Accepted Manuscript

Structure of association networks in food bacterial communities

Eugenio Parente, Teresa Zotta, Karoline Faust, Francesca De Filippis, Danilo Ercolini

PII: S0740-0020(17)30630-5

DOI: [10.1016/j.fm.2017.12.010](https://doi.org/10.1016/j.fm.2017.12.010)

Reference: YFMIC 2922

To appear in: *Food Microbiology*

Received Date: 8 July 2017

Revised Date: 18 December 2017

Accepted Date: 21 December 2017

Please cite this article as: Parente, E., Zotta, T., Faust, K., De Filippis, F., Ercolini, D., Structure of association networks in food bacterial communities, *Food Microbiology* (2018), doi: 10.1016/j.fm.2017.12.010.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1 **Title**

2 **Structure of association networks in food bacterial communities**

3

4 **Authors**

5 **Eugenio Parente^a, Teresa Zotta^b, Karoline Faust^c, Francesca De Filippis^{d,e},**

6 **Danilo Ercolini^{4,5}**

7

8 ^a Dipartimento di Scienze, Università degli Studi della Basilicata, 85100 Potenza,

9 Italy

10 ^b Istituto di Scienze dell'Alimentazione, CNR, 83100 Avellino, Italy

11 ^c Department of Microbiology and Immunology, REGA Institute, KU Leuven,

12 3000, Belgium

13 ^d Department of Agricultural Sciences, Division of Microbiology, University of

14 Naples "Federico II", 80055 Portici, Italy

15 ^e Task Force on Microbiome Studies, University of Naples "Federico II", Naples,

16 Italy

17

18

19 **Corresponding author**

20 Prof. Eugenio Parente

21 Dipartimento di Scienze

22 Università degli Studi della Basilicata

23 Viale dell'Ateneo Lucano 100, 85100 Potenza, Italy

24 eugenio.parente@unibas.it

25

Abstract

The structure of microbial association networks was investigated for seventeen studies on food bacterial communities using the CoNet app. The results were compared with those for host and environmental microbiomes. Microbial association networks of food bacterial communities shared several properties with those of host microbiomes, although they were less complex and lacked a scale-free, small world structure that is characteristic of environmental microbial communities. This may depend on both the initial contamination pattern, whose main source is the raw material microbiome, and on the copiotrophic nature of food environments, with lack of well defined, specific niches. The selective factors which are characteristic of fermentation and spoilage drastically simplified microbial association networks and showed the emergence of negative hubs. Co-presence and mutual exclusion networks had a radically different structure, with high clustering coefficient in the first and high heterogeneity in the latter. Node properties (degree, positive degree, betweenness centrality, abundance) can be combined in plots, which allow a rapid identification of hub species. The combined use of three network inference tools (CoNet, SparCC, and SPIEC-EASI) confirmed that microbial association network detection is method specific, but several coherent copresence or mutual exclusion relationships were detected by at least two different methods.

Key words: 16S DNA profiling; microbial association networks; bacterial communities

1. Introduction

Association networks are increasingly being used for the characterization of the structure of microbial relationships in a variety of ecosystems. These relationships can range from true biological interactions (either positive, such as mutualism or commensalism, or negative, such as competition and amensalism) to simple co-occurrence or co-exclusion patterns due to niche sharing or changes in microbial communities due to successions or competitive exclusion (Faust and Raes, 2012). The inference of microbial association networks (MAN) adds layers of complexity to the more common approaches based on the study of changes of abundance patterns in microbial communities. The theory and terminology used in network analysis of microbial communities has been reviewed by several authors (Deng et al., 2012; Faust and Raes, 2012; Layeghifard et al., 2017). MAN have been found to share structural properties with other complex networks, such as protein and gene networks, but also with social and computer networks. MAN for environmental microbial communities have been found to be scale-free (i.e. their node degree distribution follows a power law, with few highly connected nodes and a large number of nodes with few connections) and to have short average path lengths and high clustering coefficients (i.e. they have a small-world structure, with hub nodes, which connect densely interconnected group of nodes) (Chaffron et al., 2010; Deng et al., 2012; Peura et al., 2015; Steele et al., 2011; Zhou et al., 2011). A highly modular structure with scale-free, small world topology has also been identified in host microbiomes-associated networks (Duran-Pinedo et al., 2011; Faust et al., 2012).

The analysis of MAN may help in uncovering previously unexplored interactions between microbial taxa (Faust et al., 2012; 2015), detecting species which act as

76 "hubs" (which have a large number of interactions with other species) or
77 "bottlenecks" (which have a high betweenness centrality, a measure of the
78 number of shortest paths which pass through a node in a network), and whose
79 removal may cause significant disturbance in the ecosystem (Berry and Widder,
80 2014; Layeghifard et al., 2017; Peura et al., 2015). Current approaches and best
81 practices for inference of MAN and their limitations have been reviewed (Berry
82 and Widder, 2014; Faust and Raes, 2012; Layeghifard et al., 2017). The inferred
83 microbial association network is method dependent, as demonstrated by
84 analysis on both real (Weiss et al., 2016) and simulated (Wang et al., 2017; Weiss
85 et al., 2016) microbial communities, and only some interactions may be
86 detectable by all algorithms (Weiss et al., 2016). In addition, only in a few cases
87 the interactions inferred with network analysis have been confirmed with
88 independent methods (Biswas et al., 2016; Cardinale et al., 2015; Kastman et al.,
89 2016; Lima-Mendez et al., 2015).

90 Application of microbial association analysis to the study of food microbial
91 communities is poorly exploited. Food microbial communities are less complex
92 than environmental and host microbial communities (Cocolin and Ercolini, 2015;
93 Ercolini, 2013; Parente et al., 2016a). Food ecosystems are characterized by a
94 biotic matrix where nutrients are usually abundant. The primary microbial
95 community originates from raw materials of animal or plant origins, and from
96 contamination arising from the food processing environment. Moreover,
97 microorganisms may be added as starter cultures, and microbial communities
98 often evolve rapidly in food in response to intrinsic (pH, a_w , etc.) or extrinsic
99 factors (temperature, atmosphere, etc.) with 100-10,000 fold changes in
100 numbers in the range of a few hours or days (ICMSF, 1980). Interactions among

microorganisms are important in both food spoilage (Gram et al., 2002b) and fermentations (Smid and Lacroix, 2013). However, positive or negative interactions may be significant only when a certain density threshold is exceeded or when colonies or micro-colonies of interacting microbes are very close (Malakar et al., 2003).

We recently inferred MAN for food bacterial communities using data from a small number of studies for dairy and meat products (Parente et al., 2016a) and found that the networks were less complex compared to those for environmental and host microbiomes and did not show the characteristics of scale-free, small world networks. The use of association networks in food microbiology studies may help to shed light on biologically meaningful relationships, highlight possible targets for improvement of food quality and safety, and provide new approaches for identifying hub species in food microbial communities. Here, using a significantly larger dataset, we inferred microbial association network using the CoNet app (Faust and Raes, 2016), in order to investigate the structure and properties of a variety of bacterial association networks in foods and to compare such features with those of microbial communities in host or environmental biomes (Faust et al., 2015).

2. Materials and methods

2.1 Datasets

The list of datasets used in this work is shown in **Table 1**. The data were obtained from FoodMicrobionet v 1.1 (Parente et al., 2016b; Studies 1-22) or from sequence data downloaded from public databases (Studies 29-31). Most studies were obtained from data generated at the same sequencing center, using

the same platform, target gene, and bioinformatic pipeline. The number of sequences per sample was in a relatively narrow range: median 3855, with 90% of the samples between 2923 and 8851 sequences. Subsets were extracted from several studies (on the basis of target gene, type of samples, etc.) and for a few studies OTUs were also aggregated at the genus level. A total of 42 OTU abundance tables were used (**Supplementary Table S1**). In addition, data on network structure for host and soil biomes were obtained from supplementary Table 1 of Faust et al. (2015) and used for comparison purposes.

