FoodMicrobionet 2: Structure of edges and nodes tables.

Eugenio Parente. Dipartimento di Scienze, Università degli studi della Basilicata.

This document describes the structure of the edge and node tables for FoodMicrobionet. Several new fields have been introduced (and a few have been removed) compared to version 1 (Parente et al., 2016). Most of the new fields are somewhat redundant for use in network visualisation software (Gephi, Cytoscape) but they do make the selection of specific subsets of sample or environmental nodes easier. In addition, most of the new features were designed to facilitate processing of the tables with other software packages (Microsoft Excel, the CoNet App of Cytoscape, the Bipartite package of R). To learn more on FoodMicrobionet visit our web site at: http://www.foodmicrobionet.org

The edges table includes a list of all relationships between Operational Taxonomic Units and samples (food or food environment). The fields included in the current version of FoodMicrobionet are:

1) **Source** (Food/food environment sample): the label of the food sample
2) **Target** (OTU): the taxonomic assignment including all the lineage in the format Root:k__Bacteria__Firmicutes:c__Bacilli:f__Lactobacillaceae:g__Lactobacillus:s__Lactobacillusdelbrueckii. Current version does not include order as a taxonomic rank
3) **Type**: the type of the relationship; in FoodMicrobionet all edges are set to Undirected
4) **id**: a numerical id for the edge, automatically created importing in Gephi
5) **Gtarget1**: the target used for sequencing, either 16S_DNA or 16S_RNA
6) **Gtarget2**: the variable regions of 16S rRNA gene targeted by PCR amplification
7) **Study**: the number of the study, see FMBN Project Table
8) **OTULabel**: a label for the target node (species if available, otherwise the first valid taxon; Other if unidentified). This field is not needed for Gephi but can facilitate the use of tables with the library bipartite
9) **Weight**: the OTU abundance in the samples (as %)

The nodes table includes metadata for nodes. The current version of FoodMicrobionet includes the following fields:

1) **Id**: a unique value matching either a source or target node.
2) **Label**: this column contains the node display label; same as Target for the OTUs; constructed by combining Source and further information for food/environment samples
3) **Type**: set to "OTU" for nodes corresponding to OTUs, "Sample" for nodes corresponding to food samples, and "Environment" for samples from food environments.
4) **Llabel**: a "long" label. For Sample and Environment nodes incorporates information on FoodID and Food status (see below) in a 8 character format (5 characters for the FoodID

---

1 changes from version 1.0 are in red
2 technically, both these fields could be moved to the node table; this set up is equivalent in Gephi (the same sample can have results on both DNA and RNA extraction) but may be somewhat awkward with other software packages
3 the number of the study is assigned as studies are introduced in FoodMicrobionet. It can be used for rapid selectoin of edges. Although it is redundant in this table (nodes can be selected more easily by using the same field in the node page) it is useful to extract abundance tables from the edge table file.
and 3 for food status, see below for abbreviations). Useful for labelling nodes in some instances and to pass information to files to be used for the bipartite package.

5) **Study**: the number of the study (see above) (for Sample and Environment nodes, set to NA for OTU nodes): this can be used for rapid selection.

6) **Nreads**: number of reads after clean-up, set to NA for OTUs, #N/D if not available

7) **Shannon**: Shannon diversity index, set to NA for OTUs, #N/D if not available

8) **Chao1**: Chao1 diversity index, set to NA for OTUs, #N/D if not available

9) **Nreads2**: a dummy value (2999) to be used for samples (food or environment) if Nreads is not available, otherwise Nreads; this value is used to rebuild OTU tables with absolute frequencies starting from the edge and node tables

10) **Bioproject**: the accession number of the sequences if available, set to NA for OTU and NONE when not available

11) **Study**: see edge table, set to NA for sample nodes

12) **Ontology columns, OTUs**: these fields are set to NA for sample nodes.
   a) **Domain**: only needed in studies in which more than one domain is retrieved, Set to Bacteria
   b) **Phylum**: if available, otherwise set to the value for Domain;
   c) **Class**: if available, otherwise set to "Other";
   d) **Family**: if available, otherwise set to "Other";
   e) **Genus**: if available, otherwise set to "Other";
   f) **Species**: Species, if available, otherwise set to the first valid taxon
   g) **Taxlevel**: a string field representing the identification level achieved (species, genus, family, class, phylum) for filtering purposes

