1 # BURRITO: An interactive multi-omic tool for visualizing

2 # taxa-function relationships in microbiome data

3 Colin P McNally[1,†], Alexander Eng[1,†], Cecilia Noecker[1,†], William C

4 Gagne-Maynard[2], and Elhanan Borenstein[1,3,4,]*

5 [1] Department of Genome Sciences, University of Washington, Seattle, Washington, USA

6 [2] Institute for Health Metrics and Evaluation, Seattle, Washington, USA

7 [3] Department of Computer Science and Engineering, University of Washington, Seattle, Washington, USA

8 [4] Santa Fe Institute, Santa Fe, New Mexico, USA

9 [†] These authors contributed equally to this work

10 * Corresponding Author (elbo@uw.edu)

11

Abstract

The abundance of both taxonomic groups and gene categories in microbiome samples can now be easily assayed via various sequencing technologies, and visualized using a variety of software tools. However, the assemblage of taxa in the microbiome and its gene content are clearly linked, and tools for visualizing the relationship between these two facets of microbiome composition and for facilitating exploratory analysis of their co-variation are lacking. Here we introduce *BURRITO*, a web tool for interactive visualization of microbiome multi-omic data with paired taxonomic and functional information. BURRITO simultaneously visualizes the taxonomic and functional compositions of multiple samples and dynamically highlights relationships between taxa and functions to capture the underlying structure of these data. Users can browse for taxa and functions of interest and interactively explore the share of each function attributed to each taxon across samples. BURRITO supports multiple input formats for taxonomic and metagenomic data, allows adjustment of data granularity, and can export generated visualizations as static publication-ready formatted figures. In this paper, we describe the functionality of BURRITO, and provide illustrative examples of its utility for visualizing various trends in the relationship between the composition of taxa and functions in complex microbiomes.

2

Background

31 Microbial communities are complex ecosystems with important impacts on human health and on the

32 environment. High-throughput DNA sequencing has enabled comprehensive profiling of these

33 communities in terms of their composition and structure. Traditionally, microbial ecology studies resort

34 to one of two primary approaches for profiling the composition of a given community, focusing either

35 on its taxonomic composition (e.g., using targeted 16S rRNA gene sequencing or a marker-gene based

36 approach) or on its functional composition (e.g., using metagenomic shotgun sequencing and assessing

37 the abundance of various gene families) (Figure 1A). Obtained taxonomic or functional profiles are then

38 often visualized as simple stacked bar or area plots of relative abundances, or via specialized data

39 visualization tools designed for exploring these data. For example, Explicet (Robertson et al., 2013) and

40 Krona (Ondov et al., 2011) aid analysis by displaying abundance data while simultaneously presenting

41 the hierarchical relationships between entities. Other tools have gone beyond relative abundance

42 visualization to enable exploration of specific, comparative aspects of microbiome data. EMPeror

43 (Vázquez-Baeza et al., 2013), for example, allows users to generate 3D principal coordinate analysis

44 plots to visualize clustering of, or variation in, taxonomic compositions coupled with trends in any

45 associated metadata. Other tools, including Community-Analyzer (Kuntal et al., 2013) and MetaCoMET

46 (Wang et al., 2016), provide access to multiple types of visualizations of the same microbiome data,

47 each highlighting different aspects of community structure or between-sample relationships.

48   Importantly, however, recent years have witnessed an explosion of microbiome *multi-omic*

49 studies that aim to describe simultaneously multiple aspects of community structure, including

50 specifically both taxonomic and functional compositions (Greenblum et al., 2015; Huttenhower et al.,

51 2012; Lloyd-Price et al., 2017; Pedersen et al., 2016; Taxis et al., 2015; Zhernakova et al., 2016).

52 Moreover, recently developed methods can now determine both the taxonomic and functional profile

53 of a given community from the same sequencing data, for example, by assigning shotgun metagenomic

3

54   reads both taxonomic and functional annotations (Abubucker et al., 2012). Notably, these two facets of

55   microbiome composition are not independent since the set of genes found in a metagenome and their

56   abundances is a direct result of the set of genes (and their copy number) encoded by each community

57   member and the relative abundance of each member in the community (Figure 1B). Put differently, the

58   abundance of each gene family (or 'function') in the metagenome can be deconvolved into taxon-

59   specific functional profiles in which shares of the gene family's total abundance are attributed to specific

60   taxa of origin (Carr, Shen-Orr, & Borenstein, 2013)(Figure 1C). This link between taxonomic and

61   functional compositions can be used, for example, to predict functional abundances from 16s rRNA-

62   based taxonomic profiles (Langille et al., 2013) or to identify taxonomic drivers of disease-associated

63   functional shifts (Manor and Borenstein, 2017b). Importantly, this inherent relationship between

64   taxonomic and functional profiles must also be considered when exploring how functional capacity co-

65   varies with taxonomic composition across samples, since differences in gene abundances between

66   samples are mainly derived from differences in taxonomic composition (Bäckhed et al., 2015; Oh et al.,

67   2014; Turnbaugh et al., 2009). Yet, despite the growing appreciation for this link between taxonomic

68   and functional compositions, an integrative tool that can simultaneously visualize both taxonomic and

69   functional data and that can *account for* and *expose* the relationships between taxonomic and functional

70   variation is lacking.