2.2 Association network inference with CoNet

Network inference was carried out with the CoNet app (Faust and Raes, 2016) with Cytoscape 3.4.0, using an approach similar to that used by Faust et al. (2015), with the following modifications: initial filtering was performed by removing OTUs which had a relative abundance of <0.05% and occurred in less than 50% of the samples. Pearson and Spearman correlation, Bray-Curtis and Kullback-Leibler dissimilarity were used to mine interactions and p-value threshold for Benjamini and Hochberg's multiple- testing correction was set to 0.001. The copresence and mutual exclusion networks were analyzed using the Network analyzer tool of Cytoscape and exported for further analysis, together with node and edge tables for the whole association network. Network statistics and fit of the power law to node degree distribution were calculated separately for the whole association network and the copresence and mutual exclusion networks. Modules were detected in the copresence networks using the Cytocluster app of Cytoscape with the OH-PIN algorithm. The Allegro spring-electric layout (<http://apps.cytoscape.org/apps/allegrolayout>) was used to

graphically organize nodes in association networks.

Alpha (Shannon H, Chao1, Pielou's J) and beta diversity analyses (Bray-Curtis and Jaccard) were carried out using the vegan package (Oksanen et al., 2017) of R (R core team, 2016). Network density was calculated as $2E/(N(N-1))$ where E is the number of edges and N the number of nodes. Clustering coefficient was normalized by dividing the clustering coefficient by the density, which approximates the clustering coefficient of a random network of the same size and average degree. Average path length for a random network of the same size and density of a given network was calculated as described by (Fronczak et al., 2004) and used to normalize calculated average path lengths.

Further tests of the node degree distribution on the whole association networks were carried out using the powerLaw (Gillespie, 2015) and the bipartite (Dormann et al., 2008) packages. Principal Component Analysis with varimax rotation was carried out using the functions `faparallel()` and `principal()` of the psych package (Revelle, 2016).

To test the hypothesis that the observed frequency of copresence interactions between nodes belonging to the same family or class was higher than the expected frequency, null distributions were built by random resampling (1,000 iterations) of node pairs and randomly assigning a copresence or mutual exclusion relationship with the constraint that the proportion of positive edges was equal to that of the original network. The same was done to test the hypothesis that the observed frequency of mutual exclusion interactions between nodes belonging to different families or classes was higher than expected. The observed frequency was then compared with the empirical cumulative distribution.

176

177 *2.3 Comparison of network inference tools*

178 Interaction networks inferred using CoNet were compared with networks
179 inferred using SparCC (Sparse Correlations for Compositional data; Friedman
180 and Alm, 2012) or SPIEC-EASI (SParse Inverse Covariance Estimation for
181 Ecological Association Inference; Kurtz et al., 2015).

182 As a first step, phyloseq class objects were created using the OTU tables and
183 taxonomic data for the studies listed in Table 1 with package phyloseq
184 (McMurdie and Holmes, 2013), using the same filters for OTU abundance and
185 occurrence which had been used for CoNet network inference. Then networks
186 were inferred using the functions spiec.easi() and sparcc() of the SpiecEasi
187 package (<https://github.com/zdk123/SpiecEasi>). For the first function, the
188 neighborhood selection (MB) method was used with lambda minimum ration
189 1×10^{-2} and 50 replicates; nlambd was adjusted by a trial and error approach by
190 testing values between 20 and 50. For sparcc estimation the sparccboot()
191 function with 100 replicates was used to generate pseudo-p values and only
192 edges whose p value was ≤ 0.01 were retained.

193 The inferred networks were transformed in igraph objects (Csardi and Nepusz,
194 2006) and edge tables were exported. Venn diagrams of common edges between
195 networks inferred using Conet, SPIEC-EASI and SparCC were generated using
196 functions from package gplots (Warnes et al., 2016).

197

198 **3. Results**

199

200 *3.1 A comparison of microbial association networks from food, host and soil biomes*

We explored the properties of microbial association networks (MAN) for food bacterial communities (**Table 1**) and compared them with those of host and soil biomes described by Faust et al. (2015) using the CoNet app.

With a few exceptions, we found that node degree distribution of food MAN did not fit the power law (**Supplementary Fig. S1**) nor the Poisson distribution. A significantly better fit was obtained by using a truncated power law distribution with adjusted R^2 values between 0.97-0.99.

Density decreased non linearly with the number of nodes and was highest for food biomes (**Supplementary Fig. S2**). The median density of food and host biomes were 0.26 and 0.12, respectively, and the former was significantly higher (Wilcoxon Rank Sum test, $W = 168$, $p\text{-value} = 2.05 \times 10^{-5}$).

The relationship between the normalized clustering coefficient and the number of nodes in the networks is shown in **Fig. 1**, while the relationship of average clustering coefficients with networks size and the distribution of average clustering coefficients are shown in **Supplementary Fig. S3**. The normalized clustering coefficient for food biomes ranged from 0.65 to 2.81, with a median value of 1.75. The type of biome had a significant effect on the location of the distribution of the normalized clustering coefficient (Kruskal Wallis test, $\chi^2 = 23.624$, $df = 4$, $p\text{-value} = 9.5 \times 10^{-5}$) and the median value for food biomes was significantly lower than that of host biomes retrieved from the Human Microbiome Project (Wilcoxon Rank Sum test, $W = 13$, $p\text{-value} = 2.79 \times 10^{-5}$).

However, when copresence networks alone were analyzed, the normalized clustering coefficient (median 3.75, median absolute deviation 0.75) was significantly higher than 1 ($p < 0.001$).

The relationship between average path length and number of nodes in the

network is shown in **Fig. 2**, while the distribution of average path length (APL) values and normalized average path length values are shown in **Supplementary Fig. S4**.

Average path length was significantly correlated with the logarithm of node count (Pearson's $r = 0.81$, $p = 1.09 \times 10^{-9}$) when all biomes were considered. However, this was due only to host and soil biomes, while the correlation for food biomes was low and was not significantly different from 0 ($r=0.30$, $p=0.23$). The median value of normalized average path length for Food biomes was 1.01 (range 0.95 - 1.10) and was significantly higher ($p<0.01$) than that of soil biomes but not than that of host biomes. The values of average path length for soil biomes from the EMP were in the range of those reported by (Deng et al., 2012) for association networks calculated using the MENA pipeline, while those of host and food biomes were lower.

Faust et al. (2015) found that the proportion of positive edges (PEP) was significantly different in different biomes and that PEP was significantly correlated with average clustering coefficient and density. When food biomes were included, only the correlation between clustering coefficient and PEP was significantly different from 0 (Spearman's $\rho = 0.57$, $p = 0.0002$), which is not surprising, given the high value of clustering coefficients in copresence networks (see below). The distribution of PEP in different biomes is shown in **Fig. 3**.

The shape and location of distribution of PEP for different biomes was significantly different (Kruskal Wallis rank sum test, $\chi^2 = 15.33$, $df = 4$, $p\text{-value} = 0.0041$); however, food biomes were not significantly different from host biomes. Differences with soil biomes were only due to the QIIME data set.

To get a better insight in the structure of MAN in foods we analyzed separately

251 copresence and mutual exclusion networks. In addition, for several studies it was
252 possible to extract OTU tables including more homogeneous groups of samples
253 (**Supplementary Table S1**).

254 Some topological indices for the copresence (node count) and mutual exclusion
255 networks (node count, density, average path length) were significantly
256 correlated with the number of samples in the study (data not shown). However,
257 this is likely to be due to biases in the dataset. In fact, studies with large number
258 of samples (>75) were also those with the largest beta-diversity (as measured by
259 the median value of the Bray-Curtis dissimilarity matrix) and the highest values
260 of evenness (as measured by the median value of Pielou's evenness index).

261 Density, normalized cluster coefficient and average path length were compared
262 for whole association networks, copresence and mutual exclusion networks. The
263 results are shown graphically in **Supplementary Fig. S5, S6 and S7**.

