

UC Davis

UC Davis Previously Published Works

Title

omu, a Metabolomics Count Data Analysis Tool for Intuitive Figures and Convenient Metadata Collection

Permalink

<https://escholarship.org/uc/item/41w9532k>

Journal

Microbiology Resource Announcements, 8(15)

ISSN

2169-8287

Authors

Tiffany, Connor R

Bäumler, Andreas J

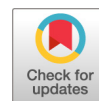
Publication Date

2019-04-11

DOI

10.1128/mra.00129-19

Peer reviewed



omu, a Metabolomics Count Data Analysis Tool for Intuitive Figures and Convenient Metadata Collection

 Connor R. Tiffany,^a Andreas J. Bäuml^a

^aDepartment of Medical Microbiology and Immunology, School of Medicine, University of California at Davis, Davis, California, USA

ABSTRACT Metabolomics is a powerful tool for measuring the functional output of the microbiota. Currently, there are few established workflows for analysis downstream of metabolite identification. Here, we introduce omu, an R package designed for assigning compound hierarchies and linking compounds to corresponding enzyme and gene annotations for organisms of interest.

The omu R package is designed to analyze processed metabolomics count data. The central idea behind omu is assigning hierarchical metadata from the Kyoto Encyclopedia of Genes and Genomes (KEGG) (1) to each metabolite in order to help users create intuitive figures for visualizing their data. To do this, omu provides a suite of graphing and statistical functions centered around the use of `assign_hierarchy`, which provides the metadata for each metabolite based on a KEGG identifier in the user's data. omu comes with an example data set of the fecal metabolome of nitric oxide synthase 2 (NOS2)-deficient C57BL/6J mice 3 days after mock treatment (gavage with sterile water) or oral gavage with a single dose of streptomycin (20 mg/animal). This data set was used to generate the figure in this paper.

Initially, users can use `assign_hierarchy` to provide metadata for each metabolite that has a KEGG compound number provided (Table 1). KEGG data are annotated by compound class, subclass 1, subclass 2, subclass 3, and subclass 4. This annotation provides several options for the users to analyze and visualize their data. For example, users can create a subset to a compound class they are particularly interested in, such as carbohydrates. The user can use the `omu_summary` function to perform a student's *t* test on each metabolite between two experimental groups, provide a fold-change value for each compound, and use these data in conjunction with the hierarchical compound annotation to create intuitive figures. The `count_fold_changes` function can be used to provide a count table of every metabolite within a KEGG compound category that significantly increased or decreased between groups to make bar plots (Fig. 1A) that show how many compounds increased or decreased between experiment groups by a metabolite class or subclass. Alternatively, since `omu_summary` also provides fold-change data, the user can incorporate effect size into a figure by creating a volcano plot using the `plot_volcano` function, which allows the user to highlight points in the plot by the metadata (Fig. 1B).

The omu R package can also help users generate hypotheses from their metabolomics data. The `KEGG_gather` function can get all known enzyme orthology data and gene data associated with each metabolite from the KEGG database as long as the computer is connected to the Internet. This produces high-dimensional data, but `assign_hierarchy` also provides metadata for enzymes and for genes in the form of organism data, allowing the user to reduce the data to items of interest. For example, if a user is interested in genes that *Pseudomonas* spp. have to metabolize organic acids, the `KEGG_gather` function in conjunction with the `assign_hierarchy` function can be used to generate a table containing this information (Table 1).

Citation Tiffany CR, Bäuml AJ. 2019. omu, a metabolomics count data analysis tool for intuitive figures and convenient metadata collection. *Microbiol Resour Announc* 8:e00129-19. <https://doi.org/10.1128/MRA.00129-19>.

Editor Irene L. G. Newton, Indiana University, Bloomington

Copyright © 2019 Tiffany and Bäuml. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Andreas J. Bäuml, ajbauml@ucdavis.edu.