71   Here we introduce BURRITO (Browser Utility for Relating micRobiome Information on Taxonomy

72   and functiOn), a web-based visualization tool that enables easy and intuitive exploratory analysis of the

73   relationships between taxonomic and functional abundances across microbiome samples. BURRITO

74   simultaneously provides a traditional interface for exploring taxonomic and functional abundances

75   independently while also visualizing the links between these two microbiome facets and highlighting the

76   share of each function's total abundance that is attributed to each taxon (Figure 1D). Through an

77   interactive interface, BURRITO also provides ample and precise information about such *attributions* (*i.e.*,

4

78   the share of each function's total abundance attributed to each taxon), as well as various summary

79   statistics. To facilitate interactive data exploration and publication-quality figure generation for a wide

80   audience, BURRITO further offers multiple options for data input and supports customizing various

81   aspects of the visualization.

Methods and Implementation

82

User input and taxa-function mapping

83

84   BURRITO accommodates multiple types of input data and, depending on the provided data, uses

85   different approaches to attribute the provided or inferred function abundances to taxa of origin (see

86   Figure 1D). Specifically, the user can select one of three options for input data and for determining taxa-

87   function attributions (Figure 2). In the first option, which requires the bare minimum in terms of input

88   data, the user can simply provide a table of taxonomic abundances across samples (measured as either

89   absolute read counts or relative abundances) using Greengenes 97% Operational Taxonomic Unit (OTU)

90   IDs for each taxon (DeSantis et al., 2006). Given these taxonomic data, BURRITO will automatically

91   predict the functional profile of each sample and will determine each function's taxonomic attributions

92   using a database of pre-annotated genomic content (following the approach described in Figure 1B-C).

93   Briefly, in this approach, taxonomic abundances for each sample are first corrected by dividing each

94   taxon's abundance by its 16S rRNA copy number. The abundance of a given function that is attributed to

95   each taxon (and ultimately the total abundance of that function) is then inferred by multiplying the

96   corrected taxonomic abundances by the number of genes associated with that function in each taxon's

97   genomes. Data on the gene content in each taxon's genome is obtained from PICRUSt (Langille et al.,

98   2013b) and functional annotations are based on KEGG Orthology groups (Kanehisa et al., 2015; Kanehisa

99   and Goto, 2000).

100   The second available option is similar to the one described above, but relies on user-provided

101   genomic content annotations (instead of PICRUSt-based inferred content) to calculate functional profiles

102   and the share of the functional profile attributed to each taxon. This approach is appropriate, for

103   example, for exploring communities whose members may not be adequately represented by PICRUSt-

104   inferred genomes but can be better characterized based on user proprietary data. As in the first option,

105   the user is required to provide a taxonomic abundance table, but also provides a custom genomic

106   content table describing the set of genes (and their copy number) encoded by each taxon. Given these

107   data, functional abundances and attributions are calculated in the same manner as described above.

108   When using this option, it is also assumed that taxonomic abundances are already corrected for 16S

109   rRNA copy number. Moreover, using this approach the user is not limited to Greengenes OTU IDs or to

110   KEGG Orthology groups, and alternately can use their taxonomic and/or functional classification of

111   choice (as long as the same IDs are used in all relevant input files). Notably, when using custom

112   taxonomic classification or custom hierarchical function relationships, the user can also provide files

113   describing these custom systems to allow taxonomic and functional data to be grouped at different

114   levels (see section 'Exploring attributions at different taxonomic and functional levels' below).

115   The third and final option relies on a pre-determined table of taxon-specific functional

116   attributions rather than calculating attributions from taxonomic composition. This approach may be

117   appropriate, for example, in cases where the user wishes to introduce specific custom modifications to a

118   pre-calculated attribution table. Using this option requires a taxonomic abundance table (as above) and,

119   instead of a genomic content table, a table of taxa-specific functional abundance attributions. As in the

120   second approach, any taxonomic classification or functional hierarchy system can be used.

121   The specific format for each data input file and the specific restrictions associated with each

122   approach are all noted on BURRITO's upload page and are described in more detail in BURRITO's

123   documentation. Example data files are also available for download from the upload page. Notably, in all

124   three approaches, the user can also upload an independent, paired dataset of functional abundance

125   profiles (i.e. based on functional annotation of shotgun metagenomic sequencing), which will be

126   visualized as described below alongside the calculated taxa-based functional profiles.

127   In addition to the above primary input, BURRITO's upload page (Figure 2) further includes

128   several visualization options, allowing users to better control the way data will be displayed. Specifically,

129   BURRITO supports grouping samples based on user-provided labels (e.g., cases vs. control or conditions).

130   Additionally, the user can select the minimum taxonomic and functional resolution to be displayed.

131   Using a finer resolution allows exploring the data in more depth, but could slow visualization

132   performance due to the large number of elements that need to be calculated, managed, and displayed.

133   Finally, users can choose between hierarchical or random color schemes to distinguish taxa and

134   functions.

135   **Visualizing the relationship between taxonomic and functional abundances**

136   BURRITO uses two stacked bar plots, one for taxonomy (Figure 3A) and one for function (Figure 3B), to

137   provide a standard visualization of taxonomic and functional relative abundances in each sample.