264 A Wilcoxon signed rank test was used to compare the properties of copresence
265 and mutual exclusion networks. Density was significantly ($p = 0.0003$) higher in
266 copresence networks, and was negatively correlated with the logarithm of the
267 size for copresence (Spearman's $\rho = -0.59$, $p < 0.01$) and mutual exclusion ($\rho = -$
268 0.81 , $p < 0.01$) networks. The normalized clustering coefficient was significantly
269 higher in copresence networks ($p = 1.9 \times 10^{-6}$) and was significantly correlated
270 with the logarithm of the size ($\rho = 0.70$, $p < 0.001$) for mutual exclusion but not
271 for copresence networks. Clustering coefficients in mutual exclusion networks
272 were low (0-0.08) and lower than expected for random networks (normalized
273 clustering coefficient < 1).

274 Average path length was correlated with network size for both copresence ($\rho =$
275 0.561 , $p < 0.001$) and mutual exclusion ($\rho = 0.564$, $p < 0.0095$). Average path

length was significantly higher for mutual exclusion networks, while normalized average path length was lower (Wilcoxon signed rank test, $p < 0.001$).

The distribution of heterogeneity (a measure of the tendency of the network for having hub nodes; Dong and Horvath, 2007) by interaction type is shown in **Supplementary Fig. S8**. Heterogeneity was not correlated with network size for any of the interaction types and was significantly higher in mutual exclusion networks (Wilcoxon signed rank test, $p = 0.0014$).

Finally, we carried out module detection in copresence networks. The number of modules varied from 1 to 5 and was significantly correlated with average path length and size of the network (logarithm of the number of nodes).

3.2 Exploring copresence networks

This study includes microbial association networks from a wide variety of food environments (**Table 1**) and is, as a consequence, exploratory in nature. In addition, network parameters are known to be related in a complex way to size, alpha- and beta-diversity of the OTU count matrices (Berry and Widder, 2014; Faust et al., 2015; see also section 3.1). Therefore, a direct comparison of networks from different environments is difficult. However, a few studies were sufficiently large and diverse to allow the extraction of OTU tables with more than 24 samples having more homogeneous properties (**Supplementary Table S1**), whereas some studies were relatively homogeneous and included only fermented foods, food environments or raw food samples. A Principal Component Analysis with varimax rotation was carried out using both network topological parameters (logarithm of the node count, density, normalized clustering coefficient, heterogeneity) and information on the alpha- and beta-

diversity of the OTU count matrix (median values for Chao1, PielouJ and Bray Curtis dissimilarity). The first two rotated components explained 46 and 33% of the total variance and a score and loading plot is shown in **Fig. 4**. The properties of the OTU count matrix were associated only with the first component, while, with the exception of node counts (whose loading correlated with measures related to the OTU count matrix) all the other network topological indexes loaded on the second component. The following features can be highlighted:

- a. networks for fermented foods or spoiled samples were only found in the left side of the graph; they were characterized by low average values of alpha-diversity and beta-diversity, by a lower evenness and by smaller size (number of nodes);
- b. for study 10, data obtained on the same samples using both 16S DNA and cDNA as target were available; the results were highly divergent, with the cDNA network showing higher normalized clustering coefficient and heterogeneity and lower density;
- c. food environment and food samples can be compared for studies 19 and 21. While for the latter, which include both meat and environmental samples from butcheries (Stellato et al., 2016) the results were quite similar, food environments for RTE foods (Pothakos et al., 2015) had copresence networks with a comparatively larger number of nodes, lower density and higher heterogeneity, as well as a slightly higher normalized clustering coefficient.

Finally, for selected studies, networks were inferred for OTU tables where the identification had been carried out at the species level or above or at the genus level or above. We used non-parametric tests for paired data to compare the

properties of copresence networks inferred using information at the species level or above or at the genus level or above. Besides the obvious differences in network size, no significant difference was found for any of the other network topological parameters for any type of interaction.

3.3 *The structure of selected association networks*

A detailed analysis on two selected studies, one for fermented foods with low diversity and network size (Study 10; De Filippis et al., 2016) and one with higher values of alpha and beta diversity and of evenness (Study 11; De Filippis et al., 2013), is shown as an example.

Study 10 is a longitudinal study including 42 samples obtained during the manufacture of Caciocavallo Silano, a hard pasta filata cheese (De Filippis et al., 2016); community profiling data obtained using cDNA as a target were also available for 25 samples. The study is characterized by a low alpha (median Chao1 40 and 35 for DNA and cDNA, respectively) and beta (Bray-Curtis dissimilarity matrix median±median absolute deviation 0.23 ± 0.14 and 0.18 ± 0.12 for DNA and cDNA, respectively) diversity and by intermediate evenness (median value for Pielou's evenness 0.36 and 0.29 for DNA and cDNA, respectively). The complete association networks based on DNA and cDNA profiling are shown in **Supplementary Fig. S9**, while a degree plot is shown in **Fig. 5** for the network obtained from DNA profiling.

The two networks shared approximately two thirds of the nodes while twelve and nine nodes were specific of the cDNA and DNA network respectively. The cDNA network had an unusually low value for PEP. Both the copresence network for DNA and cDNA had two modules (not shown), one essentially including

species associated with the raw milk environment, and another, smaller, which included lactobacilli associated with cheese manufacture and ripening. In addition, the two networks had a radically different structure. In the network based on DNA as a target, a group of mesophilic and psychrotrophic species whose origin is from raw milk, and which formed the main module in the copresence network, behaved as positive hubs in the network and were strongly connected, with relatively high values of betweenness centrality (i.e. they are also bottlenecks).

Fig. 5 is designed to highlight keystone taxa. The group of nodes in the upper right of the graph can be considered as hub species in the network (because of high positive and total degree; Peura et al., 2015). Although these are relatively less abundant, they are present in most or all samples and their source is likely to be from raw milk. They also share relatively high values of betweenness centrality (i.e. beside being highly connected, their removal would remove a high proportion of shortest paths in the networks, disrupting it). Nodes with the largest distance from the diagonal dotted line showing the average ratio of positive interactions have a relatively larger number of positive (below the line) or negative (above the line) interactions. The group of species on the left of the vertical dotted line (which marks the average value for positive degree) are negative hubs and include species of the genus *Lactobacillus* which dominate the cheese during ripening. The species in the lower left quadrant have a lower than average degree, and are less important in the association network. *Streptococcus thermophilus*, one of the most abundant species (it was part of the natural starter culture for the cheese), had a relatively low degree and was connected to some psychrotrophic species by negative edges; its only positive edge was with *L.*

helveticus, the other member of the natural starter culture (De Filippis et al., 2016).

In the network obtained from the cDNA, the group of negative hubs associated with the mature cheese had negative edges with primary starters and species associated with raw milk. The difference between the two networks is likely to be due to the permanence of extracellular DNA or DNA from inactive cells in the samples, which overemphasizes the role of species which mark only a short phase of the succession of this ecosystem.

Study 11 is a large study on contamination of beef carcasses (De Filippis et al., 2013), including both unspoiled and spoiled samples. It also included a limited number of samples from food environments that were excluded from the association network analysis. A large number of OTUs was observed and the study was characterized by intermediate values for median alpha diversity (median value for Chao1 28; the distribution was bimodal and the median value was affected by the low Chao1 values for spoiled samples), high values for beta diversity (median value of Bray Curtis dissimilarity matrix 0.686) and intermediate values of evenness (median value for Pielou J 0.464). The association network is shown in **Supplementary Fig. S10**, while a degree plot is shown in **Fig. 6**.

The high diversity resulted in a complex association network. Two modules were detected in the copresence network, one including the species constantly present in the unspoiled samples and one, smaller, including the species dominating the spoiled samples (data not shown). The most abundant species did not have large values for total degree and betweenness centrality, which are more frequent among the members of the shared core microbiota. On the other hand, species

which were more frequently associated with the spoiled samples had disproportionately high values of negative interactions, because they tended to exclude the core microbiota. *Acinetobacter guilloiae* showed a disproportionate ratio of degree to positive degree: this species, which appeared at low abundance, was systematically present in spoiled samples and was not detected in fresh carcasses, thus acting as a marker for spoilage. Another striking feature of the graph compared to that of Fig. 5 is that a large number of nodes tend to have a PEP ratio close to average (they are distributed along the diagonal line). Two features emerged from the analysis of these and other networks (data not shown): abundant species tended to have a high ratio of negative interactions to total interactions, and positive interactions among members of the same family or class were often more frequent than expected from random sampling (**Fig. 7**). The probability of copresence relationships within members of the same class was significantly ($p < 0.05$) higher than random expectations only for 10 networks out of 17. As to mutual exclusion relationships, the probability of their occurrence between members of different families was also significantly higher than that expected from random sampling. In the same networks the probability of mutual exclusion relationships among members of different classes was also significantly higher than expected.

