



HAL
open science

Integration of metataxonomic datasets into microbial association networks highlights shared bacterial community dynamics in fermented vegetables

Romane Junker, Florence Valence, Michel-Yves Mistou, Stéphane Chaillou,
Helene Chiapello

► To cite this version:

Romane Junker, Florence Valence, Michel-Yves Mistou, Stéphane Chaillou, Helene Chiapello. Integration of metataxonomic datasets into microbial association networks highlights shared bacterial community dynamics in fermented vegetables. 2023. hal-04286431

HAL Id: hal-04286431

<https://hal.inrae.fr/hal-04286431>

Preprint submitted on 15 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



Journals.ASM.org

1 **Integration of metataxonomic datasets into** 2 **microbial association networks highlights** 3 **shared bacterial community dynamics in** 4 **fermented vegetables**

5 **Romane Junker,^{1*} Florence Valence,² Michel-Yves Mistou,¹ Stéphane Chaillou,³ Hélène**
6 **Chiapello¹**

7 ¹Université Paris-Saclay, INRAE, MaIAGE, Jouy-en-Josas, France

8 ²INRAE, Agrocampus Ouest, STLO, Rennes, France

9 ³Université Paris-Saclay, INRAE, MICALIS, Jouy-en-Josas, France

10 * Address correspondence to Romane Junker, romane.junker@inrae.fr.

11 **ABSTRACT**

12 The management of food fermentation is still largely based on empirical knowledge, as
13 the dynamics of microbial communities and the underlying metabolic networks that
14 produce safe and nutritious products remain beyond our understanding. Although these
15 closed ecosystems contain relatively few taxa, they have not yet been thoroughly
16 characterized with respect to how their microbial communities interact and dynamically
17 evolve. However, with the increased availability of metataxonomic datasets on different
18 fermented vegetables, it is now possible to gain a comprehensive understanding of the
19 microbial relationships that structure plant fermentation.

20 In this study, we present a bioinformatics approach that integrates public
21 metataxonomic 16S datasets targeting fermented vegetables. Specifically, we developed a
22 method for exploring, comparing, and combining public 16S datasets in order to perform
23 meta-analyses of microbiota. The workflow includes steps for searching and selecting
24 public time-series datasets and constructing association networks of amplicon sequence
25 variants (ASVs) based on co-abundance metrics. Networks for individual datasets are
26 then integrated into a core network of significant associations. Microbial communities are
27 identified based on the comparison and clustering of ASV networks using the “stochastic
28 block model” method. When we applied this method to 10 public datasets (including a
29 total of 931 samples), we found that it was able to shed light on the dynamics of vegetable
30 fermentation by characterizing the processes of community succession among different
31 bacterial assemblages.

32 **IMPORTANCE**

33 Within the growing body of research on the bacterial communities involved in the
34 fermentation of vegetables, there is particular interest in discovering the species or
35 consortia that drive different fermentation steps. This integrative analysis demonstrates
36 that the reuse and integration of public microbiome datasets can provide new insights
37 into a little-known biotope. Our most important finding is the recurrent but transient
38 appearance, at the beginning of vegetable fermentation, of ASVs belonging to
39 *Enterobacterales* and their associations with ASVs belonging to *Lactobacillales*. These
40 findings could be applied in the design of new fermented products.

41 **INTRODUCTION**

42 Over the last 20 years, the development of low-cost sequencing technologies has led to
43 the creation of a large number of microbiome datasets, mainly generated using
44 metataxonomic analyses based on 16S rRNA metabarcoding technology. For example, the
45 number of papers using metataxonomic or metagenomic approaches to study the

46 microbial communities of food increased six-fold between 2015 and 2021, and currently
47 exceeds 600 [1]; similarly, within the NCBI database, the Taxonomy ID "Food
48 metagenome" (NCBI:txid870726) is associated with 770 BioProjects. In keeping with the
49 principles of Open Science, most of these publication-associated datasets are available in
50 public repositories such as SRA (the Sequence Read Archive of NCBI), ENA (the European
51 Nucleotide Archive of EBI), or DDBJ (the DNA Data Bank of Japan). To promote the reuse
52 of certain kinds of datasets, specialized databases have been developed, such as MGNIFY
53 for microbiome data [2]. The availability of such vast amounts of metataxonomic data
54 provides an unprecedented opportunity to develop new integrative tools for comparing
55 and better understanding various microbial ecosystems. However, these efforts face
56 numerous challenges related to data reusability (e.g., data availability, metadata quality,
57 data preprocessing) and the most appropriate ways of identifying biologically informative
58 features in a collection of metataxonomic studies. In this work, we address these
59 challenges by developing a method for exploring public datasets related to the microbiota
60 of fermented vegetables and performing meta-analyses of previous research (i.e., reusing
61 independent datasets, integrating them into a larger analysis to generate new knowledge).

62 Our choice of ecosystem was motivated by current interest in the bacterial
63 communities involved in the fermentation of vegetables [3, 4, 5]. Plant-based fermented
64 foods diversify human diets and possess interesting properties in terms of sustainability
65 and nutritional quality. These products require little energy to produce and preserve, and
66 their consumption confers several benefits on human health [6, 7]. With this study, we
67 wanted to assess whether public datasets that are already available for fermented
68 vegetables could help to improve our knowledge on the ecological dynamics taking place
69 in these products. Fermented vegetables are created through the (usually spontaneous)
70 activity of heterofermentative and homofermentative lactic acid bacteria (LAB) naturally
71 present on the raw material [8]. In Europe, the most popular example of this kind of food
72 is sauerkraut, for which the use of pre-selected starter strains remains uncommon even for
73 large-scale production [9]. A combination of low pH and the anaerobic conditions
74 resulting from the fermentation process are the main factors that select for the beneficial
75 anaerobic LAB essential in the production of good-quality fermented vegetables [3].

76 These bacteria are a broad and diverse group of species classified in phylum *Firmicutes*,
77 class *Bacilli*, and order *Lactobacillales*, and include representatives from the families
78 *Lactobacillaceae*, *Streptococcaceae*, *Enterococcaceae*, *Carnobacteriaceae*, and *Aerococcaceae* [10].

79 It should be noted that, to date, most studies have focused on describing the microbial
80 communities present at the end of the fermentation process [4, 5], while the dynamic
81 succession of various microbial populations during fermentation has received little
82 attention. This represents an important gap in knowledge, especially when compared, for
83 example, to research on cheese microbial communities which has revealed that the proper
84 succession of microbial populations is important to the quality of the final product [11, 12].
85 Two separate metataxonomic analyses that have revealed important changes in microbial
86 dynamics during vegetable fermentation. A study on carrot juice reported a succession
87 process involving *Enterobacteriaceae*, *Leuconostoc*, and *Lactobacillus*, while work on Suan Cai
88 (Chinese pickles) showed that the dominant species changed from early stages of
89 fermentation (*Leuconostoc mesenteroides*) to later ones (*Lactiplantibacillus plantarum*) Wuyts
90 et al. [13], Yang et al. [14]. The little information that can be gathered on the subject does
91 not allow us to identify species or consortia that might be responsible for controlling
92 various stages of fermentation among different vegetables. In this context, the use of
93 metataxonomic data to carry out meta-analysis could prove illuminating.

94 The use and comparison of amplicon data (such as the 16S-based data considered in
95 the present work) raises certain difficulties. First, sequencing technology may vary among
96 studies, as may the region amplified or PCR primers employed. Second, taxonomic
97 assignment based on the 16S variable region is considered valid only to the genus level,
98 limiting species-level interpretations [4]. There are therefore two possibilities for carrying
99 out a comparative study of multiple datasets: comparing genus-level taxonomic profiles
100 or comparing exact sequences, specifically, amplicon sequence variants (ASVs). The
101 advantages of the first approach include the ability to compare different sequenced
102 regions and to reduce the sparsity of the count matrices, while the use of ASVs enables
103 intra-genus diversity to be taken into account [15, 16]. In both cases, the aim of this type
104 of meta-analysis is often to identify core taxa based on criteria of abundance and
105 prevalence [17].

106 The analytical design of such a study is also important. One promising approach for
107 meta-analysis is the construction of microbial association networks, which provide
108 additional and complementary information to classic analyses of alpha- and
109 beta-diversity [18]. Association networks enable the identification of hub species [19, 20],
110 taxa clusters [21], and core networks, the last of which corresponds to the intersection of
111 several microbial association networks and can be used to identify taxa and associations
112 shared by most networks [22]. Association networks were originally designed for
113 macroscopic ecosystems and have only recently been adapted for the investigation of
114 interactions within microbial assemblages [21]. They are constructed using count data
115 from the sequenced environment, which are compositional [23], high-dimensional, and in
116 the form of sparse matrices, thus increasing the difficulty of analysis [21]. However,
117 compared to networks from other assemblages, the association networks in fermented
118 ecosystems appear to be significantly smaller [16], making them easier to construct,
119 visualize, and compare. According to Chen et al. [24], association networks can be
120 divided into four categories, which are built using different approaches: correlation
121 networks (CoNet [25], SparCC [26]), conditional correlation networks (SPIEC-EASI [27]),
122 mixture networks (MixMPLN [28]), and differential networks (DCDTr). Due to the
123 complexity of microbial interactions, all these approaches have important limitations, and
124 no method has yet managed to capture all of the aspects of interest. Indeed, studies have
125 even shown that classical measures such as Pearson and Spearman correlations can
126 perform just as well as computationally time-consuming methods based on more
127 sophisticated statistical models [29, 30].