138   Taxonomic abundances are taken from the user-provided taxonomic abundance table. Functional

139   abundances are calculated as described in the previous sections and are displayed as the sums of all

140   taxon-specific attributions for each function and sample. Precise abundance values for each taxon or

141   function in each sample can be viewed in a *tooltip* that appears when the user hovers over any bar

142   segment (Figure 3C). Additionally, hovering over a bar segment highlights the corresponding taxon's or

143   function's bar segment across all samples to aid visual comparison of abundances between samples

144   (Figure 3D).

145 The most innovative component of BURRITO is the visualization of how function abundance

146 shares are attributed to the various taxa. This information is revealed when the user hovers over a bar

147 segment in the taxonomic abundance bar plot, which, in addition to highlighting taxon abundances as

148 noted above, also highlights the portion of each function abundance bar segment (in each sample) that

149 is attributed to this taxon (Figure 3E). To view the exact function abundance share attributed to a given

150 taxon, the user can click on (rather than hover over) a taxon's bar segment to lock this taxon-specific

151 attribution highlighting, and then hover over the highlighted portion of a function bar segment,

152 revealing a tooltip with the corresponding information (Figure 3F).

153 BURRITO also displays a "control panel" that can be used to investigate specific taxa and/or

154 functions and explore average abundances across samples (Figure 3G). Specifically, taxa and functions

155 are represented by bars on the left and right sides of a bipartite graph. Hovering over the control panel

156 performs similar highlighting as the bar plots, highlighting the linked abundances of individual taxa and

157 functions and displaying (via a tooltip) the average relative abundance of each taxon or function.

158 Moreover, when a taxon is highlighted, edges that link that taxon to all the functions it encodes are

159 displayed, providing an easy reference for identifying the functions with shares attributed to that taxon.

160 Similarly, when a function is highlighted, edges that indicate which taxa encode that function (and hence

161 have shares of that function attributed to them) are displayed. The width of an edge between a taxon

162 and a function represents the average share of that function's abundance that is attributed to that taxon

163 across all samples (Figure 3H). Clicking on a specific taxon or function in the control panel (or bar plots)

164 locks the selection of that taxon or function, allowing the user to hover over each edge and view (via a

165 tooltip) the exact taxon-function attribution values (Figure 3I). Additionally, when a taxon or function

166 selection has been clicked (i.e., selected) and edges connecting taxa and functions are displayed, the

167 user can highlight the abundances of a single function attributed to a single taxon by clicking on the

168 edge between them.

8

169   BURRITO also supports comparison between taxa-based function abundances (calculated based

170   on taxonomic profiles and genomic content as noted above) and a separate dataset of user-provided

171   function abundances (typically obtained by functional annotation of shotgun metagenomic sequencing

172   reads). If such a dataset is included in the input, these user-provided function abundances are displayed

173   adjacent to the taxa-based function abundances for comparison (Figure 3J).

174   **Exploring attributions at different taxonomic and functional levels**

175   BURRITO provides an interactive and intuitive interface for exploring abundance and attribution data at

176   varying taxonomic and functional resolutions. Taxonomic classification and hierarchical function

177   relationships (either those used by default or custom systems provided by the user as noted above) are

178   each represented as a tree above the corresponding abundance bar plot (Figure 3K and 3L, respectively).

179   To aid the user in understanding how different taxa or functions are related in the bipartite graph and

180   bar plots, these trees are aligned to the left and right of the bipartite graph with all taxa and functions

181   appearing in the same vertical order across all components of the visualization. These trees also indicate

182   average taxon and function abundances across samples by the size of leaf nodes in the trees (Figure

183   3M). Beyond visualizing hierarchical relationships, these trees also provide a tool for interactive data

184   exploration by allowing the user to expand or collapse different leaves or branches of each tree. Clicking

185   on a leaf node reveals all taxonomic or functional subcategories of that node in the tree, and

186   correspondingly expands the bipartite graph and subdivides the relevant relative abundance bars in the

187   bar plot into the relative abundances of those subcategories. Alternatively, clicking on a non-leaf node

188   within a tree performs the reverse operation, collapsing all visible descendants of that node, making the

189   clicked node a leaf node, and aggregating the abundance bars for those descendants into the abundance

190   bars for the clicked node. Importantly, these interactive features allow the user to dynamically drill up or

191   down in both taxonomic and functional resolution across the different branches of either tree as they

192   explore the data.

193 **Exporting visualization plots and processed data**

194 BURRITO also provides options for exporting a static version of the visualization (e.g., for including in

195 presentations or publications) and for downloading the function abundance table or attribution table

196 underlying the displayed function plot. These options can be accessed from the visualization screen via a

197 hidden menu (Figure 3N). Exported figures will maintain any currently-selected highlighting and

198 taxonomic or functional tree expansion. In addition to exporting the full visualization, users can choose

199 to individually export either bar plot of relative abundances and either half of the bipartite graph, which

200 can serve as a legend for the color-coding of taxa or functions in the bar plots. All images can be

201 exported in PNG or SVG format. If users wish to further explore the predicted function abundance or

202 attribution data in more detail, they can also download the tables underlying the visualization. Function

203 abundance and attribution tables can be downloaded at the minimum functional resolution (and

204 minimum taxonomic resolution for the attribution table) specified on the upload page.

205 **Technical Implementation**

206 BURRITO's client is a browser page written in HTML and Javascript, utilizing the d3.js library to display

207 data. User-submitted data are uploaded to an R Shiny server for processing (including, specifically,

208 calculation of attributions) and the results are sent back to the browser for visualization. Additional

209 details concerning BURRITO implementation can be found in the Supplementary Text.