3.4 A comparison of network inference tools

We compared networks inferred by three tools (CoNet, which uses an ensemble of approaches and provides a convenient interface with the Cytoscape app and has a low false positive rate; SparCC, which is particularly useful with sparse, compositionally biased data; and SPIEC-EASI, which has recently been shown to

perform well on both synthetic and real data) on the seventeen studies of **Table 1**. Venn diagrams for co-presence and mutual exclusion relationships are shown in **Supplementary Fig. S11** and **S12**. Both SparCC and SPIEC-EASI failed to detect an association network for Study 3 (which also resulted in the smallest association network, 8 nodes and 11 edges with CoNet). No association network was detected by SPIEC-EASI for study 10. In general, CoNet detected the largest number of interactions, followed by SparCC, while SPIEC-EASI detected only a few significant interactions for each study, and these were often parent-child relationships (i.e. relationships between members of a lower taxon with those of the corresponding higher taxon; for example *Acinetobacter johnsonii*--*Acinetobacter*). However, all relationships detected by SPIEC-EASI were also detected by SparCC and, to a lesser extent, by CoNet. A large number of interactions was specific to either CoNet or SparCC. The 40 most frequent copresence and mutual exclusion interactions detected by one or more tools across all datasets are shown in **Table 2** and **3** respectively.

Copresence relationships were frequent among bacteria used as starters in the dairy products used for the studies listed in **Table 1**, and among psychrotrophic spoilage organisms which are frequent in raw meat, fish, dairy or vegetable products. Conversely, mutual exclusion relationships were frequent between beneficial starter organisms and spoilage organisms (and, in general, between *Bacilli* and *Proteobacteria*).

4. Discussion

Several important network properties are usually compared to evaluate the general structure and complexity of association networks: node degree

distribution, network density, average clustering coefficient and average path length. A fit of the power law to the degree distribution indicates a scale-free nature of the network, with a low number of nodes with a high degree and a high number of nodes with low degree, which is a characteristic of many complex networks (Newman, 2010), including complex food webs and microbial association networks (Deng et al., 2012; Dunne et al., 2002; Faust et al., 2012, 2015; Wang et al., 2017). The node degree distribution of microbial association networks (MAN) for food bacterial communities did not fit the power law nor the Poisson distribution, which is typical of random networks (Newman, 2010). This is in agreement with what has been found for other microbial ecological networks (Steele et al., 2011; Zhou et al., 2011) and reflects both the occurrence of highly connected "hub" species and a constraint to the maximum number of edges, whose nature may depend on analytical limitations (i.e. sequencing depth).

MAN for food bacterial communities had, on average, a higher density, a lower average path length and a lower average clustering coefficients compared to those of host and, especially, soil biomes, but the proportion of positive interactions was similar to that of host biomes. This is in part related to network size, which was systematically lower for food networks.

We conclude that MAN for food are less complex than soil biomes and do not show the properties of scale-free, small-world networks. Their properties are close to those of host biomes of the HMP dataset (Faust et al., 2015). In fact, one of the main sources of bacteria in foods, especially in terms of diversity, is the microbiota of raw material of animal or plant origin, although other factors (contamination from food processing environments, addition of starter cultures)

may contribute. The lack of a scale-free, small-world structure is more difficult to explain and may be related to a number of factors. Because of abundance of nutrients, especially in raw foods of both animal and plant origin, modules of nodes associated to specific trophic interactions or niches, which are found in soil or host biomes (Faust et al., 2015) may be obscured by copresence patterns related to a common contamination pattern. In addition, large time series studies are rare for food bacterial communities, and the nature of data obtained from 16S profiling studies for foods (which are invariably relative abundance rather than absolute count data) is not completely adequate for the evaluation of interactions in microbial communities (Berry and Widder, 2014; Layeghifard et al., 2017). This may be further aggravated by the fact that DNA from inactive cells may persist for a relatively long time and mask the relationships between active fractions of the microbiota. In our data set, studies based on cDNA did have normalized clustering coefficients for the copresence network that were in the upper quartile of the distribution, but APL was only rarely shorter than that expected from random networks. Differential detection of active and inactive cells using molecular methods has been shown to provide completely different pictures of the evolution of microbial populations and communities in food (Erkus et al., 2016). Finally, confounding factors, related to the heterogeneity of the studies included in this work may also contribute to obscure the nature of copresence and mutual exclusion relationships and *ad hoc* planned experiments may be helpful to prove that the structure of MAN in foods is systematically different from that of environmental MAN.

The presence of both positive (copresence) and negative interactions in microbial association networks may complicate the interpretation of their

structure. Therefore, copresence and mutual exclusion networks for food microbial communities were compared. Copresence and mutual exclusion networks for food bacterial communities had a radically different structure, with high density and normalized clustering coefficient and lower average path length and heterogeneity in the former and the reverse in the latter. Copresence networks, while lacking the properties of scale-free networks which were detected in association networks for complex environments (Deng et al., 2012; Faust et al., 2012, 2015; Wang et al., 2017), have high clustering coefficients compared to random networks and short average path lengths, showing some of the properties of small-world networks, although their heterogeneity was low. Low clustering coefficients in mutual exclusion networks have been attributed to the high proportion of positive interactions in the neighbors of nodes (Faust et al., 2015) but this might also be related to the nature of the clustering coefficient: if two OTU nodes tend to be excluded by the same node, they tend to be positively, rather than negatively associated. The high heterogeneity of mutual exclusion networks may be due to the emergence of blooming taxa in fermented and spoiled foods, which disproportionately dominate the microbial community and appear to exclude most of the other taxa.

Given the heterogeneity of the data used in our study and the complex relationship between network properties and network size (Berry and Widder, 2014; Faust et al., 2015; see also section 3.1), we compared homogeneous and representative subsets of data from selected studies. We confirmed that fermentation and spoilage significantly decreased (Parente et al., 2016a) the size of microbial association networks and found that the structure of networks inferred at two different taxonomic resolution levels were similar. It has been

suggested (Berry and Widder, 2014) that network inference should always be carried out at the highest resolution possible and, indeed, species belonging to the same genus may play very different roles in food manufacturing or spoilage processes (Ercolini, 2013). However, this may be due to the fact that, with few exceptions, only one species was important (in terms of either abundance or of node properties) in each genus.

We confirmed that copresence relationships are relatively more frequent among members of the same family, while the reverse was true for mutual exclusion relationships. This may indicate a lineage-dependent ecological coherence (Koeppel and Wu, 2012) with members of the same taxa sharing similar niches and therefore co-occurring at a significantly higher rate.

A variety of methods have been proposed and used for the inference of MAN (Layeghifard et al., 2017). Recently, Weiss et al. (2016) compared eight correlation detection strategies on synthetic datasets and found that the ability of different methods in detecting positive and negative interactions varies significantly in sensitivity, specificity and accuracy, and proposed that ensemble approaches using a combination of methods may be more effective in detecting associations. In fact, a combined approach has been found to be successful in detecting meaningfully significant associations in paddy soils (Wang et al., 2017).

