taxon.

Usage

```
TaxSEA_db
```

Format

A list of vectors. Each vector contains character strings representing taxa.

Source

See READ ME.

Examples

```
data(TaxSEA_db)
all_sets <- names(TaxSEA_db)
GABA_producers<-TaxSEA_db[["MiMeDB_producers_of_GABA"]]
```

`TaxSEA_test_data` *TaxSEA Test Data*

Description

A dataset containing taxon ranks and taxon IDs.

Usage

```
TaxSEA_test_data
```

Format

A data frame with two columns:

rank Character vector representing taxon ranks

id Character vector representing taxon IDs

A data frame with columns 'rank' and 'id' representing taxon ranks and taxon IDs, respectively.

Source

See READ ME.

Examples

```
data(TaxSEA_test_data)
test_results <- TaxSEA(TaxSEA_test_data)
```

Index