Case Studies & Discussion

211 To demonstrate the utility of BURRITO, we describe below its application to two microbiome datasets

212 with varying properties.

213 **Case study 1: Exploring the effects of antibiotic treatment and recovery on inferred**

214 **functional composition in the mouse cecum**

215 We used BURRITO to visualize a publicly available dataset of 16S rRNA sequencing data, describing cecal

216 samples from mice treated with antibiotics 2 days and 6 weeks after treatment (labeled 'Abx Day 2' and

217 'Abx Day 42', respectively), and time-matched controls (labeled 'Control Day 2' and 'Control Day 42',

218 respectively) (Theriot et al., 2014). This study, which focused on associations of the microbiome with

219 metabolomic data and colonization resistance, confirmed significant community perturbations in

220 response to antibiotics. This dataset is also used in BURRITO's *Preview* option on the upload page (see

221 Figure 2), allowing users to examine BURRITO's visualization and functionality (and to compare those to

222 the examples provided in this case study) without the need to provide any additional data.

223 Given this dataset, we used the first input approach described above, allowing BURRITO to

224 predict functional abundances (and taxon-specific attributions) based on the default PICRUSt-derived

225 genomic content table. BURRITO's visualization of taxonomic and functional profiles revealed relatively

226 subtle functional variation despite drastic taxonomic variation across samples, a pattern commonly

227 observed in microbiome studies (Manor and Borenstein, 2017a). Specifically, while Abx Day 2 samples

228 are markedly different from, for example, Control Day 2 samples (with the former being dominated by

229 species from the class *Bacilli* and the latter by species from the classes *Clostridia* and *Bacteroidia*), their

230 predicted functional profiles are relatively similar (Figure 4A). Hovering over the various taxa in the

231 taxonomic profiles highlighted the shares of the functional profile in the various sample groups that are

232 attributed to microbes from these three classes (*Bacilli*, *Clostridia,* and *Bacteroidia*), demonstrating, for

233 example, that attributions to the *Bacteroidia* species were relatively small compared to their abundance.

234 See, for instance, sample NonAbx29, in which *Bacteroidia* is more abundant than *Clostridia* (54.93% vs.

235 44.13%), but has a smaller share attributed to it than to *Clostridia* in every functional category (Figure

236 4B). To further illustrate BURRITO's functionality, we then visually searched for functions that are

237   nonetheless differentially abundant between Abx Day 2 samples and other samples, focusing specifically

238   on pathways in the metabolism category. We observed, for example, that Abx Day 2 samples were

239   generally depleted for amino acid metabolism genes (Figure 4C). Selecting this pathway (by clicking on it

240   in the bar plot or in the control panel) and examining the share of this pathway attributed to each taxon

241   (by then clicking on the edge connecting this pathway to each taxon in the control panel), suggested

242   that its depletion in Abx Day 2 samples could be explained by the fact that *Bacilli* contribute less than

243   *Clostridia* to this pathway compared to their abundances. Specifically, we noted that the share of this

244   pathway attributed to *Bacilli* in Abx Day 2 samples is smaller than the share of this pathway attributed to

245   *Clostridia* in Abx Day 42 samples, even though the abundance of *Bacilli* and *Clostridia* in Abx Day 2 and in

246   Abx Day 42 samples, respectively, is comparable (close to 100%) (Figure 4D). Similarly, the share of this

247   pathway attributed to *Bacilli* in Abx Day 2 samples is comparable to the share of this pathway attributed

248   to *Clostridia* in Control samples, even though the abundance of *Bacilli* in Abx Day 2 samples is higher

249   than the abundance of *Clostridia* in control samples. This lower proportional contribution indicates a

250   lower number of genes involved in this pathway in *Bacilli* compared to *Clostridia*. Lastly, we searched for

251   additional functions with higher or lower share attributed to *Bacilli* by selecting this taxon (again, by

252   hovering over or clicking on it in the control panel or in the bar plot) and examining the width of the

253   attribution edges connecting it to each function. In this setting, it was easy to note that a relatively small

254   share of the cell motility function is attributed to this taxon, compared to, for example, the shares of

255   metabolic functions attributed to it (Figure 4E).

256   **Case Study 2: Taxa-function relationships in the Human Microbiome Project**

257   We additionally used BURRITO to visualize data from 21 supragingival plaque samples with both 16S

258   rRNA and shotgun metagenomic data downloaded from the Human Microbiome Project (Huttenhower

259   et al., 2012). We first used the 16S rRNA data alone (*i.e.*, again using the first input approach), examining

260   inferred functional profiles and the taxa they are attributed to. As noted above, expanding the

261   taxonomic tree can provide additional details about specific genera to which each function is attributed.

262   Similarly, expanding the functional tree can offer insights into differentially abundant pathways and

263   subpathways. For example, drilling down into subpathways in the Environmental Information Processing

264   category and examining the average share from each subpathway attributed to the various phyla, we

265   noted that the abundance of ABC transporter genes is attributed primarily to Actinobacteria, (average

266   attribution across samples 24.87% of this function abundance), Firmicutes (24.65%), and Proteobacteria

267   (29.55%). Indeed, samples with high relative abundance of these phyla tended to have higher

268   abundance of this subpathway. Similarly, examining the shares of pyruvate metabolism (a subpathway

269   of carbohydrate metabolism) attributed to genera from the phylum Proteobacteria revealed a

270   specifically large attribution of the genus Neisseria, a prominent acid producer in the oral microbiome

271   (Figure 5A). This is consistent with the capacity of strains from the genus Neisseria to metabolize lactate

272   (via reactions included in the pyruvate metabolism pathway) (Hoshino and Araya, 1980; McLean et al.,

273   2012).