128 This study presents an integrative bioinformatics approach for the meta-analysis of
129 public amplicon datasets. The workflow includes steps designed to search for and select
130 public time-series datasets and construct ASV association networks based on
131 co-abundance metrics. Microbial communities are then analyzed by comparing and
132 clustering the ASV networks. We applied this workflow to 10 publicly available datasets
133 on the microbial assemblages of fermented vegetables. Here, we describe the value of this
134 approach for discovering core bacterial taxa and core associations shared by different
135 vegetables during the process of fermentation.

136 **RESULTS**

137 **Design of a bioinformatics workflow for integration of metataxonomic datasets**

138 Figure 1 depicts the main steps of the bioinformatics workflow designed to analyze
139 and integrate the amplicon datasets. The first step involved the careful selection of public
140 datasets focused on the microbial communities of fermented vegetables. Next, ASV count
141 tables were constructed for each of the selected studies. Using these count tables, we then
142 produced ASV association networks for each study that were based on four sensitive and
143 computationally efficient metrics: Jaccard distance, Pearson and Spearman correlations
144 between relative abundances, and a proportionality measure calculated from
145 clr-transformed abundances. The purpose of the networks was to help visualize how
146 microbial communities interact and evolve dynamically. Finally, the various networks
147 were integrated together. A core network was constructed that identified which bacterial
148 ASVs were common to most fermentations and which associations between ASVs were
149 significantly shared among networks. In addition, a multiple SBM clustering method was
150 used to identify a set of ASVs that were associated with each other across the different
151 networks.

152 **Selection of metataxonomic studies on fermented vegetables**

153 Ten datasets meeting our selection criteria (see Materials and Methods section) were
154 obtained out of 1443 studies from SRA (NCBI), 10 studies from MGnify (ENA), and 3
155 studies from FoodMicrobioNet. All datasets contained sequences of the V3–V4 or V4
156 hypervariable region, enabling ASV comparison. The selected datasets originated from
157 studies on five different varieties of vegetables (cucumber, carrot, cabbage, pepper, radish,
158 used alone or in a mixture) and comprised between 18 and 310 samples each, for a total of
159 931 samples (Table 1). The time scales that were examined varied among studies, as the
160 datasets included between 2 and 12 time points. Depending on the study in question,
161 monitoring began between 0 and 30 days after the beginning of fermentation and ended
162 between 3 and 720 days after. All studies were conducted on spontaneous fermentations,
163 with the exception of PRJNA751723 and PRJNA662831, which included samples from

164 spontaneous fermentations as well as samples inoculated with various LAB
165 (*Latilactobacillus curvatus*, *Leuconostoc gelidum*, *Latilactobacillus sakei*, or *Weissella koreensis*).
166 Dataset PRJEB15657 contained data from two sets of experiments (samples from a
167 laboratory experiment and samples from a citizen science experiment), which we divided
168 into two subsets.

169 **Visualization of microbial succession during fermentation through the construction of** 170 **association networks**

171 Historically, bar graphs have been used to visualize changes in the taxonomic
172 composition of bacterial communities between samples. However, this method does not
173 reflect the evolution of ASV associations over time. Microbial association networks, on the
174 other hand, highlight these temporal taxonomic associations and can visually present
175 information that is complementary to bar graphs. For each of the 10 datasets, we built
176 association networks, of which one is presented in Figures 2A and 2B (study
177 PRJNA689239, paocai fermentation over 30 days, captured at six timepoints). This
178 network appeared to be composed of two subnetworks: one containing a high diversity of
179 ASVs (including *Pseudomonadales* and *Enterobacterales*) with a weighted mean age (WMA)
180 between 0 and 10 days, and the other containing a lower diversity of ASVs belonging to
181 *Enterobacterales* and *Lactobacillales*, with a higher WMA (between 8 and 30 days). These
182 observations suggest that there is a shift during fermentation from a broad initial diversity
183 of ASVs to an assemblage dominated by LAB. Interestingly, we observed the same
184 patterns in the PRJNA564474 study (Fig. 2D; kimchi fermentation over 50 days and eight
185 timepoints). However, a notable difference from the paocai study was that the first
186 subnetwork was present at WMAs ranging from 0 to 50 days, and the second, composed
187 only of *Lactobacillales* ASVs, appeared at 10 to 50 days. This structure suggests that some
188 of the samples failed to ferment, as observed for sample SRR10127549 in the bar graph.

189 A similar network pattern was observed for 8 of the 11 networks analyzed (Fig. S1).
190 The overall pattern could be described as follows: samples initially contained a high
191 diversity of ASVs (featuring *Pseudomonadales* in particular) with a low WMA; then, as the
192 WMA increased, nodes corresponding to ASVs from *Enterobacterales* and then

193 *Lactobacillales* appeared, with numerous associations between them. However, we would
194 like to emphasize a few points to keep in mind when interpreting these networks. The
195 WMA of an ASV does not reflect the exact time point at which the ASV first appears.
196 Indeed, during each of the vegetable fermentations, all ASVs were present from the
197 beginning of the fermentation process. This measurement may also represent both living
198 and dead bacterial populations because the DNA of dead bacteria may be recovered and
199 sequenced as well. Hence, the use of WMA to organize an association network merely
200 provides a general picture of the temporal dynamics of ASVs over a fermentation process,
201 highlighting the main "peaks" of presence and potential species associations.

202 We also analyzed the three networks that did not exhibit this succession of
203 communities (PRJNA473189, PRJNA662831, PRJNA544161; see Fig. S1). A common
204 feature of these three studies was a shift in timing compared to the others: more precisely,
205 sampling did not start until three days after the onset of fermentation. Therefore, it is
206 possible that the successional shift in microbial communities took place before the first
207 sampling point. This hypothesis is supported by the observation that the taxonomic
208 profile of the pepper and sauerkraut samples (PRJNA473189 and PRJNA662831) did not
209 change over time. In the case of doubanjiang (PRJNA544161), a fermented product
210 containing numerous ingredients (beans, soya, rice, spices), ASVs belonging to
211 *Enterobacteriales* appeared to proliferate relatively late, as observed on the bar graph (Fig.
212 S1).

213 **Comparison of association networks to identify a core network of bacterial** 214 **communities**

215 To integrate the 11 association networks, we constructed a core network, i.e., the
216 intersection of several networks (Fig. 3). In Figure 3A, it can be seen that the 11 networks
217 shared 3 vertices (ASVs) overall, and pairwise analyses revealed between 10 and 58
218 vertices that were shared between a given pair of networks. Similarly, pairwise analyses
219 detected between 3 and 296 edges that were shared by two networks, but no edges were
220 shared by more than nine networks (Fig. 3B). To evaluate the statistical significance of the
221 edge intersections, we compared them with a null model using a Kolmogorov-Smirnov

222 test; the results rejected the null hypothesis that our set of networks followed the same
223 distribution as the null model for intersections between two, three, four, five, or six
224 networks (p -value < 0.05 for 100 cases). This means that those network subsets share
225 associations in a significant way.

226 We then constructed core networks based on the intersections between two to six
227 networks (all shown in Fig. S2). The core network built using microbial associations
228 present in at least three networks (Fig. 3C) included 97 ASVs (out of a total of 975 used to
229 construct the 11 networks). Among them, 13 were affiliated with order *Pseudomonadales*,
230 17 with *Enterobacterales*, and 25 with *Lactobacillales*. In representing the core network, we
231 used the scaled WMA on the x-axis. The rationale of the scaled WMA was to normalize
232 time data and to establish a common time scale between the various studies. Indeed, the
233 WMAs are not directly comparable between studies because the time points measured
234 varied from one study to another.

235 Analysis of the different significant core networks revealed that, despite all of the
236 differences between experiments (type of sequencing, fermentation conditions, time
237 scale), there appeared to be a common temporal structure in the microbial dynamics of
238 fermented vegetables. In particular, after a mean scaled WMA of 0.5, *Lactobacillales* ASVs
239 tended to predominate. Furthermore, we also observed a shift from the initial microbial
240 population of vegetables to one dominated by *Enterobacterales*, and then a second shift to
241 *Lactobacillales*.