Therefore, we compared three network inference tools, two of which (CoNet, Faust et al. 2016; and SparCC, Sparse Correlations for Compositional data; Friedman and Alm, 2012) have frequently been used for the reconstruction of microbial association networks, while the other (SPIEC-EASI) has been only proposed recently and has been shown to perform well on data from the American Gut Project (Kurtz et al., 2015). We confirmed that the association

networks detected were method specific, as shown for synthetic datasets by Weiss et al. (2016) and on environmental datasets by Wang et al. (2017). Although CoNet systematically predicted the highest number of interactions and SPIEC-EASI the lowest, several interactions were detected by at least two different methods. The most frequent interactions were commonly copresence relationships between starter and non-starter beneficial microorganisms or between psychrotrophic spoilage organisms, or mutual exclusion relationships between beneficial and spoilage microorganisms. The negative relationships between starter and spoilage organisms, and the positive relationships between starter species are likely to reflect the changes (lowering of pH and redox potential, production of organic acids) caused by fermentation (Caplice and Fitzgerald, 1999), with known competition and amensalism relationships between starter species and Gram negative spoilage organisms, and both niche sharing and true proto-cooperation relationships for starter bacteria (Ivey et al., 2013; Smid and Lacroix, 2013). Copresence relationships among spoilage organisms reflect the common composition of specific spoilage associations for raw foods (Gram et al., 2002a; Remenant et al., 2015). They may also imply metabiosis and commensalistic relationships, but direct proof for this must be obtained independently.

5. Conclusions

Inference of microbial association networks in foods adds a layer of complexity to the common approach based on alpha- and beta-diversity analysis and may assist in the identification of hub species in food microbial ecosystems. However, this approach is still underexploited in food microbiology and several issues

remain to be solved. Network inference tools applied to studies which have not been specifically designed to detect microbial interactions are useful as mining tools, and especially when an ensemble of methods is used, can help in the identification of consistent patterns of ecologically relevant association, that are however to be confirmed with independent experiments in time series studies. The CoNet app for Cytoscape offers a convenient interface which greatly simplifies the tasks of network inference and representation. However, in our hands it consistently detected more association relationships compared to other tools (such as SparCC and SPIEC-EASI). This may be due, in part, to the use of an ensemble of methods (in our case Pearson and Spearman correlation, Bray-Curtis and Kullback-Leibler dissimilarity), but also to the detection of indirect associations in copresence networks, which are claimed not to be detected by other tools (SPIEC-EASI, Kurtz et al., 2015). In addition, we found that the structure of association networks inferred from 16S DNA or cDNA profiling are completely different. The use of DNA as a target is very common and gives a useful picture of the structure of the microbial diversity. However, from the MAN perspective, DNA analysis may detect the “ghosts” of the associations between members of the community contaminating the raw food, which may include species that will give little contribution to the spoilage or fermentation process. Several MAN inference methods are generally considered robust toward the compositional nature of OTU abundance tables (Kurtz et al., 2015; Layeghifard et al., 2017; Weiss et al., 2016) but independent confirmation of the inferred relationships using direct or indirect approaches in simulated co-culture experiment is rare, especially for food microbial communities (see Katsman et al., 2016 for a recent example). Quantitative methods based on qPCR or on flow

cytometry each have their limitations (Elizaquível et al., 2014; Sohier et al., 2014) and are definitely low throughput compared to 16S profiling. However, combination of 16S or metagenomic profiling and quantitative methods have been used satisfactorily to monitor the dynamics of active and inactive fractions of relatively simple microbial communities in starter cultures and cheese (Erkus et al., 2016) or the dynamics of simple simulated spoilage communities (Cauchie et al., 2017). A more systematic approach, integrating different analytical approaches with modeling and inference tools would most definitely benefit our knowledge of microbial interactions in foods.

Acknowledgements

This work was partially funded by project 'Smart Basilicata' which was approved by the Italian Ministry of Education, University and Research (MIUR n.84/Ric 2012, PON 2007-2013 of 2 March 2012) and was funded with the Cohesion Fund 2007–2013 of the Basilicata Regional authority. The authors declare they have no conflict of interest.

References

- Alessandria, V., Ferrocino, I., De Filippis, F., Fontana, M., Rantsiou, K., Ercolini, D., Cocolin, L., 2016. Microbiota of an Italian Grana like cheese during manufacture and ripening unraveled by 16S rRNA-based approaches. *Appl. Environ. Microbiol.* 82, 3988–3995.
- Berry, D., Widder, S., 2014. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Front. Microbiol.* 5, 219.
- Biswas, S., McDonald, M., Lundberg, D.S., Dangl, J.L., Jojic, V., 2016. Learning

- 626 Microbial Interaction Networks from Metagenomic Count Data. *J. Comput.*
627 *Biol.* 23, 526–535.
- 628 Calasso, M., Ercolini, D., Mancini, L., Stellato, G., 2016. Relationships among house,
629 rind and core microbiotas during manufacture of traditional Italian cheeses
630 at the same dairy plant. *Food Microbiol.* 54, 115–126.
- 631 Caplice, E., Fitzgerald, G.F., 1999. Food fermentations: role of microorganisms in
632 food production and preservation. *Int. J. Food Microbiol.* 50, 131–149.
- 633 Cardinale, M., Grube, M., Erlacher, A., Quehenberger, J., Berg, G., 2015. Bacterial
634 networks and co-occurrence relationships in the lettuce root microbiota.
635 *Environ. Microbiol.* 17, 239–252.
- 636 Cauchie, E., Gand, M., Kergourlay, G., Taminiau, B., Delhalle, L., Korsak, N., Daube,
637 G., 2017. The use of 16S rRNA gene metagenetic monitoring of refrigerated
638 food products for understanding the kinetics of microbial subpopulations at
639 different storage temperatures: the example of white pudding. *Int. J. Food*
640 *Microbiol.* 247, 70–78.
- 641 Chaillou, S., Chaulot-Talmon, A., Caekebeke, H., Cardinal, M., Christieans, S., Denis,
642 C., Desmonts, M.-H., et al. 2015. Origin and ecological selection of core and
643 food-specific bacterial communities associated with meat and seafood
644 spoilage. *ISME J.* 9, 1105–1118.
- 645 Chaffron, S., Rehrauer, H., Pernthaler, J., von Mering, C., 2010. A global network of
646 coexisting microbes from environmental and whole-genome sequence data.
647 *Genome Res.* 20, 947–959.
- 648 Cocolin, L., Ercolini, D., 2015. Zooming into food-associated microbial consortia: a
649 “cultural” evolution. *Curr Opin Food Sci.* 2, 43–50.

- 650 Csardi, G., Nepusz, T., 2006. The igraph software package for complex network
651 research, InterJournal, Complex Systems 1695. <http://igraph.org>.
- 652 De Filippis, F., Genovese, A., Ferranti, P., Gilbert, J.A., Ercolini, D., 2016.
653 Metatranscriptomics reveals temperature-driven functional changes in
654 microbiome impacting cheese maturation rate. *Sci. Rep.* 6, 21871.
- 655 De Filippis, F., La Stora, A., Stellato, G., Gatti, M., Ercolini, D., 2014. A selected core
656 microbiome drives the early stages of three popular italian cheese
657 manufactures. *PLoS One* 9, e89680.
- 658 De Filippis, F., La Stora, A., Villani, F., Ercolini, D., 2013. Exploring the sources of
659 bacterial spoilers in beefsteaks by culture-independent high-throughput
660 sequencing. *PLoS One* 8, e70222.
- 661 Deng, Y., Jiang, Y.H., Yang, Y., He, Z., Luo, F., Zhou, J., 2012. Molecular ecological
662 network analyses. *BMC Bioinformatics* 13, 30-113.
- 663 Dolci, P., De Filippis, F., La Stora, A., Ercolini, D., Cocolin, L., 2014. rRNA-based
664 monitoring of the microbiota involved in Fontina PDO cheese production in
665 relation to different stages of cow lactation. *Int. J. Food Microbiol.* 185, 127–
666 135.
- 667 Dong, J., Horvath, S., 2007. Understanding network concepts in modules. *BMC*
668 *Syst. Biol.* 4, 1-24.
- 669 Dormann, C.F., Gruber, B., Fruend, J., 2008. Introducing the bipartite Package:
670 Analysing Ecological Networks. *R news* Vol 8/2, 8–11.
- 671 Dunne, J.A., Williams, R.J., Martinez, N.D., 2002. Food-web structure and network
672 theory: The role of connectance and size. *Proc. Natl. Acad. Sci. USA* 99,
673 12917–12922.
- 674 Duran-Pinedo, A.E., Paster, B., Teles, R., Frias-Lopez, J., 2011. Correlation network

