

Title	Strain-level metagenomic analysis of the fermented dairy beverage nunu highlights potential food safety risks
Authors	Walsh, Aaron M.;Crispie, Fiona;Daari, Kareem;O'Sullivan, Orla;Martin, Jennifer C.;Arthur, Cornelius T.;Claesson, Marcus J.;Scott, Karen P.;Cotter, Paul D.
Publication date	2017-06-16
Original Citation	Walsh, A. M., Crispie, F., Daari, K., O'Sullivan, O., Martin, J. C., Arthur, C. T., Claesson, M. J., Scott, K. P. and Cotter, P. D. [2017] 'Strain-level metagenomic analysis of the fermented dairy beverage nunu highlights potential food safety risks', Applied and Environmental Microbiology. In Press, doi: 10.1128/aem.01144-17
Type of publication	Article (peer-reviewed)
Link to publisher's version	10.1128/aem.01144-17
Rights	Open access - http://creativecommons.org/licenses/by/4.0/ © 2017 American Society for Microbiology
Download date	2025-04-17 18:37:24
Item downloaded from	https://hdl.handle.net/10468/4188

1 Strain-level metagenomic analysis of the fermented dairy beverage nunu highlights potential
2 food safety risks

3 Aaron M. Walsh^{a,b,c}, Fiona Crispie^{a,b}, Kareem Daari^d, Orla O'Sullivan^{a,b}, Jennifer C. Martin^d,
4 Cornelius T. Arthur^e, Marcus J. Claesson^{b,c}, Karen P. Scott^d, Paul D. Cotter^{a,b,*}.

5

6 ^aTeagasc Food Research Centre, Moorepark, Fermoy, Co. Cork, Ireland

7 ^bAPC Microbiome Institute, University College Cork, Co. Cork, Ireland

8 ^cMicrobiology Department, University College Cork, Co. Cork, Ireland

9 ^dRowett Institute, University of Aberdeen, Aberdeen, Scotland, UK. AB25 2ZD

10 ^eAnimal Research Institute, Accra, Ghana

11 * Corresponding author.

12

13 Address correspondence to:

14 Paul D. Cotter,

15 Principal Research Officer,

16 Teagasc Food Research Centre,

17 Moorepark, Fermoy, Co. Cork,

18 Ireland.

19 Email: paul.cotter@teagasc.ie

20 Telephone: +353 (0)25 42694

21 **Abstract**

22 The rapid detection of pathogenic strains in food products is essential for the prevention of
23 disease outbreaks. It has already been demonstrated that whole metagenome shotgun
24 sequencing can be used to detect pathogens in food but, until recently, strain-level detection
25 of pathogens has relied on whole metagenome assembly, which is a computationally
26 demanding process. Here, we demonstrate that three short read alignment-based methods,
27 MetaMLST, PanPhlAn, and StrainPhlAn, can accurately, and rapidly, identify pathogenic
28 strains in spinach metagenomes which were intentionally spiked with Shiga toxin-producing
29 *Escherichia coli* in a previous study. Subsequently, we employ the methods, in combination
30 with other metagenomics approaches, to assess the safety of nunu, a traditional Ghanaian
31 fermented milk product which is produced by the spontaneous fermentation of raw cow milk.
32 We show that nunu samples are frequently contaminated with bacteria associated with the
33 bovine gut, and worryingly, we detect putatively pathogenic *E. coli* and *Klebsiella*
34 *pneumoniae* strains in a subset of nunu samples. Ultimately, our work establishes that short
35 read alignment-based bioinformatics approaches are suitable food safety tools, and we
36 describe a real-life example of their utilisation.

37

38 **Importance**

39 Foodborne pathogens are responsible for millions of illnesses, annually. Here, we
40 demonstrate that short read alignment-based bioinformatics tools can accurately, and rapidly,
41 detect pathogenic strains in food products from shotgun metagenomics data. The methods
42 used here are considerably faster than both traditional culturing methods and alternative
43 bioinformatics approaches that rely on metagenome assembly, and thus they can potentially
44 be used for more high-throughput food safety testing. Overall, our results suggest that whole
45 metagenome sequencing can be used as a practical food safety tool to prevent diseases or link
46 outbreaks to specific food products.