242 This observation was confirmed by a clear difference in scaled WMA among all ASVs
243 corresponding to *Pseudomonadales*, *Enterobacterales*, and *Lactobacillales*, as shown in Figure
244 3D. Figure 3E highlights this trend and also shows that the ASVs with the lowest and
245 highest scaled WMAs were less often shared among studies (less than three graphs when
246 WMA was lower than -1 or higher than 2) than those with median WMA values. This
247 suggests that the initial flora, as well as the LAB present mainly at the end of fermentation,
248 tended to be more specific to a given experiment than other ASVs. Moreover, the ASVs
249 belonging to *Enterobacterales* were more likely to be shared between networks than those
250 corresponding to *Lactobacillales* (non-parametric Wilcoxon-Mann-Whitney test, p -value =

251 0.02). In fact, of the three ASVs that were detected in all experiments, all belonged to the
252 *Enterobacterales* (*Klebsiella*, *Pectobacterium*, and an unidentified *Enterobacterales*).

253 Finally, we investigated distinctions between different genera within family
254 *Lactobacillaceae* (following the new taxonomy of Zheng et al. [31]) based on the type of
255 fermentation performed. In the core network, ASVs belonging to genera that perform
256 hetero-lactic fermentation were more numerous than those belonging to genera that
257 perform homo-lactic fermentation. Moreover, most members of the *Lactobacillales* were
258 found in only one graph (143 out of 208, i.e., 69%), and among those shared by more than
259 two graphs, 18 perform heterofermentation and 7 perform homofermentation. We can
260 therefore conclude that LABs are generally highly specific to a fermentation process, and
261 the ASVs that are shared among different processes are mostly heterofermentative. There
262 was no significant difference between the scaled WMA of heterofermentative and
263 homofermentative genera, but we did detect some expected successional shifts in genera
264 (Fig. 3F: *Leuconostoc* and *Lactiplantibacillus*, p-value = 0.02).

265 **Multiple clustering to identify putative bacterial consortia shared among studies**

266 To identify sets of ASVs that were connected in similar ways across the 11 microbial
267 association networks, we applied the multiplex stochastic block model (SBM) graph
268 clustering method. Ten different clusters were identified, which varied in their size and
269 the prevalence and taxonomy of their member ASVs. All ASVs within a cluster shared
270 similar intra-cluster and inter-cluster connection patterns. Clusters 1 to 5 contained few
271 ASVs (between 5 and 45) that were shared between two or more networks, while clusters
272 6 to 10 contained many ASVs (between 94 and 463) that were mainly specific to one
273 network (Fig. 4A). ASVs affiliated with *Lactobacillales* predominantly belonged to clusters
274 5, 9, and 10; this last group contained most of the *Lactobacillales* ASVs and those
275 corresponding to the diverse initial microflora, i.e., those that were specific to each
276 experiment.

277 Among the different clusters, clusters 1 and 5 were particularly interesting, as they
278 included the majority of ASVs that were shared by more than five networks and they

279 were the predominant clusters in the core network (Fig. 4B). Moreover, ASVs in the two
280 clusters differed significantly in their scaled WMA (p-value = 0.013). It is possible that the
281 ASVs in these two clusters correspond to successive bacterial communities that are
282 common in vegetable fermentation: cluster 1 appeared to be highly associated with ASVs
283 in the orders *Pseudomonadales* and *Enterobacterales*, while cluster 5 represented an
284 assemblage of *Enterobacterales* and *Lactobacillales*. This community of conserved ASVs
285 could potentially represent a shared core-consortia of early fermentation; its detailed
286 composition is shown in Table S1.

287 **DISCUSSION**

288 This work presents an integrative bioinformatics approach that utilizes association
289 networks to combine different sets of publicly available data on the microbial dynamics of
290 fermentation in vegetables. By combining ASV networks from different studies, we
291 obtained valuable insights into bacterial community structure during different phases of
292 fermentation. Historically, association networks have been used to detect potential
293 inter-species interactions; here, we adapted this strategy to identify and visualize ASVs
294 with similar temporal dynamics. To our knowledge, this work is the first to construct a
295 core network representing the fermentation of different vegetables based on sequence
296 data from multiple independent datasets. By integrating several public datasets together,
297 we were able to characterize two successional shifts that were conserved among different
298 fermentation ecosystems: the first from the initial microbial population of vegetables to
299 *Enterobacterales*, and the second to an assemblage dominated by *Lactobacillales*. To test the
300 significance of the core network we obtained, we used an approach based on comparison
301 to a null model, which was similar to that developed by Röttjers et al. [22], with a
302 sampling of random graphs similar to Doane et al. [32]. Indeed, the identification of core
303 networks is a more challenging task than computation of the global intersection network
304 [21]. With these tests, we determined that some intersections between networks would
305 not be expected by random chance, and thus that some edges may correspond to genuine
306 ASV dynamics shared among several studies. Finally, we complemented this approach by
307 using the SBM method for ASV clustering, which is a technique applicable to multiplexes

308 (a type of multi-layer network) that does not require any a priori assumptions regarding
309 connectivity patterns. The SBM model has been used for community detection in various
310 fields, such as sociology. More recently, it has been applied to taxonomic profiling of the
311 human microbiome in order to uncover patterns of community structure. Specifically, it
312 was used as a bipartite model for clustering samples and taxa [33]. In another study, the
313 simple SBM enabled the detection of OTU clusters based on their connectivity patterns in
314 a co-occurrence network [34]. In the present work, we applied the multiplex version of
315 this model to a collection of networks in order to identify clusters of ASVs that share
316 similar patterns of associations across the different networks. We were able to identify 10
317 clusters of ASVs, which could be used to guide the exploration and delineation of new
318 bacterial consortia in fermented vegetables [35].

319 With respect to the microbial ecology of fermented vegetables, our most important
320 finding was the recurring and transient appearance, at the beginning of fermentation, of
321 ASVs belonging to *Enterobacterales* and their association with ASVs affiliated with
322 *Lactobacillales*. This raises the question of their ecological function in vegetable
323 fermentation and their impact on the properties of the final product. The hypothesis of
324 bacterial succession in vegetable fermentation, from *Enterobacteria* to heterolactic and
325 homolactic acid bacteria, is not entirely new. However, due to the small number of studies
326 carried out on the subject and the extensive variability in the methodologies used, most
327 reports have not generated convincing conclusions on the impact of *Enterobacterales* and
328 their possible interactions with LAB. Nevertheless, based on the existing literature,
329 several hypotheses can be put forward. *Enterobacterales* may have fermentative properties,
330 or they may participate in nutritional mutualism that is beneficial to the development of
331 LAB. Indeed, certain trophic relationships between LAB and *Enterobacteriaceae* have
332 already been described. For example, some LAB generate metabolic energy using an
333 agmatine deiminase pathway that relies on agmatine produced by *Enterobacteriaceae* [36].
334 In the wet coffee fermentation process, the first phase involves interactions between
335 *Enterobacteriaceae* (with pectinolytic activity), acetic acid bacteria, and some yeasts [37].
336 *Enterobacteriaceae* have also been found in two other studies on fermented vegetables

337 [38, 39], of which the former hypothesizes that the presence of *Erwinia sp.* may reflect its
338 ability to invade compromised plant tissues or its potential ability to ferment sugar.

339 The meta-analysis we designed is particularly well-suited to fermented vegetable
340 ecosystems: since these ecosystems are closed, contain relatively few taxa, and undergo a
341 temporal succession of communities, the representation of ASV association networks is
342 fairly easy to visualize and interpret. This approach could be easily applied to amplicon
343 or shotgun metagenomic data for other fermented foods characterized by closed
344 ecosystems with community shifts. One limitation of the present meta-analysis is that it
345 was carried out on a relatively small scale (on 10 independent datasets including a total of
346 931 samples), due to the small number of reusable public metabarcoding datasets on
347 fermented vegetables. This is mainly due to difficulties in accessing raw data (some
348 samples are missing, some data are pre-processed, etc.) and metadata (sometimes
349 incomplete and inconsistent, with manual extract from paper required). Indeed, these
350 limitations were highlighted in a recent article [40], which recommended that data be
351 deposited in public repositories together with assay metadata (technical features of the
352 experiment) and biological metadata (environmental conditions of the biosamples). This,
353 along with the adoption of other best practices, will enable wider reuse and integration of
354 microbiome datasets on a broader scale.

355 This study is based on 16S metataxonomic data, more specifically, the V4
356 hypervariable region because it was used in the majority of the datasets found. This
357 region is the most frequent target of studies focused on food ecosystems, along with the
358 V3–V4 region of the 16S rRNA gene [1]. Unfortunately, this gene region has poor
359 discriminatory power; it is able to provide reliable taxonomic assignment at the
360 genus-level only and cannot be used to study species-level diversity (unlike, for instance,
361 the V1–V3 region [41]). Therefore, although it is interesting to discover ASVs that are
362 shared between different studies, this approach is ill-suited for characterizing the species-
363 and strain-level diversity of *Lactobacillales* and *Enterobacterales*. Furthermore, the read
364 count tables obtained for the different studies can be shaped by many biases, including
365 differences in sample collection and storage, DNA extraction method and primer choice,
366 variation in the number of rRNA operons [41, 42], amplification of extracellular DNA, and

367 errors in taxonomic affiliations. Therefore, the results of any individual ASV count table
368 must be interpreted cautiously. However, in the context of our study, the use of ASVs
369 enabled direct comparison of sequences between studies and reduced the influence of
370 taxonomic misclassifications [43, 44]. In addition, integrating ASVs into association
371 networks allowed comparisons of similar dynamics between ASVs in different studies,
372 and limited the biases that might arise from direct comparison of relative abundances.