Received 1 February 2019

Accepted 8 March 2019

Published 11 April 2019

TABLE 1 Data table created from the example data set, showcasing the metadata collection that omu can perform

Gene	Metabolite	KEGG no.	Metabolite			Species, strain, serotype	
			count	Class	Subclass 1		Subclass 2
PA4091 (<i>hpaA</i>)	4-Hydroxyphenylacetic acid	C00642	32,307	Organic acids	None	None	<i>Pseudomonas aeruginosa</i> PAO1
N297_4221 (<i>hpaB</i>)	4-Hydroxyphenylacetic acid	C00642	32,307	Organic acids	None	None	<i>P. aeruginosa</i> PAO1, VE13
N296_4221 (<i>hpaB</i>)	4-Hydroxyphenylacetic acid	C00642	32,307	Organic acids	None	None	<i>P. aeruginosa</i> PAO1, VE2
PA14_11000 (<i>hpaA</i>)	4-Hydroxyphenylacetic acid	C00642	32,307	Organic acids	None	None	<i>P. aeruginosa</i> UCBPP, PA14
PSPA7_1007 (<i>hpaB</i>)	4-Hydroxyphenylacetic acid	C00642	32,307	Organic acids	None	None	<i>P. aeruginosa</i> PA7
PP4_31900 (<i>gbd</i>)	4-Hydroxybutyric acid	C00989	315	Organic acids	Carboxylic acids	Hydroxycarboxylic acids	<i>Pseudomonas putida</i> NBRC 14164
APT59_05510	4-Hydroxybutyric acid	C00989	315	Organic acids	Carboxylic acids	Hydroxycarboxylic acids	<i>Pseudomonas oryzae</i> habitaans
PverR02_11545	4-Hydroxybutyric acid	C00989	315	Organic acids	Carboxylic acids	Hydroxycarboxylic acids	<i>Pseudomonas veronii</i>
BJP27_20245	4-Hydroxybutyric acid	C00989	315	Organic acids	Carboxylic acids	Hydroxycarboxylic acids	<i>Pseudomonas psychrotolerans</i>

In summary, omu is a novel metabolomics analysis tool that helps users describe their data by incorporating metabolite metadata into intuitive figures and creating tables with genes and enzymes associated with metabolites of interest.