47

48 **Introduction**

49 In recent years, high-throughput sequencing (HTS) has become an important tool in food
50 microbiology (1). HTS enables in-depth characterisation of food-related microbial isolates,

51 *via* whole genome sequencing (WGS), and it facilitates culture-independent analysis of
52 mixed microbial communities in foods, *via* metagenomic sequencing.

53 WGS has provided invaluable insights into the genetics of starter cultures (2, 3), and it is
54 routinely used in epidemiology to identify outbreak-associated foodborne pathogens isolated
55 from clinical samples, by comparing the single nucleotide polymorphism (SNP) profiles of
56 outbreak strain genomes versus non-outbreak strain genomes (4-6). Metagenomic sequencing
57 enables the elucidation of the roles of microorganisms during food production (7-9), and it
58 can be used to track microorganisms of interest through the food production chain, as
59 illustrated by Yang *et al.* (10), who used whole metagenome shotgun sequencing to track
60 pathogenic species in the beef production chain. Indeed, metagenomic sequencing can be
61 used to detect pathogens in foods to monitor outbreaks of foodborne illnesses (11), but few
62 studies have done so, because of the limited taxonomic resolution achievable using these
63 methods. Typically, 16S rRNA gene sequencing provides genus-level taxonomic resolution
64 (12), and although sub-genus-level classification is achievable using species-classifiers (13)
65 or oligotyping (14, 15), these methods cannot accurately discriminate between strains.
66 Similarly, metagenome sequence classification tools usually provide species-level resolution
67 (16). However, strain-level resolution is necessary for the accurate identification of pathogens
68 in food products (17). Leonard *et al.* successfully achieved strain-level resolution of Shiga
69 toxin producing *Escherichia coli* strains in spinach samples using metagenome shotgun
70 sequencing (18). However, the bioinformatics methods used in that study were based on
71 metagenome assembly, which is a computationally demanding process (19, 20), and thus
72 alternative strain-level identification methods are needed.

73 Since 2016, several short read alignment based software applications, including MetaMLST
74 (20), StrainPhlAn (21), and PanPhlAn (19), have been released that can achieve strain-level
75 characterisation of microorganisms from metagenome shotgun sequencing data. All three
76 applications are considerably faster than metagenome assembly based methods. To date,
77 these programs have not been employed to detect pathogens in food products, but there is
78 strong evidence to suggest that they have considerable potential for this purpose: MetaMLST
79 accurately predicted that the strain responsible for the 2011 German *E. coli* outbreak
80 belonged to *E. coli* ST678 (20), and similarly, PanPhlAn accurately predicted that the strain
81 was a Shiga toxin producer (19), based on the analysis of the gut metagenomes of infected
82 patients (22). StrainPhlAn has so far not been used for epidemiological purposes, but a recent

83 study demonstrated that it can be used to predict the phylogenetic relatedness of bacterial
84 strains from different samples (21).

85 MetaMLST aligns sequencing reads against a housekeeping gene database to identify
86 sequence types present in metagenomic samples based on multilocus sequence typing
87 (MLST). The MetaMLST database contains all currently known sequence types, but it can be
88 updated as required to include newly identified sequence types. MetaMLST does not require
89 any prior knowledge of the microbial composition of sample and it can simultaneously detect
90 different species' sequence types. PanPhlAn aligns sequencing reads against a species
91 pangenome database, constructed from reference genomes, to functionally characterise
92 strains present in metagenomic samples. PanPhlAn allows the user to generate customisable
93 pangenome databases for any species. StrainPhlAn extracts species specific marker genes
94 from sequencing reads and it aligns the markers against reference genomes to identify the
95 strains present in metagenomic samples. StrainPhlAn requires output from MetaPhlAn2, and
96 both programs use the same database.