373 This work demonstrates the effectiveness of using association networks for temporal
374 meta-analysis. The approach we developed could easily be applied to new datasets or
375 extended to incorporate new tools for association network inference, core network
376 detection, and clustering. In the future, it could be interesting to integrate additional
377 sample metadata (such as temperature, lactic acid concentration, pH, and/or salinity) if
378 they were available in a standardized format and could be easily integrated to an
379 association network. This approach could lead to the design of ideal consortia that could
380 make vegetable fermentation safer [45] (Capozzi et al., 2017), more reproducible, and
381 exploitable on a large scale [46].

382 Finally, the taxonomic profile inferred from 16S rRNA is not able to provide insights
383 into the functional profile of bacterial communities or into the part(s) played by other
384 microorganisms (even if their presence is minor, e.g., less than 5% relative abundance for
385 fungi and *Archaea* in brine food according to Leech et al. [4]). Ultimately, there is a need
386 for complementary functional studies (shotgun metagenomics, metatranscriptomics) to
387 improve our understanding of vegetable fermentation and assess the functional
388 interactions taking place during this process.

389 **MATERIALS AND METHODS**

390 **Study selection**

391 Datasets were obtained from three repositories: the MGnify database (on microbiome
392 data), the FoodMicrobioNet database (on food ecosystems), and the NCBI SRA database.
393 Studies focused on the microbial ecosystems in fermented vegetables were identified in

394 MGnify by selecting the biome "Food production" and filtering with the term "Fermented
395 vegetables", while in FoodMicrobionet, we used the spoilage filter "Fermented" from
396 studies labeled with "Vegetables and vegetable products". From NCBI/SRA, we retrieved
397 studies with the Taxonomy IDs "Food metagenome" (870726), "Fermentation
398 metagenome" (1326787), and "Food fermentation metagenome" (1154581). Of the resulting
399 studies, the only ones that were considered were those whose "SRA Run Selector"
400 metadata contained the words "day", "week", "month", "hour," or "time", and that had an
401 associated publication on fermented vegetables.

402 We included only studies that examined at least two time points, contained more than
403 10 samples, and were associated with a publication (to ensure access to extensive
404 metadata). Finally, we retained only studies that sequenced the V4 or V3–V4
405 hypervariable region of the 16S rRNA gene to permit comparisons of ASVs. Raw
406 sequencing data of the resulting selected studies were retrieved from the NCBI SRA
407 repository using home-made scripts.

408 **Construction of ASV count tables**

409 Sequencing data from each study were processed using the dada2 pipeline [43] for
410 read quality control, read filtering and trimming (with parameters truncLen = 240 or 220
411 depending on read length, maxN = 0, maxEE = 2, truncQ = 2), error rate learning and ASV
412 inference, paired read assembly (with parameter minOverlap = 3), chimera removal, and
413 taxonomic assignment to kingdom, phylum, class, order, family, and genus (using Silva
414 database nr 99 v 138. 1). For the five studies in which the V3–V4 region of the 16S rRNA
415 gene was sequenced, only the V4 region was retained. The ASV count table for each
416 sample, the ASV taxonomy table, and the sample metadata were combined into one
417 phyloseq object [47] for each study. ASVs matching mitochondrial or chloroplast DNA
418 and samples from negative fermentation controls were excluded from the count tables.

419 **Inference of microbial association networks**

420 For each study a count table was filtered to create a microbial association network.
421 Only non-control samples with more than 15,000 reads and ASVs found in at least three
422 samples and with an average relative abundance greater than $1e-5$ were included. We
423 chose association metrics that take into account co-presence, with Jaccard distance, as well
424 as co-abundance, with Pearson and Spearman correlations based on relative abundances.
425 The proportionality measure proposed by Lovell et al. [48] ($\Phi(a, b) = \frac{\text{var}(\text{clr}(a) - \text{clr}(b))}{\text{var}(\text{clr}(a) + \text{clr}(b))}$) was
426 also used following centered log-ratio transformation, performed using the function
427 `aldex.clr` from the package `ALDEx2` Quinn et al. [49]. Edges were traced if at least one of
428 these four measures reached a non-stringent threshold (0.4 for Jaccard distance and 0.5 for
429 the three other measures). The thickness of each edge reflected the number of combined
430 metrics supporting it. A force-driven algorithm (Fruchterman-Reingold) was used to
431 calculate the layout of each association network. This layout was preserved on the y-axis,
432 but the x-axis was modified: the position of each ASV was the mean age of the samples in
433 which the ASV was present, weighted by its relative abundance (hereafter named WMA
434 for weighted mean age).

435 **Core network construction**

436 The core network was constructed based on the intersections of the independent
437 association networks created for each study. To account for the different sampling time
438 points and fermentation rates among studies, the x-axis position of each ASV in the core
439 network corresponded to the average of its centered and scaled positions in the original
440 networks. A null-model statistical test was used to assess the significance of the core
441 networks constructed from edges shared by a subset of networks or by all networks. First,
442 we generated 100 sets of networks with the same nodes as the networks of interest but
443 with random edges, using the “rewire” function of the `igraph` R package with `prob = 1`.
444 Next, the distribution of edges shared by a given subset of networks or by all networks
445 was compared between each null model and the studied set of networks with a
446 Kolmogorov-Smirnov test.

447 **SBM multiplex clustering**

448 Multiplex networks refer to a collection of networks involving the same sets of nodes
449 but originating from different types of relationships. Here, each network corresponded to
450 a specific study and each node corresponded to an ASV. SBM clustering was applied to
451 multiplex networks to assign each ASV to a community (or block) according to its
452 connection patterns. The estimateMultiplexSBM function from the R package sbm [50]
453 was used with a Poisson model describing the relationship between the nodes. The
454 number of blocks was chosen using a penalized likelihood criterion (ICL), and the
455 likelihood maximization was obtained via a variational version of the
456 Expectation-Maximization algorithm.

457 **Statistical analysis and figure construction**

458 To compare WMA or the prevalence among studies of ASVs belonging to different
459 groups (taxonomic rank or SBM cluster), the non-parametric Wilcoxon-Mann-Whitney
460 test was performed. To create figures, the R packages ggplot2, viridis, ggpubr, and
461 ComplexHeatmap were used [51, 52].

462 **ACKNOWLEDGMENTS**

463 We are grateful to INRAE MICA Division and the ENS Paris-Saclay for the funding of
464 this PhD and to INRAE MIGALE bioinformatics facility (MIGALE, INRAE, 2020. Migale
465 bioinformatics Facility, doi:10.15454/1.5572390655343293E12) for providing computing
466 and storage resources.

467 **DATA AVAILABILITY STATEMENT**

468 No new research data was generated in the preparation of this article.

469 **CONFLICTS OF INTEREST**

470 The authors declare no conflict of interest.

471 **REFERENCES**

472 **References**

- 473 [1] **Parente E, Zotta T, Ricciardi A.** Jul 2022. FoodMicrobionet v4: A large, integrated,
474 open and transparent database for food bacterial communities. *International Journal of*
475 *Food Microbiology* 372:109696. doi:[10.1016/j.ijfoodmicro.2022.109696](https://doi.org/10.1016/j.ijfoodmicro.2022.109696).
- 476 [2] **Mitchell AL, Almeida A, Beracochea M, Boland M, Burgin J, Cochrane G, Crusoe**
477 **MR, Kale V, Potter SC, Richardson LJ, Sakharova E, Scheremetjew M,**
478 **Korobeynikov A, Shlemov A, Kunyavskaya O, Lapidus A, Finn RD.** Jan 2020.
479 MGnify: the microbiome analysis resource in 2020. *Nucleic Acids Research* 48
480 (D1):D570–D578. doi:[10.1093/nar/gkz1035](https://doi.org/10.1093/nar/gkz1035).
- 481 [3] **Thierry A, Baty C, Marché L, Chuat V, Picard O, Lortal S, Valence F.** Sep 2023.
482 Lactofermentation of vegetables: An ancient method of preservation matching new
483 trends. *Trends in Food Science & Technology* 139:104112. doi:[10.1016/j.tifs.2023.07.009](https://doi.org/10.1016/j.tifs.2023.07.009).
- 484 [4] **Leech J, Cabrera-Rubio R, Walsh AM, Macori G, Walsh CJ, Barton W, Finnegan L,**
485 **Crispie F, O’Sullivan O, Claesson MJ, Cotter PD.** Nov 2020. Fermented-Food
486 Metagenomics Reveals Substrate-Associated Differences in Taxonomy and
487 Health-Associated and Antibiotic Resistance Determinants. *mSystems* 5 (6):e00522–20.
488 doi:[10.1128/mSystems.00522-20](https://doi.org/10.1128/mSystems.00522-20). Publisher: American Society for Microbiology.
- 489 [5] **Xu M, Su S, Zhang Z, Jiang S, Zhang J, Xu Y, Hu X.** 2022. Two sides of the same coin:
490 Meta-analysis uncovered the potential benefits and risks of traditional fermented
491 foods at a large geographical scale. *Frontiers in Microbiology* 13.
492 <https://www.frontiersin.org/articles/10.3389/fmicb.2022.1045096>. Retrieved
493 2023-03-30.