274   Finally, Figure 5B demonstrates how BURRITO can also be used to compare such amplicon-

275   based inferred functional profiles with functional abundance profiles obtained directly from shotgun

276   metagenomic sequencing (when such functional profiles are provided as an additional input file). As

277   expected, shotgun metagenomic-based profiles are generally in agreement with taxa-based inferred

278   functional profiles, yet some differences can be observed, for example in the abundance of the genes

279   related to the metabolism of cofactors and vitamins (highlighted).

Conclusions

281   BURRITO is a web-based tool that addresses a major gap in currently available visualization tools for

282   microbial ecology research. Studies in this field typically analyze, explore, and visualize taxonomic and

283   functional abundances separately and fail to account for the interdependence between the two,

13

284   potentially due to a shortage of tools that support simultaneous and integrative study of taxonomic and

285   functional profiles. Our tool enables data exploration and hypothesis generation based on the

286   attribution of functional abundances to specific taxa, providing a novel view into microbiome variation

287   and dynamics.

Availability and requirements

289   • **Project name:** BURRITO

290   • **Project home page:** http://elbo.gs.washington.edu/software_burrito.html

291   • **Operating system:** Platform independent

292   • **Programming language:** HTML 5, JavaScript, D3

293   • **Other requirements:** None

294   • **Compatible browsers:** Chrome version 61.0+, Firefox version 53.0+, Opera version 48.0+, Safari

295   11.0+ (PNG exports not supported in Safari)

296   • **License:** Non-commercial research use only (available at https://github.com/borenstein-lab/Burrito)

List of abbreviations

298   BURRITO (Browser Utility for Relating micRobiome Information on Taxonomy and functiOn); OTU

299   (Operational Taxonomic Unit); KEGG (Kyoto Encyclopedia of Genes and Genomes)

Competing interests

301   None

Authors' contributions

303   CPM, AE, CN, and EB conceived and designed this study; CPM, AE, CN, and WGM implemented a

304   preliminary version of BURRITO; CPM, AE, and CN completed the implementation of BURRITO; CPM, AE,

305   CN, and EB wrote the paper. All authors read and approved this manuscript.

14

306 ## Acknowledgements

311

References

312

313 Abubucker, S., Segata, N., Goll, J., Schubert, A. M., Izard, J., Cantarel, B. L., et al. (2012). Metabolic
314   reconstruction for metagenomic data and its application to the human microbiome. *PLoS*
315   *Comput. Biol.* 8, e1002358. doi:10.1371/journal.pcbi.1002358.

316 Bäckhed, F., Roswall, J., Peng, Y., Feng, Q., Jia, H., Kovatcheva-Datchary, P., et al. (2015). Dynamics and
317   Stabilization of the Human Gut Microbiome during the First Year of Life. *Cell Host Microbe* 17,
318   690–703. doi:10.1016/j.chom.2015.04.004.

319 Brewer, C. A. (2017). ColorBrewer: A web tool for selecting colors for maps. Available at:
320   colorbrewer2.org [Accessed October 9, 2017].

321 Carr, R., Shen-Orr, S. S., and Borenstein, E. (2013). Reconstructing the genomic content of microbiome
322   taxa through shotgun metagenomic deconvolution. *PLoS Comput. Biol.* 9, e1003292.
323   doi:10.1371/journal.pcbi.1003292.

324 DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., et al. (2006). Greengenes, a
325   chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ.*
326   *Microbiol.* 72, 5069–72. doi:10.1128/AEM.03006-05.

327 Greenblum, S., Carr, R., and Borenstein, E. (2015). Extensive Strain-Level Copy-Number Variation across
328   Human Gut Microbiome Species. *Cell* 160, 583–594. doi:10.1016/j.cell.2014.12.038.

329 Hoshino, E., and Araya, A. (1980). Lactate degradation by polysaccharide-producing Neisseria isolated
330   from human dental plaque. *Arch. Oral Biol.* 25, 211–212. doi:10.1016/0003-9969(80)90023-0.

331 Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J. H., Chinwalla, A. T., et al. (2012).
332   Structure, function and diversity of the healthy human microbiome. *Nature* 486, 207–214.
333   doi:10.1038/nature11234.

334 Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.*
335   28, 27–30. doi:10.1093/nar/28.1.27.

336 Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2015). KEGG as a reference
337   resource for gene and protein annotation. *Nucleic Acids Res* 44, D457--D462.
338   doi:10.1093/nar/gkv1070.

339 Kuntal, B. K., Ghosh, T. S., and Mande, S. S. (2013). Community-Analyzer: A platform for visualizing and
340   comparing microbial community structure across microbiomes. *Genomics* 102, 409–418.
341   doi:10.1016/j.ygeno.2013.08.004.

342 Langille, M. G. I., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., et al. (2013a).
343   Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences.
344   *Nat. Biotechnol.* 31, 814–21. doi:10.1038/nbt.2676.

345 Langille, M. G. I., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., et al. (2013b).
346   Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences.
347   *Nat. Biotechnol.* 31, 814–21. doi:10.1038/nbt.2676.