97 In this study, we describe the characterisation of nunu, a traditional Ghanaian fermented milk
98 product (FMP), at the genus, species, and strain-levels, using a combination of 16S rRNA
99 gene sequencing and whole metagenome shotgun sequencing. Nunu is produced by the
100 spontaneous fermentation of raw cow milk in calabashes or plastic or metal containers under
101 ambient conditions, and it is usually consumed after 24-36 hours (23). At present, little is
102 known about nunu's microbiology, relative to other FMPs, like kefir or yoghurt (24).
103 Previously, a number of potentially pathogenic bacteria, including *Enterobacter*, *Escherichia*
104 and *Klebsiella*, were detected in nunu by culture based methods (25). Here, we carry out the
105 first culture-independent analysis of a number of nunu samples. In addition to detecting the
106 presence of a variety of lactic acid bacteria (LAB) typical of fermented dairy products,
107 MetaMLST, PanPhlAn and StrainPhlAn all indicated the presence of pathogenic *E. coli* and
108 *Klebsiella pneumoniae* in a subset of the samples. We also demonstrate that these tools can
109 accurately predict the presence of pathogenic strains in foods by testing them on food
110 metagenomes which were spiked with Shiga toxin producing *E. coli*. Ultimately, our work
111 establishes that short read alignment based methods can be used for the detection of
112 pathogens in foods.

113

114 **Results**

115 **16S rRNA gene sequencing of nunu samples**

116 Nunu samples were collected from producers with hygiene practice training (n=5) and
117 producers without hygiene practice training (n=5), respectively. 16S rRNA gene sequencing
118 analysis revealed that there were no significant differences in the alpha-diversity of nunu
119 samples from trained or untrained producers (Figure S1a), although there was a clear
120 separation in the beta-diversity of the two groups (Figure S1b).

121 The 16S rRNA data was also analysed to determine bacterial composition (Figure 1a). At the
122 family level, all of the samples were dominated by Lactobacillales, and at the genus-level,
123 most samples were dominated by *Streptococcus*, although the sample 1t2am was dominated
124 by *Lactococcus*. *Enterococcus* was detected in 4/10 samples (1 trained and 3 untrained) at
125 $\geq 3\%$ relative abundance, and it was highest in the sample 2u6am, where it was present at
126 19% relative abundance. In addition, *Staphylococcus* was detected in all 10 samples, although
127 its abundance was $\leq 1\%$ in each case. The detection of staphylococci was consistent with a
128 corresponding culture-dependent analysis of the samples (supplemental material).

129 Importantly, Enterobacteriales were also prevalent. *Enterobacter* was detected in 9/10
130 samples (4 samples from trained producers and 5 from untrained producers) at $\geq 1\%$ relative
131 abundance, and it was highest in the sample 2u8am, where it was present at 23% relative
132 abundance. *Escherichia-Shigella* was detected in 8/10 samples (4 trained and 4 untrained) at
133 $\geq 1\%$ relative abundance, and it was highest in the sample 1t7am, where it was present at 17%
134 relative abundance; this finding was again consistent with culture-dependent analysis of the
135 samples (supplemental material).

136 The Kruskal-Wallis test indicated that there were significant differences in the relative
137 abundances of *Macroccoccus* (p=0.01), which was higher in samples from trained producers,
138 and *Streptococcus* (p=0.02), which was higher in samples from untrained producers (Figure
139 1b). No other genera were significantly different.

140

141 **Species-level compositional analysis of nunu samples as revealed by shotgun sequencing**

142 MetaPhlan2-based analysis of shotgun metagenomic data provided results that were
143 generally consistent with those derived from amplicon sequencing. 11 species accounted for
144 $>90\%$ of the microbial composition of every sample (Figure 2). At the species-level, most
145 samples were dominated by *Streptococcus infantarius*, although sample 1t2am was

146 dominated by *Lactococcus lactis*. *Enterococcus faecium* was detected in 4/10 samples (2
147 trained and 2 untrained) at $\geq 1\%$ relative abundance, and it was highest in the sample 1t2am,
148 where it was present at 22% relative abundance. High abundances of Enterobacteriales were
149 again apparent. *Enterobacter cloacae* were detected in the sample 1t8am, where it was
150 present at 1% relative abundance. *Escherichia coli* was detected in 2/10 samples (2 trained) at
151 $\geq 7\%$ relative abundance, and it was highest in 1t7am, where it was present at 13% relative
152 abundance. *Klebsiella pneumoniae* was detected in 7/10 samples (4 trained and 3 untrained)
153 at $\geq 3\%$ relative abundance, and it was highest in 1t8am, where it was present at 71% relative
154 abundance. In contrast, *Klebsiella* was not detected by amplicon sequencing, and this
155 discrepancy might be due to similarities in the 16S rRNA genes from these genera(42).