13) **DOI**: the DOI of the article where the study including the data was originally published; set to NONE if the work is unpublished and NA for OTUs

14) **Outlink1**: this field contains a link to external resources. It is currently set to a pointer to the DOI for the paper where the data were originally published (NA for unpublished papers) for sample nodes or a link to NCBI taxonomy for OTU nodes.

15) **Outlink2**: this field contains a link to external resources. It is currently set to a pointer to the study (NA when not available) in SRA for sample/environment nodes or a link to bacteri0.net for OTU nodes.

16) **Ontology columns, samples/environment**: these fields are set to NA for OTU nodes; the hierarchical terms and IDs are derived from the food classification and description system FoodEx 2 (draft-revision 1) adopted by the European Food Safety Authority (EFSA, 2011). For food environments these are set to the same values of the food processed in that environment.4
   a) **FoodCat**: the food macro category
   b) **FoodGroup**: the food group
   c) **Foodsubgroup1**: the food subgroup in FoodEx 2 if available, otherwise Foodsubgroup1 is used
   d) **Foodsubgroup2**: the lowest classification level in FoodEx 2 if available, otherwise Foodsubgroup1 is used
   e) **FoodID**: the alphanumerical Food ID code in FoodEx 2

---

4 there is an issue with food samples obtained during processing of, for example, a hard, ripened cheese. A cheese curd is quite obviously a completely different environment compared to a mature hard cheese. This issue is dealt with by the "Nature" field; when the correct classification is not found in FoodEx the closest relative is used
f) **FoodType**: a shorter list of food groups with items Bakery items, Beef, Beverages, Dairy products, Eggs, Fishery products, Multi-ingredient foods, Other, Pork, Poultry, Produce (including fruits)

17) **Food status columns**: these columns describe the status of the food sample (NA for OTUs and environmental samples, shorthand for LLabel 0)

a) **Nature**: Raw (raw material or ingredient), Intermediate, Finished (including the products during storage) (shorthands r, i, f)

b) **Process**: None (raw materials and foods other than those covered by the other options); Mild (foods with no physical lethal treatment; includes foods with added preservatives, mild heat treatments and fermentation); **Type2** (foods submitted at least to a lethal treatment resulting potentially in a 6D decrease in *L. monocytogenes*, but whose treatment is lower than that of type 1 foods; in these foods the heat treatment in between 70°C for 2 min and 90°C for 10 min); **Type1** (foods submitted at least to a lethal treatment resulting potentially in a 6D decrease in *C. botulinum* type E spores, approximately 90°C for 10 min or equivalent)\(^5\). Shorthands for LLabel: n, m, 2, 1

c) **Spoilage**: NA (no information available), Unspoiled (all fresh or processed, non spoiled foods); Fermented (all unspoiled fermented foods), Spoiled (all non-fermented foods which are spoiled or are at the end of shelf-life), Fermented+spoiled (fermented foods which are also spoiled or at the end of shelf-life). Shorthands for LLabel: 0, u, f, s, b.

18) **Selection fields**: these fields are used for partitioning, searching and applying styles to both OTU and sample nodes

a) **Custom1**: set to FoodCat and phylum for sample/environment and OTU nodes, respectively

b) **Custom2**: set to FoodGroup and class for sample/environment and OTU nodes, respectively

c) **Custom3**: set to family for OTU nodes; for sample/environment nodes a description including Food subgroup terms or nature of the environmental samples with further information provided by the authors of the study

**References.**


---

5 this is applicable to ready to eat foods. Both type1 and type2 may include fermented foods which have been submitted to a heat treatment ACMSF, 2007. Report on safe cooking of burgers. Retrieved from http://www.food.gov.uk/multimedia/pdfs/acmsfburgers0807.pdf.

6 includes foods at the end of shelf life, even if they do not show clear spoilage.