- 675 analysis applied to complex biofilm communities. PLoS One 6, e28438.
- 676 Elizaquível, P., Aznar, R., Sanchez, G., 2014. Recent developments in the use of
677 viability dyes and quantitative PCR in the food microbiology field. J. Appl.
678 Microbiol. 116, 1–13.
- 679 Ercolini, D., 2013. High-throughput sequencing and metagenomics: moving
680 forward in the culture-independent analysis of food microbial ecology. Appl.
681 Environ. Microbiol. 79, 3148–3155.
- 682 Erkus, O., de Jager, V.C.L., Geene, R.T.C.M., van Alen-Boerrigter, I., Hazelwood, L.,
683 van Hijum, S.A.F.T., Kleerebezem, M., Smid, E.J., 2016. Use of propidium
684 monoazide for selective profiling of viable microbial cells during Gouda
685 cheese ripening. Int. J. Food Microbiol. 228, 1–9.
- 686 Faust, K., Raes, J., 2016. CoNet app: inference of biological association networks
687 using Cytoscape. F1000Res 5, 1519.
- 688 Faust, K., Lima-Mendez, G., Lima-Mendez, G., Lerat, J.S., Lerat, J.S.,
689 Sathirapongsasuti, J.F., Sathirapongsasuti, J.F., Knight, R., Huttenhower, C.,
690 Lenaerts, T., Lenaerts, T., Raes, J., 2015. Cross-biome comparison of microbial
691 association networks. Front. Microbiol. 6, 1200.
- 692 Faust, K., Raes, J., 2012. Microbial interactions: from networks to models. Nat.
693 Rev. Microbiol. 10, 538–550.
- 694 Faust, K., Sathirapongsasuti, J.F., Izard, J., Segata, N., Gevers, D., Raes, J.,
695 Huttenhower, C., 2012. Microbial co-occurrence relationships in the human
696 microbiome. PLoS Comput. Biol. 8, e1002606.
- 697 Ferrocino, I., Greppi, A., La Stora, A., Rantsiou, K., Ercolini, D., Cocolin, L., 2015.
698 Impact of nisin-activated packaging on microbiota of beef burgers during
699 storage. Appl. Environ. Microbiol. 82, 549–559.

- 700 Friedman, J., Alm, E.J., 2012. Inferring correlation networks from genomic survey
701 data. *PLoS Comput. Biol.* 8, e1002687.
- 702 Fronczak, A., Fronczak, P., Hołyst, J.A., 2004. Average path length in random
703 networks. *Phys. Rev. E Stat. Nonlin. Soft. Matter Phys.* 70(5 Pt 2), 056110.
- 704 Gillespie, C.S., 2015. Fitting Heavy Tailed Distributions: The powerLaw Package.
705 *Journal of Statistical Software*, 64(2), 1-16. URL
706 <http://www.jstatsoft.org/v64/i02/>.
- 707 Gram, L., Ravn, L., Rasch, M., Bruhn, J., Christensen, A., Givskov, M., 2002. Food
708 spoilage-interactions between food spoilage bacteria. *Int. J. Food Microbiol.*
709 78, 79–97.
- 710 Greppi, A., Ferrocino, I., La Stora, A., Rantsiou, K., Ercolini, D., Cocolin, L., 2015.
711 Monitoring of the microbiota of fermented sausages by culture independent
712 rRNA-based approaches. *Int. J. Food Microbiol.* 212, 67–75.
- 713 Guidone, A., Zotta, T., Matera, A., Ricciardi, A., De Filippis, F., Ercolini, D., Parente,
714 E., 2016. The microbiota of high-moisture mozzarella cheese produced with
715 different acidification methods. *Int. J. Food Microbiol.* 216, 9–17.
- 716 Hultman, J., Rahkila, R., Ali, J., Rousu, J., Björkroth, K.J., 2015. Meat processing
717 plant microbiome and contamination patterns of cold-tolerant bacteria
718 causing food safety and spoilage risks in the manufacture of vacuum-
719 packaged cooked sausages. *Appl. Environ. Microbiol.* 81, 7088–7097.
- 720 International Commission on Microbiological Specifications for Foods. 1980.
721 *Microorganisms in Foods 3: Microbial Ecology of Foods. Vol. 1: Factors*
722 *affecting life and death of microorganisms.* Academic Press, New York.
- 723 Ivey, M., Massel, M., Phister, T.G., 2013. Microbial interactions in food
724 fermentations. *Annu. Rev. Food Sci. Technol.* 4, 141–162.

- 725 Kastman, E.K., Kamelamela, N., Norville, J.W., Cosetta, C.M., Dutton, R.J., Wolfe,
 726 B.E., 2016. Biotic interactions shape the ecological distributions of
 727 *Staphylococcus* species. *mBio* 7, e01157–16.
- 728 Koeppel, A.F., Wu, M., 2012. Lineage-dependent ecological coherence in bacteria.
 729 *FEMS Microbiol. Ecol.* 81, 574–582.
- 730 Kurtz, Z.D., Müller, C.L., Miraldi, E.R., Littman, D.R., Blaser, M.J., Bonneau, R.A.,
 731 2015. Sparse and compositionally robust inference of microbial ecological
 732 networks. *PLoS Comput. Biol.* 11, e1004226.
- 733 Layeghifard, M., Hwang, D.M., Guttman, D.S., 2017. Disentangling Interactions in
 734 the Microbiome: A Network Perspective. *Trends Microbiol.* 25, 217–228.
- 735 Leff, J.W., Fierer, N., 2013. Bacterial communities associated with the surfaces of
 736 fresh fruits and vegetables. *PLoS One* 8, e59310.
- 737 Lima-Mendez, G., Faust, K., Henry, N., Decelle, J., Colin, S., Carcillo, F., Chaffron, S.,
 738 Ignacio-Espinosa, J.C., Roux, S., Vincent, F., Bittner, L., Darzi, Y., Wang, J., Audic,
 739 S., Berline, L., Bontempi, G., Cabello, A.M., Coppola, L., Cornejo-Castillo, F.M.,
 740 d'Ovidio, F., De Meester, L., Ferrera, I., Garet-Delmas, M.J., Guidi, L., Lara, E.,
 741 Pesant, S., Royo-Llonch, M., Salazar, G., Sánchez, P., Sebastian, M., Souffreau, C.,
 742 Dimier, C., Picheral, M., Searson, S., Kandels-Lewis, S., Tara Oceans
 743 coordinators, Gorsky, G., Not, F., Ogata, H., Speich, S., Stemmann, L.,
 744 Weissenbach, J., Wincker, P., Acinas, S.G., Sunagawa, S., Bork, P., Sullivan, M.B.,
 745 Karsenti, E., Bowler, C., de Vargas, C., Raes, J., 2015. Ocean plankton.
 746 Determinants of community structure in the global plankton interactome.
 747 *Science* 348, 262073.
- 748 Malakar, P., Barker, G., Zwietering, M.H., van't Riet, K., 2003. Relevance of
 749 microbial interactions to predictive microbiology. *Int. J. Food Microbiol.* 84,

- 750 263–272.
- 751 McMurdie, P.J., Holmes, S., 2013. phyloseq: an R package for reproducible
752 interactive analysis and graphics of microbiome census data. PLoS One 8,
753 e61217.
- 754 Newman, M.E.J., 2010. Networks: an introduction. Oxford University Press, New
755 York.
- 756 Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D.,
757 Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H.,
758 Szoecs E., Wagner, H., 2017. vegan: Community Ecology Package. R package
759 version 2.4-2. <https://CRAN.R-project.org/package=vegan>
- 760 Parente, E., Cocolin, L., De Filippis, F., Zotta, T., Ferrocino, I., O'Sullivan, O.,
761 Neviani, E., De Angelis, M., Cotter, P.D., Ercolini, D., 2016a. FoodMicrobionet:
762 A database for the visualisation and exploration of food bacterial
763 communities based on network analysis. Int. J. Food Microbiol. 219, 28–37.
- 764 Parente E., De Filippis F., Zotta T., Cotter P., O'Sullivan O., Puglisi E., Ercolini D.
765 2016b. FoodMicrobionet 1.1.6: a network analysis tool for the exploration
766 of food bacterial communities. Food Micro 2016 Abstract Book, Dublin, IRL,
767 July 19-22, 2016. PO2-FS-080
- 768 Parente, E., Guidone, A., Matera, A., De Filippis, F., Mauriello, G., Ricciardi, A.,
769 2016c. Microbial community dynamics in thermophilic undefined milk
770 starter cultures. Int. J. Food Microbiol. 217, 59–67.
- 771 Peura, S., Bertilsson, S., Jones, R.I., Eiler, A., 2015. Resistant microbial
772 cooccurrence patterns inferred by network topology. Appl. Environ.
773 Microbiol. 81, 2090–2097.
- 774 Pothakos, V., Stellato, G., Ercolini, D., Devlieghere, F., 2015. Processing