156 The Kruskal-Wallis test indicated that there were significant differences in the relative
157 abundances of *Macrococcus caseolyticus* ($p=0.01$), which was higher in samples from trained
158 producers, and *Streptococcus infantarius* ($p=0.01$), which was higher in samples from
159 untrained producers (Figure S2). No other species were significantly different.

160

161 **Investigation of the functional potential of the nunu microbiota**

162 SUPER-FOCUS was used to provide an overview of the functional potential of the nunu
163 metagenome. As expected, a significant proportion of the metagenome was assigned to
164 housekeeping functions like carbohydrate metabolism, nucleic acid metabolism, and protein
165 metabolism (Figure 3). However, SUPER-FOCUS also detected high levels of functions
166 associated with horizontal gene transfer and virulence in nunu. The level 1 subsystem
167 “Phages, Prophages, Transposable elements” was present at $\geq 1\%$ average relative abundance
168 in both groups, although it was significantly higher in nunu samples from trained producers
169 ($p=0.047$). Similarly, the level 1 subsystem “Virulence” was present at $\geq 3.5\%$ average
170 relative abundance in both groups.

171 HUMAnN2 was used to provide more comprehensive insights into the functional potential of
172 the nunu metagenome. Unsurprisingly, the 25 most abundant genetic pathways were
173 associated with carbohydrate metabolism, nucleic acid metabolism, and protein metabolism
174 (Figure 4a). MDS analysis of all the normalised HUMAnN2 pathway abundances suggested
175 that there were differences in the overall functional potential of the groups (Figure S3), and
176 we detected significant differences in the relative abundances of some individual pathways
177 (Table S1). Notably, we observed that histidine degradation pathways were higher in trained

178 samples ($p=0.047$) (Figure 4c). Furthermore, histidine decarboxylase genes were only
179 detected in trained samples. Several other undesirable genetic pathways were detected in both
180 groups. For example, putrescine biosynthesis pathways and polymyxin resistance genes co-
181 occurred in 7/10 samples (Figure 4c), and these pathways were all attributed to *E. cloacae*, *E.*
182 *coli*, *K. pneumoniae*, or a combination of these three species. We detected several other
183 antibiotic resistance genes, including beta-lactamase genes and methicillin resistance genes,
184 in both groups (Figure S4). In addition, we found HGT-associated genes, including plasmid
185 maintenance genes and transposition genes, in both groups.

186

187 **Application of strain-level analysis to characterise enteric bacteria in nunu**

188 Leonard *et al.* previously used metagenomic sequencing to detect *E. coli* in spinach which
189 was intentionally spiked with *E. coli* O157:H7 strain Sakai (11). We downloaded the
190 metagenomic reads from that study (16 samples) and we subjected them to StrainPhlAn,
191 MetaMLST and PanPhlAn analysis, to confirm that these tools can accurately detect
192 pathogens in food samples: MetaMLST was used for multi-locus sequence typing,
193 StrainPhlAn was used for phylogenetic identification, and PanPhlAn was used for functional
194 characterisation. MetaMLST accurately detected *E. coli* ST11 in 7/16 spinach samples (Table
195 1). StrainPhlAn detected *E. coli* strains in 5/16 samples and it showed that the *E. coli* strain in
196 each of these samples was closely related to *E. coli* O157:H7 strain Sakai (Figure 5).
197 PanPhlAn detected Shiga toxin genes in 15/16 samples (Table 1) and it indicated that the *E.*
198 *coli* strain in each of these samples was most closely related to *E. coli* O157:H7 strain Sakai.
199 Thus, overall, PanPhlAn was the most sensitive method in this instance, since it was able to
200 detect STEC in almost all of the samples, whereas the other tools detected STEC in less than
201 half of the samples. In a follow-on study, Leonard *et al.* spiked spinach with 12 different
202 Shiga toxin producing *E. coli* strains, and they detected single strains in 17 samples (18). We
203 downloaded the metagenomic reads from the 17 samples and ran PanPhlAn, and were able to
204 identify Shiga toxin genes in all 17 samples (Table S2).

