

Package ‘MAGNAMWAR’

October 12, 2022

Title A Pipeline for Meta-Genome Wide Association

Version 2.0.4

Date 2018-06-15

Description Correlates variation within the meta-genome to target species phenotype variations in meta-genome with association studies. Follows the pipeline described in Chaston, J.M. et al. (2014) <[doi:10.1128/mBio.01631-14](https://doi.org/10.1128/mBio.01631-14)>.

License MIT + file LICENSE

LazyData true

Imports ape, coxme, doParallel, dplyr, foreach, iterators, lme4, multcomp, parallel, plyr, qqman, survival, seqinr

Suggests knitr, rmarkdown

VignetteBuilder knitr

Depends R (>= 3.0)

RoxygenNote 6.0.1

NeedsCompilation no

Author Corinne Sexton [aut],
John Chaston [aut, cre],
Hayden Smith [ctb]

Maintainer John Chaston <john_chaston@byu.edu>

Repository CRAN

Date/Publication 2018-07-12 07:20:17 UTC

R topics documented:

| | |
|---|---|
| after_ortho_format | 2 |
| after_ortho_format_grps | 3 |
| AnalyzeOrthoMCL | 3 |
| CalculatePrincipalCoordinates | 6 |
| FormatAfterOrtho | 6 |
| FormatMCLFastas | 7 |
| joined_mtrx | 8 |

| | |
|----------------------------|-----------|
| joined_mtrx_grps | 9 |
| JoinRepSeq | 9 |
| ManhatGrp | 10 |
| mcl_mtrx | 11 |
| mcl_mtrx_grps | 11 |
| PDGPlot | 12 |
| PDGvOG | 13 |
| pheno_data | 13 |
| PhyDataError | 14 |
| PrintOGSeqs | 15 |
| QQPlotter | 16 |
| RASTtoGBK | 16 |
| starv_pheno_data | 17 |
| SurvAppendMatrix | 18 |
| WriteMCL | 18 |
| Index | 20 |

after_ortho_format *Formatted output of OrthoMCL.*

Description

A list created by inputting the output of OrthoMCL clusters into the FormatAfterOrtho function.

Usage

after_ortho_format

Format

List of 2: (1) presence absence matrix, (2) protein ids:

pa_matrix matrix showing taxa presence/absence in OG

proteins matrix listing protein_id contained in each OG

 after_ortho_format_grps

Formatted output of OrthoMCL.

Description

A list created by inputting the output of OrthoMCL clusters into the FormatAfterOrtho function.

Usage

```
after_ortho_format_grps
```

Format

List of 2: (1) presence absence matrix, (2) protein ids:

pa_matrix matrix showing taxa presence/absence in OG

proteins matrix listing protein_id contained in each OG

 AnalyzeOrthoMCL

Main OrthoMCL Analysis

Description

Main function for analyzing the statistical association of OG (orthologous group) presence with phenotype data

Usage

```
AnalyzeOrthoMCL(mcl_data, pheno_data, model, species_name, resp = NULL,
  fix2 = NULL, rndm1 = NULL, rndm2 = NULL, multi = 1, time = NULL,
  event = NULL, time2 = NULL, startnum = 1, stopnum = "end",
  output_dir = NULL, sig_digits = NULL, princ_coord = 0)
```

Arguments

| | |
|------------|---|
| mcl_data | output of FormatAfterOrtho; a list of matrices; (1) a presence/absence matrix of taxa per OG, (2) a list of the specific protein ids within each OG |
| pheno_data | a data frame of phenotypic data with specific column names used to specify response variable as well as other fixed and random effects |

