

Package ‘MetaComp’

October 12, 2022

Version 1.1.2

Title EDGE Taxonomy Assignments Visualization

Description Implements routines for metagenome sample taxonomy assignments collection, aggregation, and visualization. Accepts the EDGE-formatted output from GOTTCHA/GOTTCHA2, BWA, Kraken, MetaPhlAn, DIAMOND, and Pangia. Produces SVG and PDF heatmap-like plots comparing taxa abundances across projects.

URL <https://github.com/seninp-bioinfo/MetaComp>

BugReports <https://github.com/seninp-bioinfo/MetaComp/issues>

Depends R (>= 3.1.0)

Imports reshape2, plyr, dplyr, data.table, ggplot2, Cairo

Suggests testthat

LinkingTo

LazyData true

License GPL-2

RoxygenNote 6.0.1

NeedsCompilation no

Author Pavel Senin [aut, cre]

Maintainer Pavel Senin <senin@hawaii.edu>

Repository CRAN

Date/Publication 2018-06-18 20:06:45 UTC

R topics documented:

load_edge_assignment	2
load_edge_assignments	3
merge_edge_assignments	4
merge_edge_counts	5
plot_edge_assignment	5
plot_merged_assignment	6

`load_edge_assignment` *Efficiently loads an EDGE-produced taxonomic assignment from a file. An assumption has been made – since EDGE tables are generated in an automated fashion, they should be properly formatted – thus the code doesn't check for any inconsistencies except for the very file existence. Note however, the unassigned to taxa entries are removed. This implementation fully relies on the `fread` function from `data.table` package gaining performance over traditional R techniques.*

Description

Efficiently loads an EDGE-produced taxonomic assignment from a file. An assumption has been made – since EDGE tables are generated in an automated fashion, they should be properly formatted – thus the code doesn't check for any inconsistencies except for the very file existence. Note however, the unassigned to taxa entries are removed. This implementation fully relies on the `fread` function from `data.table` package gaining performance over traditional R techniques.

Usage

```
load_edge_assignment(filepath, type)
```

Arguments

`filepath` the path to EDGE-generated tab-delimited taxonomy assignment file.

`type` the assignment type. Following types are recognized: 'bwa', 'diamond', 'gottcha', 'gottcha2', 'kraken', 'metaphlan', and 'pangia'.

Value

a data frame containing four columns: TAXA, LEVEL, COUNT, and ABUNDANCE, representing taxonomically anchored sequences from the sample.

Examples

```
pa_fpath <- system.file("extdata", "HMP_even//allReads-pangia.list.txt", package="MetaComp")
pangia_assignment = load_edge_assignment(pa_fpath, type = "pangia")

table(pangia_assignment$LEVEL)

pangia_assignment[pangia_assignment$LEVEL == "phylum",]
```

load_edge_assignments *Efficiently loads a BWA (or other EDGE-like taxonomic assignment) tables from a list of files. Outputs a named list of assignments.*

Description

Efficiently loads a BWA (or other EDGE-like taxonomic assignment) tables from a list of files. Outputs a named list of assignments.

Usage

```
load_edge_assignments(filepath, type)
```

Arguments

filepath the path to tab delimited, two-column file whose first column is a project_id (which will be used to name this assignment) and the second column is the assignment filename.

type the type of assignments to be loaded. Following types are recognized: 'bwa', 'diamond', 'gottcha', 'gottcha2', 'kraken', 'metaphlan', and 'pangia'.

Value

a list of all read assignments.

Examples

```
hmp_even_fp <- system.file("extdata", "HMP_even", package="MetaComp")
hmp_stagger_fp <- system.file("extdata", "HMP_stagger", package="MetaComp")
data_files <- data.frame(V1 = c("HMP_even", "HMP_stagger"),
                        V2 = c(file.path(hmp_even_fp, "allReads-gottcha2-speDB-b.list.txt"),
                              file.path(hmp_stagger_fp, "allReads-gottcha2-speDB-b.list.txt")))
write.table(data_files, file.path(tempdir(), "assignments.txt"),
            row.names = FALSE, col.names = FALSE)
gottcha2_assignments = load_edge_assignments(file.path(tempdir(), "assignments.txt"),
                                             type = "gottcha2")

names(gottcha2_assignments)
table(gottcha2_assignments[[1]]$LEVEL)
```

merge_edge_assignments

Merges two or more EDGE-like taxonomical assignments. The input data frames are assumed to have the following columns: LEVEL, TAXA, and ABUNDANCE – these will be used in the merge procedure, all other columns will be ignored.

Description

Merges two or more EDGE-like taxonomical assignments. The input data frames are assumed to have the following columns: LEVEL, TAXA, and ABUNDANCE – these will be used in the merge procedure, all other columns will be ignored.