348 Lloyd-Price, J., Mahurkar, A., Rahnavard, G., Crabtree, J., Orvis, J., Hall, A. B., et al. (2017). Strains,
349   functions and dynamics in the expanded Human Microbiome Project. *Nature*.
350   doi:10.1038/nature23889.

351 Manor, O., and Borenstein, E. (2017a). Revised computational metagenomic processing uncovers hidden
352   and biologically meaningful functional variation in the human microbiome. *Microbiome* 5.
353   doi:10.1186/s40168-017-0231-4.

354 Manor, O., and Borenstein, E. (2017b). Systematic Characterization and Analysis of the Taxonomic
355   Drivers of Functional Shifts in the Human Microbiome. *Cell Host Microbe*.
356   doi:10.1016/j.chom.2016.12.014.

357 McLean, J. S., Fansler, S. J., Majors, P. D., McAteer, K., Allen, L. Z., Shirtliff, M. E., et al. (2012). Identifying
358   Low pH Active and Lactate-Utilizing Taxa within Oral Microbiome Communities from Healthy
359   Children Using Stable Isotope Probing Techniques. *PLoS ONE* 7, e32219.
360   doi:10.1371/journal.pone.0032219.

361 Oh, J., Byrd, A. L., Deming, C., Conlan, S., Barnabas, B., Blakesley, R., et al. (2014). Biogeography and
362   individuality shape function in the human skin metagenome. *Nature* 514, 59–64.
363   doi:10.1038/nature13786.

364 Ondov, B. D., Bergman, N. H., and Phillippy, A. M. (2011). Interactive metagenomic visualization in a
365   Web browser. *BMC Bioinformatics* 12, 385. doi:10.1186/1471-2105-12-385.

366 Pedersen, H. K., Gudmundsdottir, V., Nielsen, H. B., Hyotylainen, T., Nielsen, T., Jensen, B. A. H., et al.
367   (2016). Human gut microbes impact host serum metabolome and insulin sensitivity. *Nature*
368   advance online publication. doi:10.1038/nature18646.

369 Robertson, C. E., Harris, J. K., Wagner, B. D., Granger, D., Browne, K., Tatem, B., et al. (2013). Explicet:
370   graphical user interface software for metadata-driven management, analysis and visualization of
371   microbiome data. *Bioinformatics* 29, 3100–3101. doi:10.1093/bioinformatics/btt526.

372 Taxis, T. M., Wolff, S., Gregg, S. J., Minton, N. O., Zhang, C., Dai, J., et al. (2015). The players may change
373   but the game remains: network analyses of ruminal microbiomes suggest taxonomic differences
374   mask functional similarity. *Nucleic Acids Res.*, gkv973. doi:10.1093/nar/gkv973.

375 Theriot, C. M., Koenigsknecht, M. J., Carlson Jr, P. E., Hatton, G. E., Nelson, A. M., Li, B., et al. (2014).
376   Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility
377   to Clostridium difficile infection. *Nat. Commun.* 5. doi:10.1038/ncomms4114.

378 Turnbaugh, P. J., Hamady, M., Yatsunenko, T., Cantarel, B. L., Duncan, A., Ley, R. E., et al. (2009). A core
379   gut microbiome in obese and lean twins. *Nature* 457, 480–484. doi:10.1038/nature07540.

380 Vázquez-Baeza, Y., Pirrung, M., Gonzalez, A., and Knight, R. (2013). EMPeror: a tool for visualizing high-
381   throughput microbial community data. *GigaScience* 2, 16. doi:10.1186/2047-217X-2-16.

382 Wang, Y., Xu, L., Gu, Y. Q., and Coleman-Derr, D. (2016). MetaCoMET: a web platform for discovery and
383   visualization of the core microbiome. *Bioinformatics* 2, btw507.
384   doi:10.1093/bioinformatics/btw507.

17

385   Zhernakova, A., Kurilshikov, A., Bonder, M. J., Tigchelaar, E. F., Schirmer, M., Vatanen, T., et al. (2016).
386   Population-based metagenomics analysis reveals markers for gut microbiome composition and
387   diversity. *Science* 352, 565–569. doi:10.1126/science.aad3369.

388

389

390

Figure Legends

391

392 **Figure 1.** The taxonomic and functional compositions of a microbiome are inherently linked. **A)** Typical

393 microbiome studies quantify and report the taxonomic (colored shapes) and functional (letters) profiles

394 of a given community as separate entities. **B)** The functional profile of a community is a linear

395 combination of the taxonomic composition and the genomic content of each taxon. **C)** Functional

396 profiles can be deconvolved into taxon-specific functional profile, denoting which share of the

397 abundance of each function is *attributed* to each taxon. **D)** Such deconvolved functional profiles can be

398 visualized, illustrating the total abundance of each function as a stacked bar of taxon-specific

399 attributions.

400 **Figure 2.** BURRITO's upload page, describing the various input approaches BURRITO supports and other

401 visualization and analysis options.