| | |
|--------------|--|
| model | linear model with gene presence as fixed effect (lm), linear mixed mfect models with gene presence as fixed effect and additional variables specified as: one random effect (lmeR1); two independent random effects (lmeR2ind); two random effects with rndm2 nested in rndm1 (lmeR2nest); or two independent random effects with one additional fixed effect (lmeF2), Wilcox Test with gene presence as fixed effect (wx), Survival Tests with support for multi core design: with two random effects (survmulti), and with two times as well as an additional fixed variable (survmulticensor) |
| species_name | Column name in pheno_data containing 4-letter species designations |
| resp | Column name in pheno_data containing response variable |
| fix2 | Column name in pheno_data containing second fixed effect |
| rndm1 | Column name in pheno_data containing first random variable |
| rndm2 | Column name in pheno_data containing second random variable |
| multi | (can only be used with survival tests) Number of cores |
| time | (can only be used with survival tests) Column name in pheno_data containing first time |
| event | (can only be used with survival tests) Column name in pheno_data containing event |
| time2 | (can only be used with survival tests) Column name in pheno_data containing second time |
| startnum | number of test to start on |
| stopnum | number of test to stop on |
| output_dir | (if using survival tests) directory where small output files will be placed before using SurvAppendMatrix. Must specify a directory if choosing to output small files, else only written as a matrix |
| sig_digits | amount of digits to display for p-values and means of data; default to NULL (no rounding) |
| princ_coord | the number of principle coordinates to be included in model as fixed effects (1, 2, or 3), if a decimal is specified, as many principal coordinates as are needed to account for that percentage of the variance will be included in the analysis |

Value

A matrix with the following columns: OG, p-values, Bonferroni corrected p-values, mean phenotype of OG-containing taxa, mean pheotype of OG-lacking taxa, taxa included in OG, taxa not included in OG

Examples

```
#Linear Model

## Not run:
mcl_mtrx <- AnalyzeOrthoMCL(after_ortho_format, pheno_data, 'lm',
  'Treatment', resp='RespVar')
```

```
## End(Not run)

# the rest of the examples are not run for time's sake
#Linear Mixed Effect with one random effect
## Not run:
mcl_mtrx <- AnalyzeOrthoMCL(after_ortho_format, pheno_data, 'lmeR1',
  'Treatment', resp='RespVar', rndm1='Experiment')

## End(Not run)

#Linear Mixed Effect with two independent random effects
## Not run:
mcl_mtrx <- AnalyzeOrthoMCL(after_ortho_format, pheno_data, 'lmeR2ind',
  'Treatment', resp='RespVar', rndm1='Experiment', rndm2='Vial')

## End(Not run)

#Linear Mixed Effect with rndm2 nested in rndm1
## Not run:
mcl_mtrx <- AnalyzeOrthoMCL(after_ortho_format, pheno_data, 'lmeR2nest',
  'Treatment', resp='RespVar', rndm1='Experiment', rndm2='Vial')

## End(Not run)

#Linear Mixed Effect with two independent random effects and one additional fixed effect
## Not run:
mcl_mtrx3 <- AnalyzeOrthoMCL(after_ortho_format, pheno_data, 'lmeF2',
  'Treatment', resp='RespVar', fix2='Treatment', rndm1='Experiment', rndm2='Vial', princ_coord = 4)

## End(Not run)

#Wilcoxon Test
## Not run:
mcl_mtrx <- AnalyzeOrthoMCL(after_ortho_format, pheno_data, 'wx',
  'Treatment', resp='RespVar')

## End(Not run)

# ~ 5 minutes
#Survival with two independent random effects, run on multiple cores
## Not run:
mcl_mtrx <- AnalyzeOrthoMCL(after_ortho_format, starv_pheno_data, 'TRT', model='survmulti',
  time='t2', event='event', rndm1='EXP', rndm2='VIAL', multi=1)

## End(Not run)

# ~ 5 minutes
#Survival with two independent random effects and one additional fixed effect,
#includeing drops on multi cores
```

```
## Not run:
mcl_mtx <- AnalyzeOrthoMCL(after_ortho_format, starv_pheno_data, 'TRT', model='survmulticensor',
  time='t1', time2='t2', event='event', rndm1='EXP', rndm2='VIAL', fix2='BACLO', multi=1)

## End(Not run)
#to be appended with SurvAppendMatrix
```

CalculatePrincipalCoordinates

Show Principal Components Breakdown

Description

Function to show Principal Components statistics based on the OrthoMCL presence absence groupings.