- 775 Environment and ingredients are both sources of *Leuconostoc gelidum*, which
776 emerges as a major spoiler in ready-to-eat meals. Appl. Environ. Microbiol.
777 81, 3529–3541.
- 778 R Core Team, 2016. R: A language and environment for statistical computing. R
779 Foundation for Statistical Computing, Vienna, Austria. URL
780 <https://www.R-project.org/>.
- 781 Remenant, B., Jaffrès, E., Dousset, X., Pilet, M.F., Zagorec, M., 2015. Bacterial
782 spoilers of food: Behavior, fitness and functional properties. Food Microbiol.
783 45, 45–53.
- 784 Revelle, W., 2016. psych: Procedures for Personality and Psychological Research,
785 Northwestern University, Evanston, Illinois, USA, [https://CRAN.R-](https://CRAN.R-project.org/package=psych)
786 [project.org/package=psych](https://CRAN.R-project.org/package=psych) Version = 1.6.12.
- 787 Smid, E.J., Lacroix, C., 2013. Microbe-microbe interactions in mixed culture food
788 fermentations. Curr. Opin. Biotechnol. 24, 148–154.
- 789 Sohier, D., Pavan, S., Riou, A., Combrisson, J., Postollec, F., 2014. Evolution of
790 microbiological analytical methods for dairy industry needs. Front. Microbiol.
791 5, 16.
- 792 Steele, J.A., Countway, P.D., Xia, L., Vigil, P.D., Beman, J.M., Kim, D.Y., Chow, C.E.T.,
793 Sachdeva, R., Jones, A.C., Schwalbach, M.S., Rose, J.M., Hewson, I., Patel, A., Sun,
794 F., Caron, D.A., Fuhrman, J.A., 2011. Marine bacterial, archaeal and protistan
795 association networks reveal ecological linkages. ISME J. 5, 1414–1425.
- 796 Stellato, G., La Stora, A., De Filippis, F., Borriello, G., Villani, F., Ercolini, D., 2016.
797 Overlap of spoilage microbiota between meat and meat processing
798 environment in small-scale vs large-scale retail distribution. Appl. Environ.
799 Microbiol. 82, 4045–4054.

- Stellato, G., De Filippis, F., La Stora, A., Ercolini, D., 2015a. Coexistence of lactic acid bacteria and potential spoilage microbiota in a dairy processing environment. *Appl. Environ. Microbiol.* 81, 7893–7904.
- Stellato, G., La Stora, A., Cirillo, T., Ercolini, D., 2015b. Bacterial biogeographical patterns in a cooking center for hospital foodservice. *Int. J. Food Microbiol.* 193, 99–108.
- Wang, H., Wei, Z., Mei, L., Gu, J., Yin, S., Faust, K., Raes, J., Deng, Y., Wang, Y., Shen, Q., Yin, S., 2017. Combined use of network inference tools identifies ecologically meaningful bacterial associations in a paddy soil. *Soil Biol. Biochem.* 105, 227–235.
- Warnes, G.R., Bolker, B., Bonebakker, L., Gentleman, R., Liaw, W.H.A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M., Venables, B., 2016. *gplots: Various R Programming Tools for Plotting Data*. R package version 3.0.1. <https://CRAN.R-project.org/package=gplots>.
- Weiss, S., Van Treuren, W., Lozupone, C., Faust, K., Friedman, J., Deng, Y., Xia, L.C., Xu, Z.Z., Ursell, L., Alm, E.J., Birmingham, A., Cram, J.A., Fuhrman, J.A., Raes, J., Sun, F., Zhou, J., Knight, R., 2016. Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. *ISME J.* 10, 1669–1681.
- Zhou, J., Deng, Y., Luo, F., He, Z., Yang, Y., 2011. Phylogenetic molecular ecological network of soil microbial communities in response to elevated CO₂. *MBio* 2, e00122–11.

Legends to Figures

Fig. 1. Relationship between the normalized clustering coefficient and the network size (as number of nodes) for food microbial association networks and association networks from host (Host_H Human Microbiome Project; Host_Q QIIME) and soil (Soil_E Earth Microbiome Project; Soil_Q QIIME) (Faust et al., 2015). The fit of a power law relationship and the 95% confidence interval is shown.

Fig. 2. Relationship between the average path length and the network size (as number of nodes) for food microbial association networks and association networks from host (Host_H Human Microbiome Project; Host_Q QIIME) and soil (Soil_E Earth Microbiome Project; Soil_Q QIIME) (Faust et al., 2015). Separate regression lines are shown for food, host (Host_H and Host_Q) and soil (Soil_E and Soil_Q) biomes.

Fig. 3. Distribution of positive edge proportion (PEP) for food microbial association networks and association networks from host (Host_H Human Microbiome Project; Host_Q QIIME) and soil (Soil_E Earth Microbiome Project; Soil_Q QIIME) (Faust et al., 2015).

Fig. 4. Score and loading plots for a Principal Component Analysis on copresence network topological properties (l10nc: logarithm of the node count; dens: density; het: heterogeneity; ncc: normalized clustering coefficient) and on some properties of the OTU count matrix used for network inference (medBCm median value for the Bray Curtis dissimilarity matrix; medChao1 median value for the Chao1 index; medPielouJ median value for Pielou's evenness). See Supplementary Table S1 for sample labels.

Fig. 5. Relationship between positive degree and degree for nodes in the microbial association network for Caciocavallo cheese (De Filippis et al., 2016). The data shown are for the network obtained from DNA (16S RNA gene) profiling. Relative abundance of OTUs and betweenness centrality are shown as size and transparency of symbols respectively. Only the labels of the nodes which are in the top 20% range for the properties shown in the graph are shown. The horizontal and vertical dotted lines indicate the average degree and positive degree, respectively, while the diagonal dotted line indicates the average ratio of positive to total degree.

Fig. 6. Relationship between positive degree and degree for nodes in the microbial association network for beef carcasses (De Filippis et al., 2013). Relative abundance of OTUs and betweenness centrality are shown as size and transparency of symbols respectively. Only the labels of the nodes which are in the top 20% range for the properties shown in the graph are shown. The horizontal and vertical dotted lines indicate the average degree and positive degree, respectively, while the diagonal dotted line indicates the average ratio of positive to total degree.

Fig. 7. Distribution of the ratio between observed and expected frequency of copresence relationships among members of the same class (cop. same class) and among members of the same family (cop. same family) and of mutual exclusion relationships among members of different classes (mutE, diff. class) and family (mutE, diff. family). The probability of the values exceeding random expectation is also shown.

Table 1. List of studies used in the inference of microbial association networks. See Supplementary Table S1 for further details.