402 **Figure 3.** A layout of BURRITO's visualization. **A, B)** Stacked bar plots of taxonomic and functional

403 composition across samples. **C)** Tooltips appear when hovering over each taxon, providing information

404 about this taxon's relative abundance in each sample. **D)** Interactive highlighting of individual taxa,

405 which correspondingly highlights the shares of functional abundances that can be attributed to the

406 taxon in question **(E)**. **F)** Tooltips provide detailed function abundance and attribution data for each

407 sample. **G)** The bipartite graph control panel identifies individual taxa and functions and shows links

408 between them. **H)** Edge width in the control panel represents the *average* share of a function that is

409 attributed to a given taxon. **I)** Exact taxon-function attribution values can be seen by hovering over the

410 edge connecting the taxon to the function. **J)** An independent dataset of shotgun metagenomics-derived

411 function abundances can be provided and displayed alongside inferred taxon-specific function

412 abundances. **K, L)** The data can be viewed at a higher or lower taxonomic or functional resolution by

413 clicking on nodes in the corresponding tree diagrams. **M)** The size of each node in the tree represents

19

414   the average abundance of that entity. **N)** Opening the hidden menu allows users to export visualization

415   plots and processed data.

416   **Figure 4.** Using BURRITO to visualize the functional impacts of microbiome shifts in response to

417   antibiotics. **A)** An overall view of BURRITO's display for this dataset of 29 mouse cecal samples. **B)**

418   Selecting a taxon (e.g., *Bacteroidia*) displays the share of each function attributed to this taxon. Tooltips

419   provide exact attribution values. **C)** Expanding functional resolution and clicking on a given function

420   (e.g., Amino Acid Metabolism) highlights this function in each sample. Tooltips provide exact abundance

421   values. **D)** Once a function is selected, taxon-function edges in the control panel can be clicked,

422   displaying the average share of the function attributed to the selected taxon and the sample-specific

423   share of the function attributed to the selected taxon. **E)** Edges connecting a given taxon to all encoded

424   functions provide information about the average share of each function attributed to this taxon.

425   **Figure 5.** Taxa-function relationships across supragingival plaque samples from the Human Microbiome

426   Project. **A)** Taxa and functions can be explored at different levels, using the trees in the control panel to

427   expand taxa and functions of interest. This allows users to examine shares of specific functions that are

428   attributed to specific taxa at high resolution. Edges in the control panel display information about the

429   average share of a function attributable to each taxon across samples (with exact values provided via

430   tooltips). **B)** Inferred taxa-based functional abundances (linked to 16S rRNA taxonomic data, T) can be

431   displayed alongside measured functional relative abundance data obtained from metagenomic shotgun

432   sequencing (M) for easy comparison.
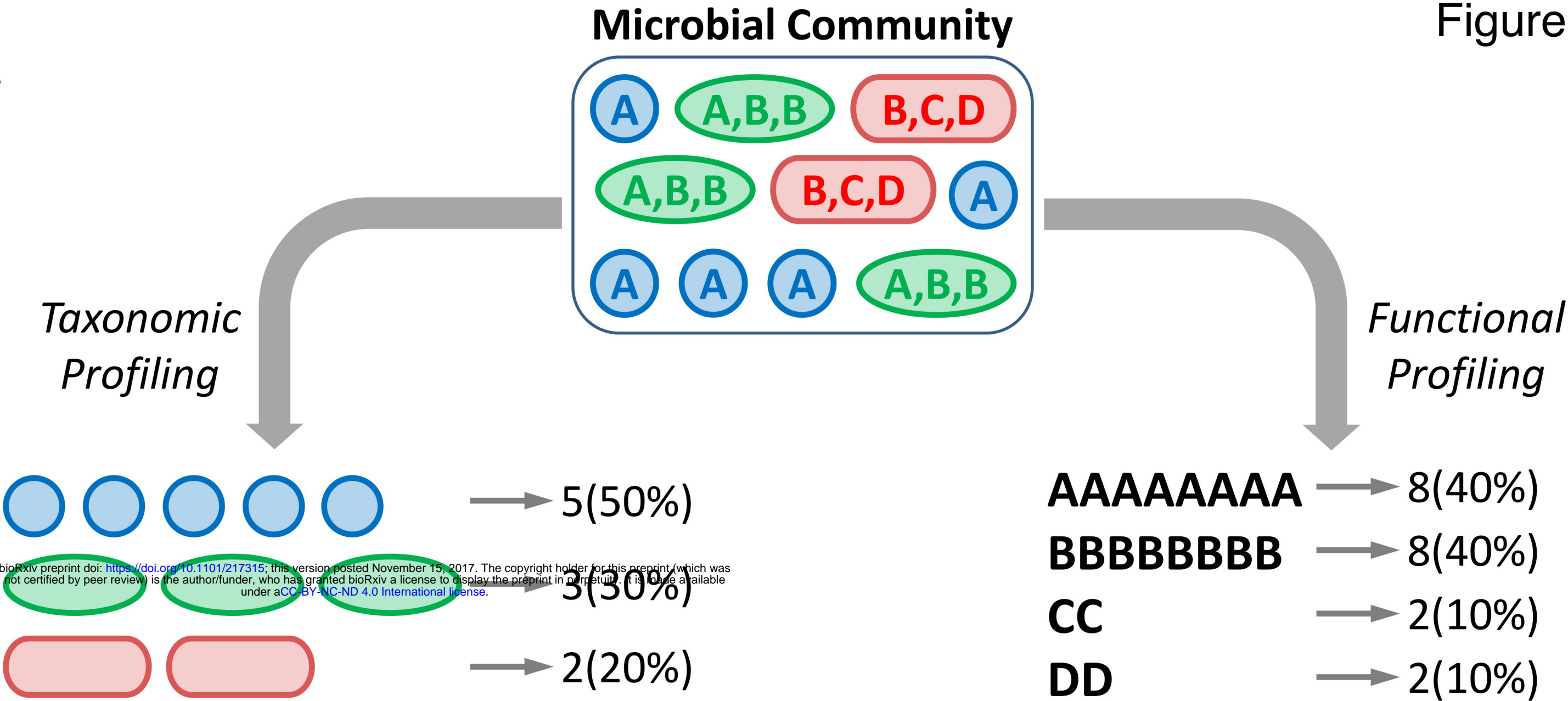
433