- 494 [6] **Lebeer S, Vanderleyden J, De Keersmaecker SCJ**. Dec 2008. Genes and Molecules of
495 Lactobacilli Supporting Probiotic Action. *Microbiology and Molecular Biology Reviews* 72
496 (4):728–764. doi:[10.1128/MMBR.00017-08](https://doi.org/10.1128/MMBR.00017-08). Publisher: American Society for
497 Microbiology.
- 498 [7] **Melini F, Melini V, Luziatelli F, Ficca AG, Ruzzi M**. May 2019. Health-Promoting
499 Components in Fermented Foods: An Up-to-Date Systematic Review. *Nutrients* 11
500 (5):1189. doi:[10.3390/nu11051189](https://doi.org/10.3390/nu11051189). Number: 5 Publisher: Multidisciplinary Digital
501 Publishing Institute.
- 502 [8] **Di Cagno R, Coda R, De Angelis M, Gobbetti M**. Feb 2013. Exploitation of vegetables
503 and fruits through lactic acid fermentation. *Food Microbiology* 33 (1):1–10. doi:
504 [10.1016/j.fm.2012.09.003](https://doi.org/10.1016/j.fm.2012.09.003).
- 505 [9] **Medina-Pradas E, Pérez-Díaz IM, Garrido-Fernández A, Arroyo-López FN**. Jan. 2017.
506 Chapter 9 - Review of Vegetable Fermentations With Particular Emphasis on
507 Processing Modifications, Microbial Ecology, and Spoilage, p 211–236. In Bevilacqua
508 A, Corbo MR, Sinigaglia M (ed), *The Microbiological Quality of Food*. Woodhead
509 Publishing Series in Food Science, Technology and Nutrition, Woodhead Publishing.
510 doi:[10.1016/B978-0-08-100502-6.00012-1](https://doi.org/10.1016/B978-0-08-100502-6.00012-1).
- 511 [10] **Ashaolu TJ, Reale A**. Aug 2020. A Holistic Review on Euro-Asian Lactic Acid
512 Bacteria Fermented Cereals and Vegetables. *Microorganisms* 8 (8):1176. doi:
513 [10.3390/microorganisms8081176](https://doi.org/10.3390/microorganisms8081176). Number: 8 Publisher: Multidisciplinary Digital
514 Publishing Institute.
- 515 [11] **Parente E, Zotta T, Ricciardi A**. May 2022. A review of methods for the inference and
516 experimental confirmation of microbial association networks in cheese. *International*
517 *Journal of Food Microbiology* 368:109618. doi:[10.1016/j.ijfoodmicro.2022.109618](https://doi.org/10.1016/j.ijfoodmicro.2022.109618).
- 518 [12] **Portik D, Dutton RJ, Ashby M, Dinh CB, Cotter PD, Saak CC, Pierce EC, Hall R**.
519 Jan 2023. Longitudinal, Multi-Platform Metagenomics Yields a High-Quality Genomic
520 Catalog and Guides an In Vitro Model for Cheese Communities. *mSystems* doi:
521 [10.1128/msystems.00701-22](https://doi.org/10.1128/msystems.00701-22).

- 522 [13] **Wuyts S, Van Beeck W, Oerlemans EFM, Wittouck S, Claes IJJ, De Boeck I, Weckx**
523 **S, Lievens B, De Vuyst L, Lebeer S.** Jun 2018. Carrot Juice Fermentations as
524 Man-Made Microbial Ecosystems Dominated by Lactic Acid Bacteria. *Applied and*
525 *Environmental Microbiology* 84 (12):e00134–18. doi:[10.1128/AEM.00134-18](https://doi.org/10.1128/AEM.00134-18).
- 526 [14] **Yang H, Wu H, Gao L, Jia H, Zhang Y, Cui Z, Li Y.** Dec 2016. Effects of *Lactobacillus*
527 *curvatus* and *Leuconostoc mesenteroides* on Suan Cai Fermentation in Northeast
528 China. *Journal of Microbiology and Biotechnology* 26 (12):2148–2158. doi:
529 [10.4014/jmb.1607.07010](https://doi.org/10.4014/jmb.1607.07010).
- 530 [15] **Berry D, Widder S.** 2014. Deciphering microbial interactions and detecting keystone
531 species with co-occurrence networks. *Frontiers in Microbiology* 5.
532 <https://www.frontiersin.org/articles/10.3389/fmicb.2014.00219>. Retrieved
533 2023-01-27.
- 534 [16] **Parente E, Zotta T, Faust K, De Filippis F, Ercolini D.** Aug 2018. Structure of
535 association networks in food bacterial communities. *Food Microbiology* 73:49–60. doi:
536 [10.1016/j.fm.2017.12.010](https://doi.org/10.1016/j.fm.2017.12.010).
- 537 [17] **Holman DB, Brunelle BW, Trachsel J, Allen HK.** May 2017. Meta-analysis To Define
538 a Core Microbiota in the Swine Gut. *mSystems* 2 (3):10.1128/msystems.00004–17. doi:
539 [10.1128/msystems.00004-17](https://doi.org/10.1128/msystems.00004-17). Publisher: American Society for Microbiology.
- 540 [18] **Lv X, Zhao K, Xue R, Liu Y, Xu J, Ma B.** May 2019. Strengthening Insights in
541 Microbial Ecological Networks from Theory to Applications. *mSystems* 4 (3):e00124–19.
542 doi:[10.1128/mSystems.00124-19](https://doi.org/10.1128/mSystems.00124-19). Publisher: American Society for Microbiology.
- 543 [19] **Banerjee S, Schlaeppli K, van der Heijden MGA.** Sep 2018. Keystone taxa as drivers
544 of microbiome structure and functioning. *Nature Reviews Microbiology* 16 (9):567–576.
545 doi:[10.1038/s41579-018-0024-1](https://doi.org/10.1038/s41579-018-0024-1). Number: 9 Publisher: Nature Publishing Group.
- 546 [20] **Röttgers L, Faust K.** Mar 2019. Can we predict keystones? *Nature Reviews*
547 *Microbiology* 17 (3):193–193. doi:[10.1038/s41579-018-0132-y](https://doi.org/10.1038/s41579-018-0132-y). Number: 3 Publisher:
548 Nature Publishing Group.
- 549 [21] **Faust K.** Nov 2021. Open challenges for microbial network construction and analysis.
550 *The ISME journal* 15 (11):3111–3118. doi:[10.1038/s41396-021-01027-4](https://doi.org/10.1038/s41396-021-01027-4).

- 551 [22] **Röttjers L, Vandeputte D, Raes J, Faust K.** Jul 2021. Null-model-based network
552 comparison reveals core associations. *ISME Communications* 1 (1):1–8. doi:
553 [10.1038/s43705-021-00036-w](https://doi.org/10.1038/s43705-021-00036-w). Number: 1 Publisher: Nature Publishing Group.
- 554 [23] **Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ.** 2017. Microbiome
555 Datasets Are Compositional: And This Is Not Optional. *Frontiers in Microbiology* 8.
556 <https://www.frontiersin.org/articles/10.3389/fmicb.2017.02224>. Retrieved
557 2022-10-24.
- 558 [24] **Chen L, Wan H, He Q, He S, Deng M.** Jul 2022. Statistical Methods for Microbiome
559 Compositional Data Network Inference: A Survey. *Journal of Computational Biology* 29
560 (7):704–723. doi:[10.1089/cmb.2021.0406](https://doi.org/10.1089/cmb.2021.0406). Publisher: Mary Ann Liebert, Inc., publishers.
- 561 [25] **Faust K, Raes J.** 2016. CoNet app: inference of biological association networks using
562 Cytoscape. *F1000Research* 5:1519. doi:[10.12688/f1000research.9050.2](https://doi.org/10.12688/f1000research.9050.2).
- 563 [26] **Friedman J, Alm EJ.** Sep 2012. Inferring Correlation Networks from Genomic Survey
564 Data. *PLOS Computational Biology* 8 (9):e1002687. doi:[10.1371/journal.pcbi.1002687](https://doi.org/10.1371/journal.pcbi.1002687).
565 Publisher: Public Library of Science.
- 566 [27] **Kurtz ZD, Müller CL, von Mering C, Bonneau RA, Blaser MJ, Littman DR, Miraldi**
567 **ER.** May 2015. Sparse and Compositionally Robust Inference of Microbial Ecological
568 Networks. *PLOS Computational Biology* 11. doi:[10.1371/journal.pcbi.1004226](https://doi.org/10.1371/journal.pcbi.1004226).
- 569 [28] **Tavakoli S, Yooseph S.** Jul 2019. Learning a mixture of microbial networks using
570 minorization–maximization. *Bioinformatics* 35 (14):i23–i30. doi:
571 [10.1093/bioinformatics/btz370](https://doi.org/10.1093/bioinformatics/btz370).
- 572 [29] **Hirano H, Takemoto K.** Jun 2019. Difficulty in inferring microbial community
573 structure based on co-occurrence network approaches. *BMC Bioinformatics* 20 (1):329.
574 doi:[10.1186/s12859-019-2915-1](https://doi.org/10.1186/s12859-019-2915-1).
- 575 [30] **Weiss S, Van Treuren W, Lozupone C, Faust K, Friedman J, Deng Y, Xia LC, Xu ZZ,**
576 **Ursell L, Alm EJ, Birmingham A, Cram JA, Fuhrman JA, Raes J, Sun F, Zhou J,**
577 **Knight R.** Jul 2016. Correlation detection strategies in microbial data sets vary widely
578 in sensitivity and precision. *The ISME Journal* 10 (7):1669–1681. doi:
579 [10.1038/ismej.2015.235](https://doi.org/10.1038/ismej.2015.235). Number: 7 Publisher: Nature Publishing Group.