Usage

```
CalculatePrincipalCoordinates(mcl_data)
```

Arguments

mcl_data output of FormatAfterOrtho –list of 2 things– 1: binary matrix indicating the presence / absence of genes in each OG and 2: vector of names of OGs

Value

returns a named list of principal components and accompanying proportion of variance for each

Examples

```
CalculatePrincipalCoordinates(after_ortho_format)
```

FormatAfterOrtho

Format file from output of OrthoMCL algorithm before use in Analyze-OrthoMCL

Description

After running OrthoMCL and/or submitting to www.orthomcl.org, formats the output file to be used in AnalyzeOrthoMCL

Usage

```
FormatAfterOrtho(file, format = "ortho")
```

Arguments

| | |
|--------|--|
| file | Path to the OrthoMCL output file |
| format | Specification of the method by which file was obtained: defaults to 'ortho' for output from orthomcl.org. Other option is 'groups' for output from local run of OrthoMCL software. |

Value

a list of matrices; (1) a presence/absence matrix of taxa per OG, (2) a list of the specific protein ids within each OG

Examples

```
file <- system.file('extdata', 'orthologGroups.txt', package='MAGNAMWAR')
after_ortho_format <- FormatAfterOrtho(file)
```

```
file_grps <- system.file('extdata', 'groups_example_r.txt', package='MAGNAMWAR')
after_ortho_format_grps <- FormatAfterOrtho(file_grps, format = 'groups')
```

| | |
|-----------------|---|
| FormatMCLFastas | <i>Format all raw GenBank fastas to single OrthoMCL compatible fasta file</i> |
|-----------------|---|

Description

Creates the composite fasta file for use in running OrthoMCL and/or submitting to www.orthomcl.org

Usage

```
FormatMCLFastas(fa_dir, genbnk_id = 4)
```

Arguments

| | |
|-----------|---|
| fa_dir | Path to the directory where all raw GenBank files are stored. Note, all file names must be changed to a 4-letter code representing each species and have '.fasta' file descriptor |
| genbnk_id | (Only necessary for the deprecated version of fasta headers) The index of the sequence ID in the GenBank pipe-separated annotation line (default: 4) |

Value

Returns nothing, but prints the path to the final OrthoMCL compatible fasta file

Examples

```
## Not run:
dir <- system.file('extdata', 'fasta_dir', package='MAGNAMWAR')
dir <- paste(dir, '/', sep='')
formatted_file <- FormatMCLFastas(dir)

## End(Not run)
```

joined_mtrx

Final output of join_repset.

Description

A data frame containing the final results of statistical analysis with protein ids, annotations, and sequences added.

Usage

```
joined_mtrx
```

Format

A data frame with 17 rows and 11 variables:

OG taxa cluster id, as defined by OrthoMCL

pval1 p-value, based on presence absence

corrected_pval1 Bonferroni p-value, corrected by number of tests

mean_OGContain mean of all taxa phenotypes in that OG

mean_OGLack mean of all taxa phenotypes not in that OG

taxa_contain taxa in that cluster

taxa_miss taxa not in that cluster

rep_taxon randomly selected representative taxa from the cluster

rep_id protein id, from randomly selected representative taxa

rep_annot fasta annotation, from randomly selected representative taxa

rep_seq AA sequence, from randomly selected representative taxa

joined_mtrx_grps *Final output of join_repsct.*

Description

A data frame containing the final results of statistical analysis with protein ids, annotations, and sequences added.

Usage

```
joined_mtrx_grps
```

Format

A data frame with 10 rows and 11 variables:

OG taxa cluster id, as defined by OrthoMCL

pval1 p-value, based on presence absence

corrected_pval1 Bonferroni p-value, corrected by number of tests

mean_OGContain mean of all taxa phenotypes in that OG

mean_OGLack mean of all taxa phenotypes not in that OG

taxa_contain taxa in that cluster

taxa_miss taxa not in that cluster

rep_taxon randomly selected representative taxa from the cluster

rep_id protein id, from randomly selected representative taxa

rep_annot fasta annotation, from randomly selected representative taxa

rep_seq AA sequence, from randomly selected representative taxa

JoinRepSeq *Join Representative Sequences*

Description

Joins the OrthoMCL output matrix to representative sequences

Usage

```
JoinRepSeq(mcl_data, fa_dir, mcl_mtrx, fastaformat = "new")
```

Arguments

| | |
|-------------|---|
| mcl_data | output of FormatAfterOrtho; a list of matrices; (1) a presence/absence matrix of taxa per OG, (2) a list of the specific protein ids within each OG |
| fa_dir | Path to the directory where all raw GenBank files are stored. Note, all file names must be changed to a 4-letter code representing each species and have '.fasta' file descriptor |
| mcl_mtrx | OrthoMCL output matrix from AnalyzeOrthoMCL() |
| fastaformat | options: new & old; new = no GI numbers included; defaults to new |