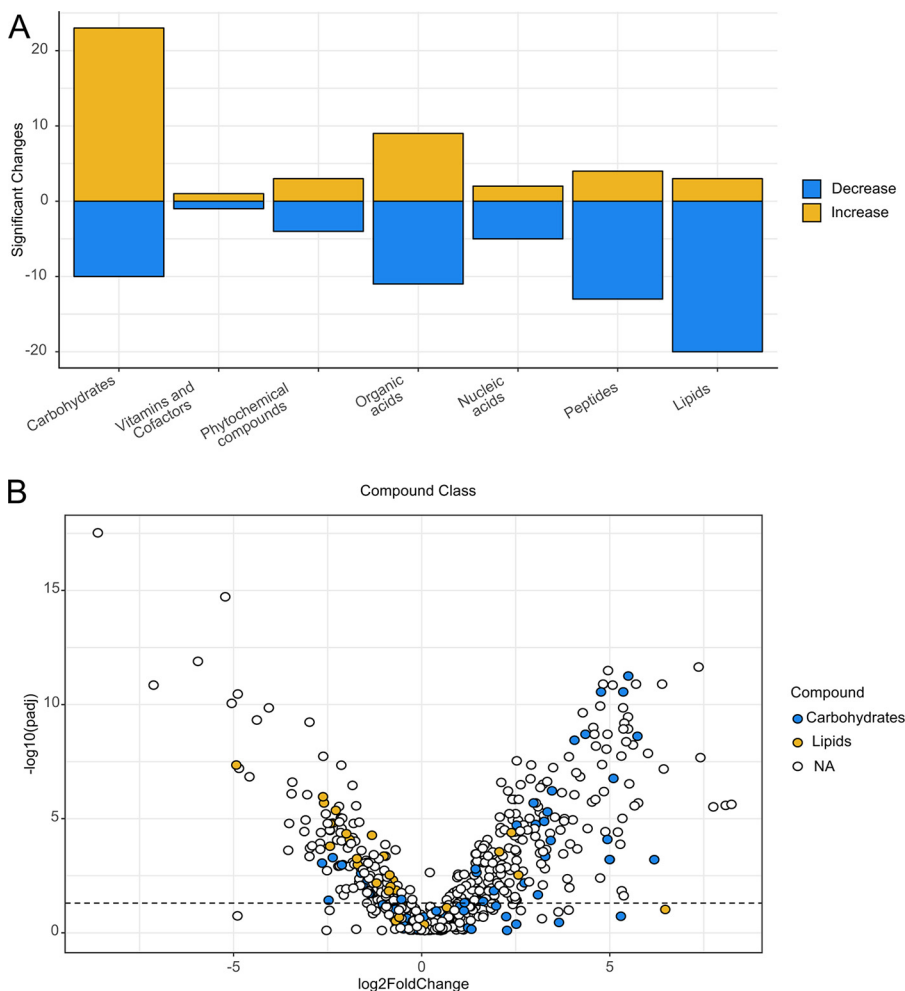


FIG 1 Data visualization in omu. (A) Bar plot showing the number of significantly different metabolites between treatment groups by class from the example data set that comes with the R package. (B) Volcano plot comparing two treatment groups from the example data set. Highlighted points in the plot correspond to metabolite classes.

Data availability. omu is available for download on the CRAN repository (<https://cran.r-project.org/web/packages/omu/index.html>) and was built using R (2), devtools (3), dplyr (4), ggfortify (5), ggplot2 (6), KEGGREST (7), knitr (8), magrittr (9), plyr (10), reshape2 (11), rmarkdown (12), stringr (13), roxygen2 (14), and tidy (15). Detailed

instructions for using omu can be found at https://cran.r-project.org/web/packages/omu/vignettes/Omu_vignette.html.

ACKNOWLEDGMENTS

Work in A.J.B.'s lab is supported by USDA/NIFA award 2015-67015-22930 and by Public Health Service grants AI044170, AI096528, AI112445, and AI112949.

REFERENCES

1. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. 1999. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 27:29–34. <https://doi.org/10.1093/nar/27.1.29>.
2. R Development Core Team. 2016. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.r-project.org/>.
3. Wickham H, Hester J, Chang W. 2018. devtools: tools to make developing R packages easier. <https://rdr.io/cran/devtools/>.
4. Wickham H, Francois R, Henry L, Müller K. 2018. dplyr: a grammar of data manipulation. *R Package version 075*. <https://rdr.io/cran/dplyr/>.
5. Tang Y, Horikoshi M, Li W. 2016. ggfortify: unified interface to visualize statistical results of popular R packages. *R J* 8:474–485. <https://doi.org/10.32614/RJ-2016-060>.
6. Wickham H. 2016. ggplot2: elegant graphics for data analysis. Springer-Verlag, New York, NY.
7. Tenenbaum D. 2018. KEGGREST: client-side REST access to KEGG. <https://rdr.io/bioc/KEGGREST/>.
8. Xie Y. 2018. knitr: a general-purpose package for dynamic report generation in R. <https://rdr.io/cran/knitr/>.
9. Bache SM, Wickham H. 2014. magrittr: a forward-pipe operator for R. R Core Development Team, Vienna, Austria. <https://cran.r-project.org/package=magrittr>.
10. Wickham H. 2011. The split-apply-combine strategy for data analysis. *J Stat Softw* 40:1–29. <https://doi.org/10.18637/jss.v040.i01>.
11. Wickham H. 2012. reshape2: flexibly reshape data: a reboot of the reshape package. *R Package version 1.4.3*. <https://rdr.io/cran/reshape2/>.
12. Allaire JJ, Xie Y, McPherson J, Luraschi J, Ushey K, Atkins A, Wickham H, Cheng J, Chang W. 2018. rmarkdown: dynamic documents for R. <https://rdr.io/cran/rmarkdown/>.
13. Wickham H. 2018. stringr: simple, consistent wrappers for common string operations. <https://rdr.io/cran/stringr/man/stringr-package.html>.
14. Wickham H, Danenberg P, Eugster M. 2014. R: Roxygen2. CRAN.
15. Wickham H, Henry L. 2018. tidyr: easily tidy data with “spread()” and “gather()” functions. <https://tidyr.tidyverse.org/>.