Label	Nature	Target	Region	Pipeline	Environment	OTUs	Reference
1	cs	DNA	V1-V3	UniNA	Dairy	79	Guidone et al., 2016
2	m	DNA	V1-V3	UniNA	Dairy	47	Parente et al., 2016c
3	cs	DNA	V1-V3	UniNA	Dairy	79	De Filippis et al., 2014
8	m	RNA	V1-V3	UniNA	Dairy	293	Dolci et al., 2014
9	ts	RNA DNA or	V1-V3	UniNA	Dairy	341	Alessandria et al., 2016
10	ts	RNA	V1-V3	UniNA	Dairy	128	De Filippis et al., 2016
11	m	DNA	V1-V3	UniNA	Meat	713	De Filippis et al., 2013
12	ts	RNA	V1-V3	UniNA	Meat	75	Greppi et al., 2015
13	ts	RNA	V1-V3	UniNA	Meat	126	Ferrocino et al., 2015
18	m	DNA	V1-V3	UniNA	Dairy	189	Stellato et al., 2015a
19	cs	DNA	V1-V3	UniNA	Miscellaneous Food	513	Pothakos et al., 2015
20	cs	DNA	V1-V3	UniNA	environment	535	Stellato et al., 2015b
21	cs	DNA	V1-V3	UniNA	Meat	403	Stellato et al., 2016
22	m	DNA	V1-V3	UniNA	Dairy	257	Calasso et al., 2016
29	ts	DNA	V1-V3	UniNA	Meat	816	Hultman et al., 2015
30	cs	DNA	V5	UniNA	Vegetables	813	Leff and Frierer, 2013
31	m	DNA	V1-V3	UniNA	Miscellaneous food	900	Chaillou et al., 2015

Nature: cs, cross-sectional; ts, time series; m, mixed. **Target:** 16S RNA gene (DNA) or 16S cDNA (RNA). **Region:** region of the target used in the study; **Pipeline:** de novo OTU picking strategy with UCLUST pipeline, taxonomy assigned by RDP, using Greengenes database. **Environment:** food or food environment. **OTUs:** number of the OTUs in the OTU count table used for inference.

Table 2. The forty most frequent copresence interactions detected in microbial association networks inferred using three methods on 17 studies (Table 1) of the food bacterial microbiome. The number of times a given association was inferred by each method is shown. Species used as starters in food fermentations are underlined while spoilage microorganisms are in bold.

Interaction	CoNet	SPIEC-EASI	SparCC
<u>Lactobacillus delbrueckii</u> <u>Streptococcus thermophilus</u>	6	3	7
<i>Propionibacterium acnes</i> <i>Staphylococcus</i>	7	4	4
<u>Lactococcus lactis</u> <u>Streptococcus thermophilus</u>	6	2	5
Acinetobacter Chryseobacterium	7		5
<i>Carnobacterium</i> <i>Lactococcus</i>	6		6
<i>Lactobacillus</i> <i>Leuconostoc</i>	7		5
Brochothrix <i>Carnobacterium</i>	5	2	4
Brochothrix <i>Pseudomonas</i>	5		6
<u>Lactobacillus helveticus</u> <u>Streptococcus thermophilus</u>	6	1	4
Acinetobacter johnsonii Chryseobacterium	5	1	4
Acinetobacter johnsonii <i>Stenotrophomonas</i>	7		3
Acinetobacter <i>Pseudomonas</i>	5	1	4
Chryseobacterium <i>Moraxellaceae</i>	6	1	3
Brochothrix <i>Pseudomonas fragi</i>	4		5
<i>Enterococcus</i> <i>Streptococcus</i>	6		3
<i>Lactobacillus delbrueckii</i> <i>Streptococcus</i>	4	1	4
<i>Actinobacteria</i> (class) <i>Paracoccus</i>	5	1	2
Brochothrix <i>Lactococcus</i>	4	1	3
Brochothrix <i>Psychrobacter</i>	5		3
Chryseobacterium <i>Stenotrophomonas</i>	6		2
Enterobacteriaceae <i>Pseudomonas</i>	3		5
<u>Lactobacillus delbrueckii</u> <u>Lactococcus lactis</u>	4		4
<u>Lactobacillus helveticus</u> <u>Lactococcus lactis</u>	5	1	2
<u>Lactococcus lactis</u> <i>Streptococcus</i>	4		4
<i>Lactococcus</i> <i>Leuconostoc</i>	4		4
<i>Microbacterium</i> Sphingomonas	6		2
<i>Propionibacterium acnes</i> <i>Streptococcus</i>	4	2	2
Pseudomonas fragi Pseudomonas veronii	4	1	3
Acinetobacter johnsonii <i>Staphylococcus</i>	5		2
Acinetobacter Enterobacteriaceae	5		2
Acinetobacter <i>Pseudomonas fragi</i>	2	1	4
Carnobacterium <i>Serratia</i>	5		2
Chryseobacterium <i>Comamonas</i>	6		1
Chryseobacterium <i>Microbacterium</i>	4		3
Chryseobacterium <i>Pseudomonas</i>	5		2
<u>Lactobacillus delbrueckii</u> <u>Lactobacillus helveticus</u>	5		2
<u>Lactobacillus helveticus</u> <i>Streptococcus</i>	5		2
<i>Lactobacillus</i> <i>Lactococcus</i>	3		4
<i>Microbacterium</i> <i>Paracoccus</i>	5		2
<i>Staphylococcus</i> <i>Streptococcus</i>	5		2

ACCEPTED MANUSCRIPT

Table 3. The forty most frequent mutual exclusion interactions detected in microbial association networks inferred using three methods on 17 studies (Table 1) of the food bacterial microbiome. The number of times a given association was inferred by each method is shown. Species used as starters in food fermentations are underlined while spoilage microorganisms are in bold.

Interaction	CoNet	SPIEC-EASI	SparCC
<i>Acinetobacter</i> _Lactobacillus	4		4
<i>Acinetobacter</i> _Leuconostoc	4	1	2
<u>Lactobacillus delbrueckii</u> Pseudomonas	4		3
Pseudomonas fragi <u>Streptococcus thermophilus</u>	3	1	3
<i>Acinetobacter johnsonii</i> _Lactobacillus	4		2
<i>Acinetobacter johnsonii</i> _Leuconostoc	2		4
<i>Acinetobacter</i> _Lactobacillus delbrueckii	4		2
<i>Acinetobacter</i> _Streptococcus thermophilus	1	1	4
Brochothrix _Lactococcus lactis	3		3
Brochothrix _Streptococcus thermophilus	3		3
Carnobacterium _Lactobacillus delbrueckii	4		2
Chryseobacterium _Lactobacillus	3		3
Chryseobacterium _Streptococcus thermophilus	2		4
<u>Lactobacillus delbrueckii</u> Pseudomonas fragi	4		2
Lactobacillus_ Pseudomonas	3		3
Lactobacillus_ Pseudomonas fragi	3		3
Pseudomonas _Sphingomonas	4		2
<i>Acinetobacter johnsonii</i> _Lactobacillus delbrueckii	4		1
<i>Acinetobacter johnsonii</i> _Streptococcus	1		4
<i>Acinetobacter</i> _Lactobacillus helveticus	3		2
<i>Acinetobacter</i> _Lactobacillus zeae	4		1
Actinobacteria (class)_ Pseudomonas fragi	2		3
Bacilli_ Pseudomonas	1		4
Bacillus_ Pseudomonas fragi	2		3
Brochothrix _Lactobacillus	2		3
Lactobacillus zeae_ Pseudomonas fragi	3		2
Moraxellaceae _Streptococcus thermophilus	3		2
Pseudomonas fragi _Sphingomonas	2	1	2
Stenotrophomonas _Streptococcus	2	1	2
<i>Acinetobacter johnsonii</i> _Lactobacillus zeae	3		1
<i>Acinetobacter johnsonii</i> _Pseudomonas veronii	2		2
<i>Acinetobacter johnsonii</i> _Streptococcus thermophilus	2		2
Bacilli_ Pseudomonas fragi			4
Brochothrix _Lactobacillus fuchuensis	1	1	2
Brochothrix _Sphingomonas	3		1
Carnobacterium _Lactobacillus helveticus	2		2
Carnobacterium _Streptococcus thermophilus	3		1
Chryseobacterium _Lactobacillus delbrueckii	2		2
Chryseobacterium _Leuconostoc	2		2
Flavobacterium _Leuconostoc	2		2

ACCEPTED MANUSCRIPT

Highlights

- microbial association networks were inferred for 17 food microbiome studies
- networks for foods were less complex than those for host and soil microbiomes
- association networks for food biomes shared several properties with host biomes
- emergence of negative hubs was characteristic of both spoilage and fermentation
- positive and negative interactions detected by three inference method are shown