Value

Returns the original OrthoMCL output matrix with additional columns: representative sequence taxon, representative sequence id, representative sequence annotation, representative sequence

Examples

```
## Not run:
dir <- system.file('extdata', 'fasta_dir', package='MAGNAMWAR')
dir <- paste(dir, '/', sep='')
joined_mtrx_grps <- JoinRepSeq(after_ortho_format_grps, dir, mcl_mtrx_grps, fastaformat = 'old')

## End(Not run)
```

ManhatGrp

Manhattan Plot of All Taxa

Description

Manhattan plot that graphs all p-values for taxa.

Usage

```
ManhatGrp(mcl_data, mcl_mtrx, tree = NULL)
```

Arguments

| | |
|----------|---|
| mcl_data | FormatAfterOrtho output |
| mcl_mtrx | output of AnalyzeOrthoMCL() |
| tree | tree file optional, used for ordering taxa along x axis |

Value

a manhattan plot

References

Some sort of reference

Examples

```
ManhatGrp(after_ortho_format, mcl_mtrx)
```

```
#@param equation of line of significance, defaults to -log10((.05)/dim(pdgs)[1])
```

| | |
|----------|--|
| mcl_mtrx | <i>Final output of AnalyzeOrthoMCL</i> |
|----------|--|

Description

A matrix containing the final results of statistical analysis.

Usage

```
mcl_mtrx
```

Format

A matrix with 17 rows and 7 variables:

OG taxa cluster id, as defined by OrthoMCL

pval1 p-value, based on presence absence

corrected_pval1 Bonferroni p-value, corrected by number of tests

mean_OGContain mean of all taxa phenotypes in that OG

mean_OGLack mean of all taxa phenotypes not in that OG

taxa_contain taxa in that cluster

taxa_miss taxa not in that cluster

| | |
|---------------|--|
| mcl_mtrx_grps | <i>Final output of AnalyzeOrthoMCL</i> |
|---------------|--|

Description

A matrix containing the final results of statistical analysis.

Usage

```
mcl_mtrx_grps
```

Format

A matrix with 10 rows and 7 variables:

OG taxa cluster id, as defined by OrthoMCL

pval1 p-value, based on presence absence

corrected_pval1 Bonferroni p-value, corrected by number of tests

mean_OGContain mean of all taxa phenotypes in that OG

mean_OGLack mean of all taxa phenotypes not in that OG

taxa_contain taxa in that cluster

taxa_miss taxa not in that cluster

 PDGPlot

Plot of a PDG and Data with Standard Error Bars

Description

Bar plot of PDG vs phenotype data with presence of taxa in PDG indicated by color

Usage

```
PDGPlot(data, mcl_matrix, OG = "NONE", species_colname, data_colname,
        xlab = "Taxa", ylab = "Data", ylimit = NULL, tree = NULL,
        order = NULL, main_title = NULL)
```

Arguments

| | |
|------------------------------|---|
| <code>data</code> | R object of phenotype data |
| <code>mcl_matrix</code> | AnalyzeOrthoMCL output |
| <code>OG</code> | optional parameter, a string with the name of chosen group (OG) to be colored |
| <code>species_colname</code> | name of column in phenotypic data file with taxa designations |
| <code>data_colname</code> | name of column in phenotypic data file with data observations |
| <code>xlab</code> | string to label barplot's x axis |
| <code>ylab</code> | string to label barplot's y axis |
| <code>ylimit</code> | optional parameter to limit y axis |
| <code>tree</code> | optional parameter (defaults to NULL) Path to tree file, orders the taxa by phylogenetic distribution, else it defaults to alphabetical |
| <code>order</code> | vector with order of taxa names for across the x axis (defaults to alpha ordering) |
| <code>main_title</code> | string for title of the plot (defaults to OG) |