- 580 [31] **Zheng J, Wittouck S, Salvetti E, Franz CM, Harris HM, Mattarelli P, O'toole PW,**
581 **Pot B, Vandamme P, Walter J, et al.** 2020. A taxonomic note on the genus
582 *Lactobacillus*: Description of 23 novel genera, emended description of the genus
583 *Lactobacillus* Beijerinck 1901, and union of *Lactobacillaceae* and *Leuconostocaceae*.
584 *International journal of systematic and evolutionary microbiology* 70 (4):2782–2858.
- 585 [32] **Doane MP, Reed MB, McKerral J, Farias Oliveira Lima L, Morris M, Goodman AZ,**
586 **Johri S, Papudeshi B, Dillon T, Turnlund AC, Peterson M, Mora M, de la**
587 **Parra Venegas R, Pillans R, Rohner CA, Pierce SJ, Legaspi CG, Araujo G,**
588 **Ramirez-Macias D, Edwards RA, Dinsdale EA.** Aug 2023. Emergent community
589 architecture despite distinct diversity in the global whale shark (*Rhincodon typus*)
590 epidermal microbiome. *Scientific Reports* 13 (1):12747. doi:[10.1038/s41598-023-39184-5](https://doi.org/10.1038/s41598-023-39184-5).
591 Number: 1 Publisher: Nature Publishing Group.
- 592 [33] **Cobo-López S, Gupta VK, Sung J, Guimerà R, Sales-Pardo M.** Jul 2022. Stochastic
593 block models reveal a robust nested pattern in healthy human gut microbiomes.
594 *PNAS Nexus* 1 (3):pgac055. doi:[10.1093/pnasnexus/pgac055](https://doi.org/10.1093/pnasnexus/pgac055).
- 595 [34] **Hall CV, Lord A, Betzel R, Zakrzewski M, Simms LA, Zalesky A, Radford-Smith**
596 **G, Cocchi L.** Nov 2019. Co-existence of Network Architectures Supporting the Human
597 Gut Microbiome. *iScience* 22:380–391. doi:[10.1016/j.isci.2019.11.032](https://doi.org/10.1016/j.isci.2019.11.032).
- 598 [35] **Ibrahim M, Raajaraam L, Raman K.** Jul 2021. Modelling microbial communities:
599 Harnessing consortia for biotechnological applications. *Computational and Structural*
600 *Biotechnology Journal* 19:3892–3907. doi:[10.1016/j.csbj.2021.06.048](https://doi.org/10.1016/j.csbj.2021.06.048).
- 601 [36] **Michael G G.** Apr 2015. Lactic metabolism revisited: metabolism of lactic acid
602 bacteria in food fermentations and food spoilage. *Current Opinion in Food Science*
603 2:106–117. doi:[10.1016/j.cofs.2015.03.001](https://doi.org/10.1016/j.cofs.2015.03.001).
- 604 [37] **Pothakos V, De Vuyst L, Zhang SJ, De Bruyn F, Verce M, Torres J, Callanan M,**
605 **Moccand C, Weckx S.** Nov 2020. Temporal shotgun metagenomics of an Ecuadorian
606 coffee fermentation process highlights the predominance of lactic acid bacteria.
607 *Current Research in Biotechnology* 2:1–15. doi:[10.1016/j.crbiot.2020.02.001](https://doi.org/10.1016/j.crbiot.2020.02.001).

- 608 [38] **Raghuvanshi R, Grayson AG, Schena I, Amanze O, Suwintono K, Quinn RA.** Aug
609 2019. Microbial Transformations of Organically Fermented Foods. *Metabolites* 9 (8):165.
610 doi:[10.3390/metabo9080165](https://doi.org/10.3390/metabo9080165). Number: 8 Publisher: Multidisciplinary Digital
611 Publishing Institute.
- 612 [39] **Yasir M, Al-Zahrani IA, Bibi F, Abd El Ghany M, Azhar EI.** Jul 2022. New insights
613 of bacterial communities in fermented vegetables from shotgun metagenomics and
614 identification of antibiotic resistance genes and probiotic bacteria. *Food Research*
615 *International* 157:111190. doi:[10.1016/j.foodres.2022.111190](https://doi.org/10.1016/j.foodres.2022.111190).
- 616 [40] **Huttenhower C, Finn RD, McHardy AC.** Oct 2023. Challenges and opportunities in
617 sharing microbiome data and analyses. *Nature Microbiology* p 1–11. doi:
618 [10.1038/s41564-023-01484-x](https://doi.org/10.1038/s41564-023-01484-x). Publisher: Nature Publishing Group.
- 619 [41] **Poirier S, Rué O, Peguilhan R, Coeuret G, Zagorec M, Champomier-Vergès MC,**
620 **Loux V, Chaillou S.** Sep 2018. Deciphering intra-species bacterial diversity of meat
621 and seafood spoilage microbiota using gyrB amplicon sequencing: A comparative
622 analysis with 16S rDNA V3-V4 amplicon sequencing. *PLoS ONE* 13 (9):e0204629. doi:
623 [10.1371/journal.pone.0204629](https://doi.org/10.1371/journal.pone.0204629).
- 624 [42] **Poirier S, Rué O, Coeuret G, Champomier-Vergès MC, Loux V, Chaillou S.** Nov
625 2018. Detection of an amplification bias associated to Leuconostocaceae family with a
626 universal primer routinely used for monitoring microbial community structures
627 within food products. *BMC Research Notes* 11 (1):802. doi:[10.1186/s13104-018-3908-2](https://doi.org/10.1186/s13104-018-3908-2).
- 628 [43] **Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP.** Jul 2016.
629 DADA2: High-resolution sample inference from Illumina amplicon data. *Nature*
630 *Methods* 13 (7):581–583. doi:[10.1038/nmeth.3869](https://doi.org/10.1038/nmeth.3869).
- 631 [44] **Callahan BJ, McMurdie PJ, Holmes SP.** Dec 2017. Exact sequence variants should
632 replace operational taxonomic units in marker-gene data analysis. *The ISME Journal* 11
633 (12):2639–2643. doi:[10.1038/ismej.2017.119](https://doi.org/10.1038/ismej.2017.119). Number: 12 Publisher: Nature Publishing
634 Group.
- 635 [45] **Capozzi V, Fragasso M, Romaniello R, Berbegal C, Russo P, Spano G.** Dec 2017.
636 Spontaneous Food Fermentations and Potential Risks for Human Health. *Fermentation*

637 3 (4):49. doi:[10.3390/fermentation3040049](https://doi.org/10.3390/fermentation3040049). Number: 4 Publisher: Multidisciplinary
638 Digital Publishing Institute.

639 [46] **Torres S, Verón H, Contreras L, Isla MI**. Jun 2020. An overview of
640 plant-autochthonous microorganisms and fermented vegetable foods. *Food Science and*
641 *Human Wellness* 9 (2):112–123. doi:[10.1016/j.fshw.2020.02.006](https://doi.org/10.1016/j.fshw.2020.02.006).

642 [47] **McMurdie PJ, Holmes S**. Apr 2013. phyloseq: An R Package for Reproducible
643 Interactive Analysis and Graphics of Microbiome Census Data. *PLOS ONE* 8
644 (4):e61217. doi:[10.1371/journal.pone.0061217](https://doi.org/10.1371/journal.pone.0061217). Publisher: Public Library of Science.

645 [48] **Lovell D, Pawlowsky-Glahn V, Egozcue JJ, Marguerat S, Bähler J**. Mar 2015.
646 Proportionality: A Valid Alternative to Correlation for Relative Data. *PLOS*
647 *Computational Biology* 11 (3):e1004075. doi:[10.1371/journal.pcbi.1004075](https://doi.org/10.1371/journal.pcbi.1004075). Publisher:
648 Public Library of Science.

649 [49] **Quinn TP, Erb I, Gloor G, Notredame C, Richardson MF, Crowley TM**. Sep 2019. A
650 field guide for the compositional analysis of any-omics data. *GigaScience* 8 (9):giz107.
651 doi:[10.1093/gigascience/giz107](https://doi.org/10.1093/gigascience/giz107).