434
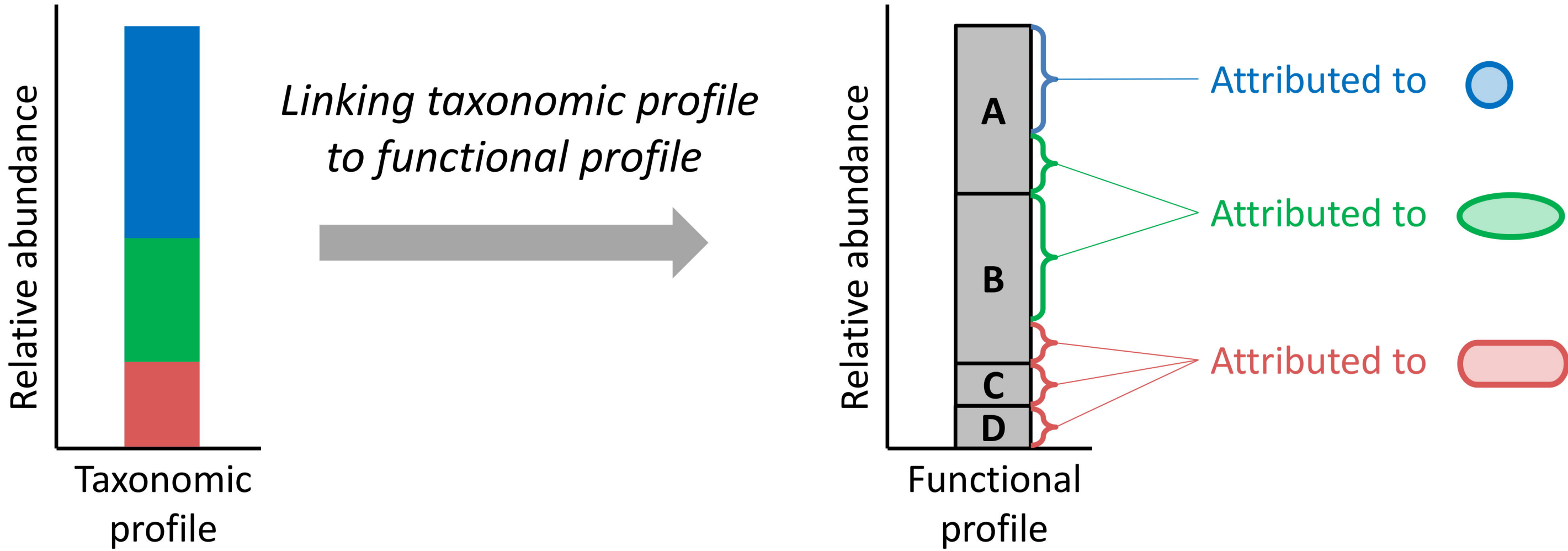
Figure 1

Burrito

Figure 2

Burrito is a tool for visualizing the relationship between taxonomy and function in microbiome datasets. A detailed description of the input parameters and options is available in the user manual. Source code is available <u>here</u>

Taxonomic Data   Manual>Taxonomy

Taxonomic abundances: ⑦   [Choose File] No file chosen   (≡ example)

Taxonomy: ⑦   ⦿ Greengenes taxonomy (default; compatible with automatic attribution calculation)
                   ○ Custom taxonomy: [Choose File] No file chosen   (≡ example)

Minimum taxonomic resolution: ⑦   [Genus ▾]

Functional Data   Manual>Function

Metagenome-based function abundances: ⑦   ⦿ No metagenome-based function abundances available (default)
                   ○ Metagenome-based function abundance table:
                   [Choose File] No file chosen   (≡ example)

Taxonomy-function linking method: ⑦   ⦿ Automatically calculate attributions (default; uses the 13_5 versions of the 16S copy number normalization and gene content tables downloaded from the PICRUSt website)
                   ○ Custom genomic content table: [Choose File] No file chosen   (≡ example)
                   ○ Function attribution table: [Choose File] No file chosen   (≡ example)

Function hierarchy: ⑦   ⦿ BRITE function hierarchy (default, compatible with automatic attribution calculation)
                   ○ Custom function hierarchy: [Choose File] No file chosen   (≡ example)

Minimum functional resolution: ⑦   [SubPathway ▾]

Sample grouping   Manual>Grouping

Sample grouping: ⑦   ⦿ No sample grouping (default) ☐ Sort samples alphabetically
                   ○ Sample group labels: [Choose File] No file chosen   (≡ example)

Label to group by: ⑦   [N/A ▾]

Visualization   Manual>Visualization

Color scheme: ⑦   [Hierarchical color scale ▾]

Figure 3

Figure 4

A

B

Figure 5