Value

a barplot with taxa vs phenotypic data complete with standard error bars

Examples

```
PDGPlot(pheno_data, mcl_mtrx, 'OG5_126778', 'Treatment', 'RespVar', ylimit=12)
```

 PDGvOG

Number of PDGs vs OGs/PDG

Description

Barplot that indicates the number of PDGs vs OGs(clustered orthologous groups) in a PDG

Usage

```
PDGvOG(mcl_data, num = 40, ...)
```

Arguments

| | |
|----------|---|
| mcl_data | FormatAfterOrtho output |
| num | an integer indicating where the x axis should end and be compiled |
| ... | args to be passed to barplot |

Value

a barplot with a height determined by the second column and the first column abbreviated to accommodate visual spacing

Examples

```
PDGvOG(after_ortho_format_grps, 2)
```

 pheno_data

Triglyceride (TAG) content of fruit flies dataset.

Description

A subset of the TAG content of fruit flies, collected in the Chaston Lab, to be used as a brief example for tests in AnalyzeOrthoMCL.

Usage

```
pheno_data
```

Format

A data frame with 586 rows and 4 variables:

Treatment 4-letter taxa designation of associated bacteria

RespVar response variable, TAG content

Vial random effect variable, vial number of flies

Experiment random effect variable, experiment number of flies

 PhyDataError

Phylogenetic Tree with Attached Bar Plot and Standard Error Bars

Description

Presents data for each taxa including standard error bars next to a phylogenetic tree.

Usage

```
PhyDataError(phy, data, mcl_matrix, species_colname, data_colname,
             color = NULL, OG = NULL, xlabel = "xlabel", ...)
```

Arguments

| | |
|-----------------|--|
| phy | Path to tree file |
| data | R object of phenotype data |
| mcl_matrix | AnalyzeOrthoMCL output |
| species_colname | name of column in data file with taxa designations |
| data_colname | name of column in data file with data observations |
| color | optional parameter, (defaults to NULL) assign colors to individual taxa by providing file (format: Taxa Color) |
| OG | optional parameter, (defaults to NULL) a string with the names of chosen group to be colored |
| xlabel | string to label barplot's x axis |
| ... | argument to be passed from other methods such as parameters from barplot() function |

Value

A phylogenetic tree with a barplot of the data (with standard error bars) provided matched by taxa.

References

Some sort of reference

Examples

```
file <- system.file('extdata', 'muscle_tree2.dnd', package='MAGNAMWAR')
PhyDataError(file, pheno_data, mcl_mtrx, species_colname = 'Treatment', data_colname = 'RespVar',
  OG='OG5_126778', xlabel='TAG Content')
```

PrintOGSeqs

Print OG Sequences

Description

Print all protein sequences and annotations in a given OG

Usage

```
PrintOGSeqs(after_ortho, OG, fasta_dir, out_dir = NULL, outfile = "none")
```

Arguments

| | |
|-------------|--------------------------------------|
| after_ortho | output from FormatAfterOrtho |
| OG | name of OG |
| fasta_dir | directory to fastas |
| out_dir | complete path to output directory |
| outfile | name of file that will be written to |

Value

A fasta file with all protein sequences and ids for a given OG

Examples

```
## Not run:
OG <- 'OG5_126968'
dir <- system.file('extdata', 'fasta_dir', package='MAGNAMWAR')
dir <- paste(dir, '/', sep='')

PrintOGSeqs(after_ortho_format, OG, dir)

## End(Not run)
```

 QQPlotter

QQPlot

Description

Makes a qqplot of the p-values obtained through AnalyzeOrthoMCL

Usage

```
QQPlotter(mcl_mtrx)
```

Arguments

mcl_mtrx matrix generated by AnalyzeOrthoMCL

Value

a qqplot of the p-values obtained through AnalyzeOrthoMCL

References

Some sore of reference

Examples

```
QQPlotter(mcl_mtrx)
```

RASTtoGBK

Write RAST files to Genbank formats OrthoMCL Analysis

Description

Useful for reformatting RAST files to GBK format

Usage

```
RASTtoGBK(input_fasta, input_reference, out_name_path)
```

Arguments

input_fasta path to input fasta file

input_reference

path to a .csv file; it should be downloaded from RAST as excel format, saved as a .csv (saved as the tab-delimited version has compatibility problems)