652 [50] **Aubert J, Barbillon P, Donnet S, Miele V**. 2022. Using Latent Block Models to Detect
653 Structure in Ecological Networks. *Statistical Approaches for Hidden Variables in Ecology* p
654 117–134.

655 [51] **Gu Z**. 2022. Complex heatmap visualization. *iMeta* 1 (3):e43. doi:[10.1002/imt2.43](https://doi.org/10.1002/imt2.43).
656 _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/imt2.43>.

657 [52] **Wickham H**. 2011. ggplot2. *WIREs Computational Statistics* 3 (2):180–185. doi:
658 [10.1002/wics.147](https://doi.org/10.1002/wics.147). _eprint:
659 <https://onlinelibrary.wiley.com/doi/pdf/10.1002/wics.147>.

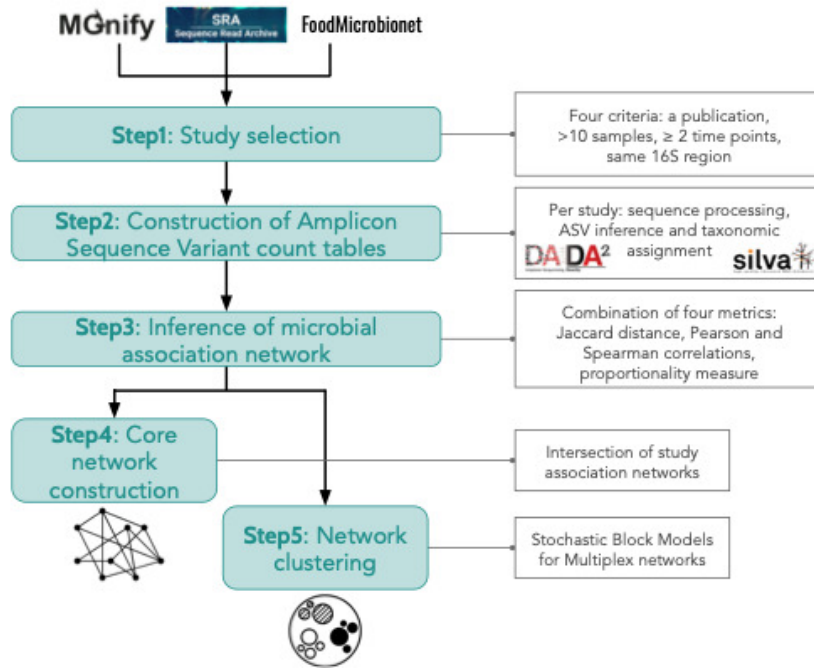
1 Tables

2 **Table 1: Publicly available metataxonomic studies on fermented vegetables used in the present work.** The table describes the main features of the
3 10 public metataxonomic studies retained for this analysis. The “study identification” column indicates the source that allowed identification of the dataset.
4 The “associated publication” column indicates the first author, title, and DOI of the article that was used to complete the dataset metadata. The “SRA identifier”
5 column indicates the Bioproject SRA study identifier that was used to download the corresponding public datasets. The “data origin and type” column indicates
6 the vegetable(s) of origin (with possible indication of additional inoculation, if applicable) of the dataset and the sequenced 16S region. The last four columns
7 indicate, respectively, the number of samples in the study, the first and last day of monitoring, and the total number of time points assessed.

Study identification	Associated publication	SRA identifier	Data origin and type	# samples	First day (time moni.)	Last day (time moni.)	# time points
MGNify (MGYS00003504)	Wuyts et al. "Carrot juice fermentations as man-made microbial ecosystems dominated by lactic acid bacteria." https://doi.org/10.1128/AEM.00134-18	PRJEB15657 ERP017487	Carrot / V4	310	1	60	12
SRA	Stoll et al. "Influence of salt concentration and iodized table salt on the microbiota of fermented cucumbers." https://doi.org/10.1016/j.fm.2020.103552	PRJNA595462 SRP237464	Cucumber / V4	60	0	64	5
SRA	Zhang et al. "Dynamics of physicochemical factors and microbial communities during ripening fermentation of Pixian Doubanjiang, a typical condiment in Chinese cuisine." https://doi.org/10.1016/j.fm.2019.103342	PRJNA544161 SRP19928	Doubanjiang (red pepper, meju) / V3-V4	28	30	720	3
SRA	Jung et al. "Role of combined lactic acid bacteria in bacterial, viral, and metabolite dynamics during fermentation of vegetable food, kimchi." https://doi.org/10.1016/j.foodres.2022.111261	PRJNA751723 SRP330976	Kimchi (kimchi cabbage, garlic, ginger, red pepper) + inoculation / V3-V4	192	0	30	6
SRA	Song et al. "Microbial niches in raw ingredients determine microbial community assembly during kimchi fermentation." https://doi.org/10.1016/j.foodchem.2020.126481	PRJNA564474 SRP221859	Kimchi (kimchi cabbage, garlic, ginger, and red pepper) / V3-V4	52	1	50	8
SRA	Wang et al. "Effects of salt concentration on the quality of paocai, a fermented vegetable product from China." https://doi.org/10.1002/jsfa.11271	PRJNA689239 SRP300069	Paocai (cabbage) / V3-V4	18	1	30	6
SRA	Li et al. "Metagenomic insights into the changes in microbial community and antimicrobial resistance genes associated with different salt content of red pepper (<i>Capsicum annuum</i> L.) sauce." https://doi.org/10.1016/j.fm.2019.103295	PRJNA473189 SRP149752	Red pepper / V4	80	3	109	10
FoodMicrobioNet	Pérez-Díaz et al. Modulation of the bacterial population in commercial cucumber fermentations by brining salt type. https://doi.org/10.1111/iam.14597	PRJNA485506 SRP157093	Cucumber / V3-V4	72	1	14	7
SRA	He et al. "Fermentation characteristics and bacterial dynamics during Chinese sauerkraut fermentation by <i>Lactobacillus curvatus</i> LC-20 under varied salt concentrations reveal its potential in low-salt suan cai production." https://doi.org/10.1016/j.jbiosc.2021.03.009	PRJNA662831 SRP282266	Chinese Sauerkraut + inoculation / V4	45	3	15	2
SRA	Zabat et al. "The impact of vegan production on the kimchi microbiome." https://doi.org/10.1016/j.fm.2018.04.001	PRJNA418790 SRP125207	Kimchi (ginger, scallion, onion, garlic, red pepper, sugar, radish, cabbage, miso paste, fish sauce) / V4	74	0	3	4