Usage

```
merge_edge_assignments(assignments)
```

Arguments

`assignments` A named list of assignments (the list element's name will be used as a resulting data frame column name).

Value

A merged table, which is a data frame whose rows are taxonomical ids and columns are the input assignments ids.

Examples

```
## Not run:
hmp_even_fp <- system.file("extdata", "HMP_even", package="MetaComp")
hmp_stagger_fp <- system.file("extdata", "HMP_stagger", package="MetaComp")
data_files <- data.frame(V1 = c("HMP_even", "HMP_stagger"),
                        V2 = c(file.path(hmp_even_fp, "allReads-gottcha2-speDB-b.list.txt"),
                              file.path(hmp_stagger_fp, "allReads-gottcha2-speDB-b.list.txt")))
write.table(data_files, file.path(tempdir(), "assignments.txt"),
            row.names = FALSE, col.names = FALSE)
gottcha2_assignments = merge_edge_assignments(
  load_edge_assignments(
    file.path(tempdir(), "assignments.txt"), type = "gottcha2"))

## End(Not run)
```

merge_edge_counts	<i>Merges two or more EDGE-like taxonomical assignments. The input data frames are assumed to have the following columns: LEVEL, TAXA, and COUNT – these will be used in the merge procedure, all other columns will be ignored.</i>
-------------------	--

Description

Merges two or more EDGE-like taxonomical assignments. The input data frames are assumed to have the following columns: LEVEL, TAXA, and COUNT – these will be used in the merge procedure, all other columns will be ignored.

Usage

```
merge_edge_counts(assignments)
```

Arguments

assignments	A named list of assignments (the list element's name will be used as a resulting data frame column name).
-------------	---

Value

A merged table, which is a data frame whose rows are taxonomical ids and columns are the input assignments ids.

plot_edge_assignment	<i>Generates a single column ggplot for a taxonomic assignment table and also outputs a PDF.</i>
----------------------	--

Description

This implementation is built upon ggplot geom_tile.

Usage

```
plot_edge_assignment(assignment, level, plot_title, column_title, filename)
```

Arguments

assignment	The EDGE-like assignment table.
level	The taxonomic level to plot (i.e., family, strain, etc...).
plot_title	The plot title, e.g., "Project XX, Run YY".
column_title	The column title.
filename	The PDF file name mask.

Value

the ggplot2 plot.

Examples

```
pa_fpath <- system.file("extdata", "HMP_even//allReads-pangia.list.txt", package="MetaComp")
pangia_assignment = load_edge_assignment(pa_fpath, type = "pangia")

plot_edge_assignment(pangia_assignment, "phylum", "Pangia", "HMP Even",
                    file.path(tempdir(), "assignment.pdf"))
```

plot_merged_assignment

Generates a single column ggplot for a taxonomic assignment table.

Description

This implementation...

Usage

```
plot_merged_assignment(assignment, taxonomy_level,
  sorting_order = "abundance", row_limit = 60, min_row_abundance = 0,
  plot_title, filename)
```

Arguments

assignment	The gottcha-like merged assignment table.
taxonomy_level	The level which need to be plotted.
sorting_order	the order in which rows shall be sorted, "abundance" is default, "alphabetical" is an alternative.
row_limit	the max amount of rows to plot (default is 60).
min_row_abundance	the minimal sum of abundances in a row required to plot. Rows whose sum is less than this value are dropped even if row_limit is specified. Ignored for "alphabetical" order. (default 0.0).
plot_title	The plot title.
filename	The output file mask, PDF and SVG files will be produced with Cairo device.

Examples

```
## Not run:
hmp_even_fp <- system.file("extdata", "HMP_even", package="MetaComp")
hmp_stagger_fp <- system.file("extdata", "HMP_stagger", package="MetaComp")
data_files <- data.frame(V1 = c("HMP_even", "HMP_stagger"),
                        V2 = c(file.path(hmp_even_fp, "allReads-gottcha2-speDB-b.list.txt"),
                              file.path(hmp_stagger_fp, "allReads-gottcha2-speDB-b.list.txt")))
write.table(data_files, file.path(tempdir(), "assignments.txt"),
            row.names = FALSE, col.names = FALSE)
gottcha2_assignments = merge_edge_assignments(
  load_edge_assignments(
    file.path(tempdir(), "assignments.txt"), type = "gottcha2"))
plot_merged_assignment(gottcha2_assignments, "family", 'alphabetical', 100, 0,
                      "HMP side-to-side", file.path(tempdir(), "assignment.pdf"))

## End(Not run)
```

Index

load_edge_assignment, [2](#)
load_edge_assignments, [3](#)

merge_edge_assignments, [4](#)
merge_edge_counts, [5](#)

plot_edge_assignment, [5](#)
plot_merged_assignment, [6](#)