out_name_path name and path of the file to write to

Examples

```
## Not run:
lfrc_fasta <- system.file('extdata', 'RASTtoGBK//lfrc.fasta', package='MAGNAMWAR')
lfrc_reference <- system.file('extdata', 'RASTtoGBK//lfrc_lookup.csv', package='MAGNAMWAR')
lfrc_path <- system.file('extdata', 'RASTtoGBK//lfrc_out.fasta', package='MAGNAMWAR')

RASTtoGBK(lfrc_fasta,lfrc_reference,lfrc_path)

## End(Not run)
```

| | |
|------------------|--|
| starv_pheno_data | <i>Starvation rate of fruit flies dataset.</i> |
|------------------|--|

Description

A subset of the Starvation rate of fruit flies, collected in the Chaston Lab, to be used as a brief example for survival tests in AnalyzeOrthoMCL.

Usage

```
starv_pheno_data
```

Format

A matrix with 543 rows and 7 variables:

EXP random effect variable, experiment number of flies

VIAL random effect variable, vial number of flies

BACLO fixed effect variable, loss of bacteria in flies

TRT 4-letter taxa designation of associated bacteria

t1 time 1

t2 time 2

event event

SurvAppendMatrix *Append Survival Test Outputs*

Description

Function used to append all .csv files that are outputted from AnalyzeOrthoMCL into one matrix.

Usage

```
SurvAppendMatrix(work_dir, out_name = "surv_matrix.csv", out_dir = NULL)
```

Arguments

| | |
|----------|---|
| work_dir | the directory where the output files of AnalyzeOrthoMCL are located |
| out_name | file name of outputted matrix |
| out_dir | the directory where the outputted matrix is placed |

Value

A csv file containing a matrix with the following columns: OG, p-values, Bonferroni corrected p-values, mean phenotype of OG-containing taxa, mean phenotype of OG-lacking taxa, taxa included in OG, taxa not included in OG

Examples

```
## Not run:  
file <- system.file('extdata', 'outputs', package='MAGNAMWAR')  
directory <- paste(file, '/', sep = '')  
SurvAppendMatrix(directory)  
  
## End(Not run)
```

WriteMCL *Print analyzed matrix*

Description

Writes a tab separated version of the analyzed OrthoMCL data with or without the joined representative sequences

Usage

```
WriteMCL(mtrx, filename)
```

Arguments

| | |
|----------|-------------------------------------|
| mtrx | Matrix derived from AnalyzeOrthoMCL |
| filename | File name to save final output |

Value

The path to the written file

Examples

```
## Not run:  
WriteMCL(mcl_mtrx, 'matrix.tsv')  
#mcl_mtrx previously derived from AnalyzeOrthoMCL() or join_repset()  
  
## End(Not run)
```

Index

* datasets

- after_ortho_format, [2](#)
- after_ortho_format_grps, [3](#)
- joined_mtrx, [8](#)
- joined_mtrx_grps, [9](#)
- mcl_mtrx, [11](#)
- mcl_mtrx_grps, [11](#)
- pheno_data, [13](#)
- starv_pheno_data, [17](#)

- after_ortho_format, [2](#)
- after_ortho_format_grps, [3](#)
- AnalyzeOrthoMCL, [3](#)

- CalculatePrincipalCoordinates, [6](#)

- FormatAfterOrtho, [6](#)
- FormatMCLFastas, [7](#)

- joined_mtrx, [8](#)
- joined_mtrx_grps, [9](#)
- JoinRepSeq, [9](#)

- ManhatGrp, [10](#)
- mcl_mtrx, [11](#)
- mcl_mtrx_grps, [11](#)

- PDGPlot, [12](#)
- PDGvOG, [13](#)
- pheno_data, [13](#)
- PhyDataError, [14](#)
- PrintOGSeqs, [15](#)

- QQPlotter, [16](#)

- RASTtoGBK, [16](#)

- starv_pheno_data, [17](#)
- SurvAppendMatrix, [18](#)

- WriteMCL, [18](#)