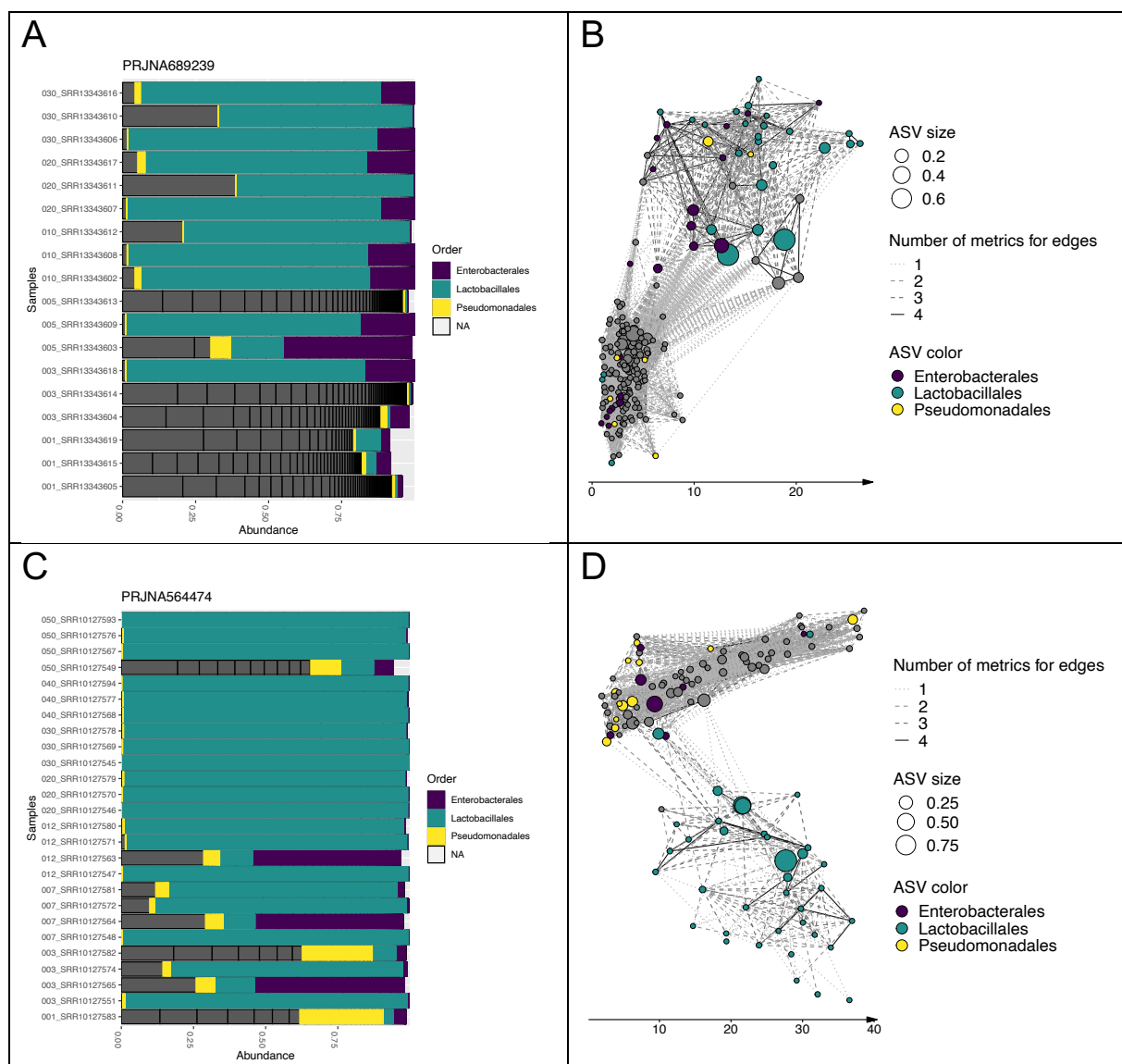
9 Figures

10
11



12
13
14
15
16

Figure 1: Meta-analysis approach for integrating amplicon datasets into microbial association networks to compare microbial communities of fermented vegetables.

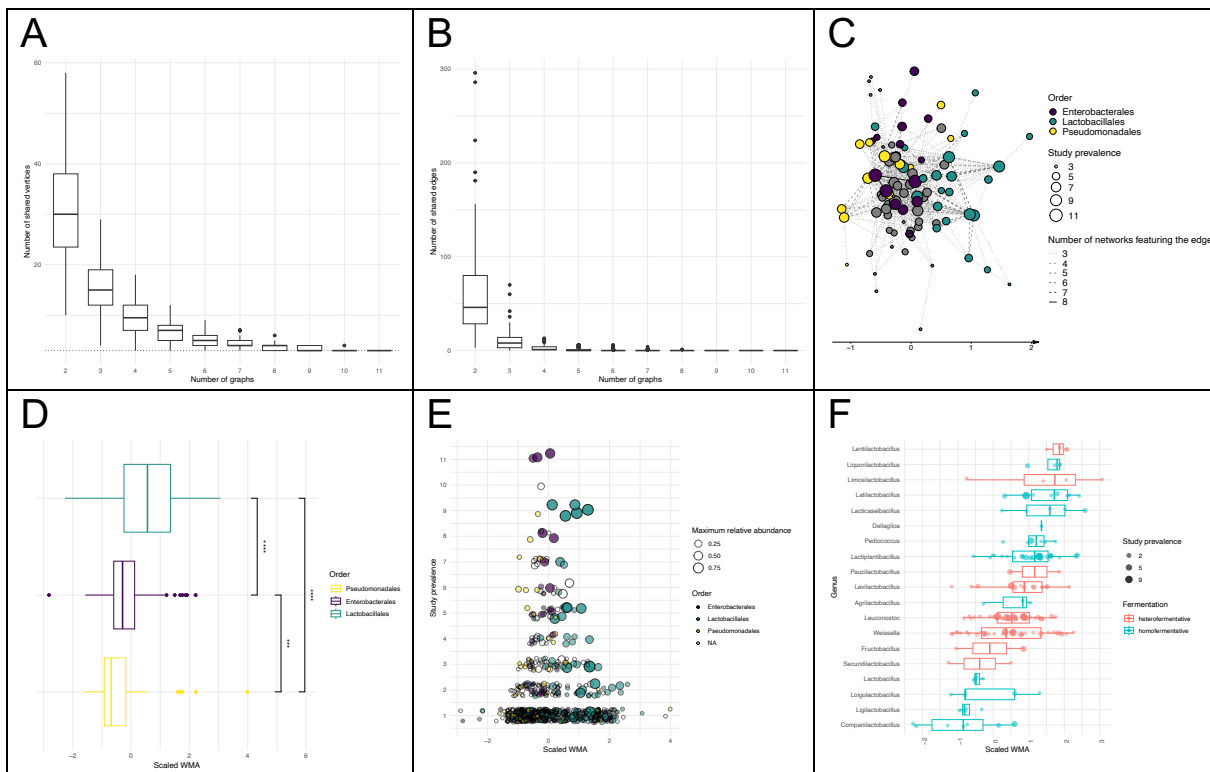


17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34

Figure 2: Microbial association networks for studies PRJNA689239 and PRJNA564474 highlight the dynamic evolution of microbial communities during fermentation.

(A) and (C): Barplots depicting relative abundances in each sample for studies PRJNA689239 and PRJNA564474, respectively. Samples are ordered by age. Gray color indicates a taxonomic order other than *Enterobacterales*, *Lactobacillales*, and *Pseudomonadales*.

(B) and (D): ASV association networks for studies PRJNA689239 and PRJNA564474, respectively. Each node represents an ASV; node size reflects its maximum relative abundance and color represents its taxonomic order. The x-axis corresponds to the weighted mean age (WMA) of the samples in which the ASV was detected, measured in days, and weighted by ASV relative abundance. An edge between two nodes indicates an association that was detected according to at least one metric.



35

36

Figure 3: Core network and succession of bacterial communities

37

(A) Boxplot showing the number of vertices in the core networks built from the intersection of 2 to 11 networks. The dotted gray line corresponds to three vertices.

38

(B) Boxplot showing the number of edges in the core networks built from the intersection of 2 to 11 networks

39

(C) Core network built from ASV associations found in at least three networks. The line type of an edge represents the number of times the ASV association was found. The node position on the x-axis is the mean scaled WMA. ASVs are colored by taxonomic order.

40

(D) Boxplot showing the differences in mean scaled WMA between ASVs affiliated with the orders *Pseudomonadales*, *Enterobacterales*, and *Lactobacillales*.

41

(E) Scatterplot of ASVs colored by taxonomic order, depicting their prevalence in relation to mean scaled WMA.

42

(F) Boxplot showing the differences in mean scaled WMA among genera in family *Lactobacillaceae*. Each dot is an ASV.

43

44

45

46

47

48

49

50

51

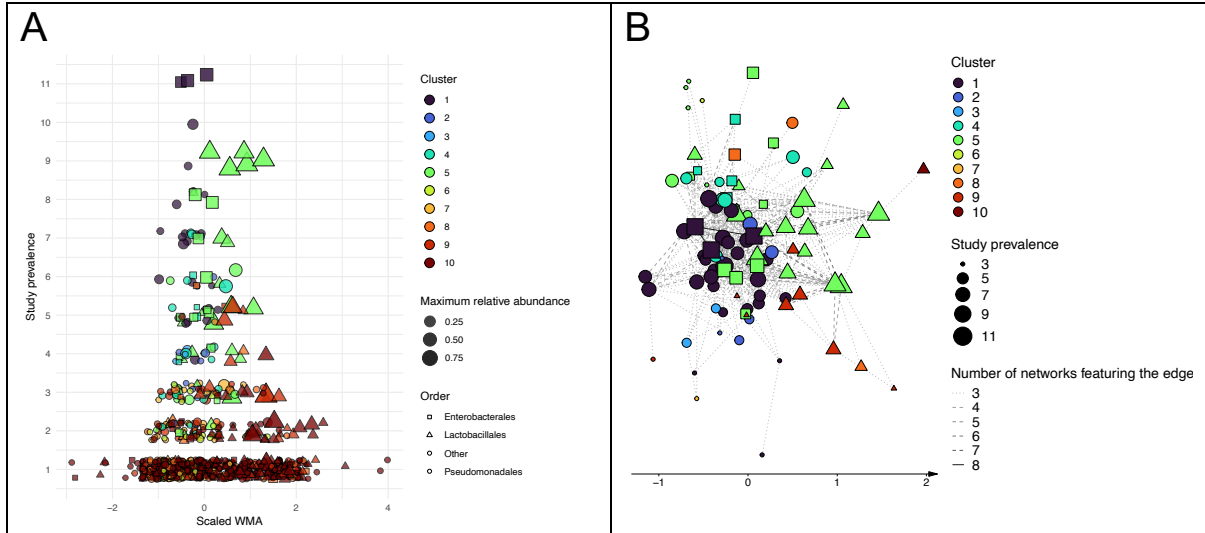
52

53

54

55

56



57

58 **Figure 4: Network clustering shows an association between highly prevalent ASVs from**
59 **orders *Lactobacillales* and *Enterobacterales*.**

60 (A) Scatterplot of ASVs colored by cluster. The shape corresponds to the taxonomic order.

61 (B) Core network built from ASV associations found in at least three networks, with ASVs colored by
62 cluster. The shape corresponds to the taxonomic order.

63

64