Package ‘SpoMAG’

May 29, 2025

Type Package

Title Probability of Sporulation Potential in MAGs

Version 0.1.0

Description Implements an ensemble machine learning approach to predict the sporulation potential of metagenome-assembled genomes (MAGs) from uncultivated Firmicutes based on the presence/absence of sporulation-associated genes.

License Artistic-2.0

Encoding UTF-8

Imports dplyr, tidyr, tibble, stats

RoxygenNote 7.3.2

Suggests testthat (>= 3.0.0), caret, kernlab, randomForest, readr

Config/testthat/edition 3

NeedsCompilation no

Author Douglas Terra Machado [aut, cre] (ORCID:
<`https://orcid.org/0000-0002-6580-7628`>),
Otávio José Bernardes Brustolini [ctb] (ORCID:
<`https://orcid.org/0000-0001-8132-9753`>),
Ellen dos Santos Corrêa [ctb],
Ana Tereza Ribeiro Vasconcelos [ctb] (ORCID:
<`https://orcid.org/0000-0002-4632-2086`>)

Maintainer Douglas Terra Machado <`dougterra@gmail.com`>

Depends R (>= 3.5.0)

Repository CRAN

Date/Publication 2025-05-29 18:20:09 UTC

Contents

`build_binary_matrix` *Build binary presence/absence matrix of sporulation genes*

Description

Transforms the output of `sporulation_gene_name()` into a wide-format matrix indicating the presence (1) or absence (0) of each sporulation-associated gene per genome.

Usage

```
build_binary_matrix(df)
```

Arguments

`df`	A data.frame from `sporulation_gene_name()` with columns `genome_ID` and `spo_gene_name`.

Value

A wide-format binary matrix with genomes in rows and genes in columns.

Examples

```
# Load package
library(SpoMAG)

# Load example annotation tables
file_spor <- system.file("extdata", "one_sporulating.csv.gz", package = "SpoMAG")
file_aspo <- system.file("extdata", "one_asporogenic.csv.gz", package = "SpoMAG")

# Read files
df_spor <- readr::read_csv(file_spor, show_col_types = FALSE)
df_aspo <- readr::read_csv(file_aspo, show_col_types = FALSE)

# Step 1: Extract sporulation-related genes
genes_spor <- sporulation_gene_name(df_spor)
genes_aspo <- sporulation_gene_name(df_aspo)

# Step 2: Convert to binary matrix
bin_spor <- build_binary_matrix(genes_spor)
bin_aspo <- build_binary_matrix(genes_aspo)
```

`predict_sporulation` *Predict Sporulation Potential*

Description

This function predicts the sporulation potential of MAGs using an ensemble learning model. It uses probabilities from Random Forest and SVM classifiers as inputs to a meta-model.

Usage

```
predict_sporulation(binary_matrix)
```

Arguments

`binary_matrix`	A binary matrix (1/0) indicating gene presence/absence for each MAG. Must include a `genome_ID` column.

Value

A tibble with predicted class and probability of sporulation for each genome.

Examples

```
# Load package
library(SpoMAG)

# Load example annotation tables
file_spor <- system.file("extdata", "one_sporulating.csv.gz", package = "SpoMAG")
file_aspo <- system.file("extdata", "one_asporogenic.csv.gz", package = "SpoMAG")

# Read files
df_spor <- readr::read_csv(file_spor, show_col_types = FALSE)
df_aspo <- readr::read_csv(file_aspo, show_col_types = FALSE)

# Step 1: Extract sporulation-related genes
genes_spor <- sporulation_gene_name(df_spor)
genes_aspo <- sporulation_gene_name(df_aspo)

# Step 2: Convert to binary matrix
bin_spor <- build_binary_matrix(genes_spor)
bin_aspo <- build_binary_matrix(genes_aspo)

# Step 3: Predict using ensemble model (preloaded in package)

result_spor <- predict_sporulation(bin_spor)
result_aspo <- predict_sporulation(bin_aspo)
```

`sporulation_gene_name` *Identify Sporulation-Associated Genes*

Description

This function identifies sporulation-associated genes in a genome annotation data frame. It searches for gene names and KEGG Orthology identifiers related to sporulation steps and returns a data frame with annotated sporulation genes and a consensus name.

Usage

```
sporulation_gene_name(df)
```

Arguments

`df`	A data frame containing MAG annotation with the columns ’Preferred_name’, ’KEGG_ko’, and ’genome_ID’.

Value

A data frame of sporulation-associated genes with standardized names and spo_process tags.

Examples

```
# Load package
library(SpoMAG)
# Load example annotation tables
file_spor <- system.file("extdata", "one_sporulating.csv.gz", package = "SpoMAG")
file_aspo <- system.file("extdata", "one_asporogenic.csv.gz", package = "SpoMAG")

# Read files
df_spor <- readr::read_csv(file_spor, show_col_types = FALSE)
df_aspo <- readr::read_csv(file_aspo, show_col_types = FALSE)

# Step 1: Extract sporulation-related genes
genes_spor <- sporulation_gene_name(df_spor)
genes_aspo <- sporulation_gene_name(df_aspo)
```

Index