205 Having established the relative merits of these tools, we subsequently employed all three
206 strategies to identify the strains of *E. coli* and *K. pneumoniae* present in the nunu samples.
207 With regard to *E. coli*, MetaMLST detected a novel *E. coli* sequence type in 1t7am (Table 2).
208 StrainPhlAn detected 24 *E. coli* marker genes in the samples and a phylogenetic tree (Figure
209 6a), which was generated by aligning these markers against 118 *E. coli* reference genomes

210 (listed in Table S3), revealed that the *E. coli* strain in one sample, 1t7am, was closely related
211 to *E. coli* O139:H28 E24377A. PanPhlAn detected *E. coli* strains in two samples: 1t7am and
212 1t8am. MDS analysis indicated that the strains from the two samples were functionally
213 distinct from one another. Notably, a ShET2 enterotoxin encoding gene was identified in the
214 *E. coli* strain from 1t7am. The same gene was found in *E. coli* O139:H28 E24377A. With
215 regard to *K. pneumoniae*, MetaMLST detected the known sequence type *K. pneumoniae*
216 ST39 in the sample 2u3am. Apparently novel *K. pneumoniae* sequence types were identified
217 in six other samples (Table 1). StrainPhlAn detected 38 *K. pneumoniae* marker genes in the
218 samples and a phylogenetic tree (Figure 6b), which was constructed by aligning these
219 markers against 40 *K. pneumoniae* reference genomes (listed in Table S4), revealed that the
220 *K. pneumoniae* strains in two samples, 1t8am and 2u3am, were closely related to *K.*
221 *pneumoniae* KpQ3. In contrast, the *K. pneumoniae* strain in 1t7am was most closely related to
222 *K. pneumoniae* UCICRE 7. MDS analysis of the PanPhlAn output showed that five of the
223 detected *K. pneumoniae* strains were functionally similar to one another (Figure 6c).
224 However, two of the detected *K. pneumoniae* strains, in samples 1t6am and 1t7am, appeared
225 to be functionally distinct from the others. In addition, PanPhlAn indicated that sample 1t6am
226 might have contained multiple strains, since an unusually high number of 5746 *K. pneumoniae*
227 gene families were detected. A TEM beta-lactamase gene was found in 1t2am using
228 PanPhlAn and, furthermore, an OXA-48 carbapenemase gene was detected in 2u8am and the
229 same gene was found in *K. pneumoniae* KpQ3.

230 Finally, we compared the time taken to process 10 nunu metagenome samples using the
231 short-read alignment tools versus the metagenome assembler IDBA-UD (Figure S5). In each
232 case, we observed that all of the short-read alignment tools were faster than IDBA-UD. It is
233 important to note that additional bioinformatics analyses (contig binning, SNP analysis, etc.)
234 are required to achieve strain-level identification from assembled metagenomes, and this
235 emphasises the superior speed of the short-read alignment tools.

236

237 Discussion

238 Foodborne pathogens are responsible for millions of cases of disease annually, in the United
239 States alone (43). High-throughput sequencing can potentially be used to detect pathogenic
240 strains in food products to prevent the occurrence of disease outbreaks. A recent proof of
241 concept study demonstrated that whole metagenome shotgun sequencing accurately detected

242 Shiga toxin producing *E. coli* (STEC) strains in spiked spinach samples (18). However, that
243 study used whole metagenome assembly-based approaches to achieve strain-level taxonomic
244 resolution of the STEC in the samples. Whole metagenome assembly is a computationally
245 intensive, time-consuming process, as illustrated by Nurk *et al.*, who recently reported that
246 metagenome assembly can take between 1.5 hours to 6 hours, with a memory footprint
247 ranging from 7.3 GB to 234.5 GB, to process a single human gut metagenomic sample,
248 depending on the chosen assembler (44). Thus, the application of more rapid, less intensive
249 bioinformatic tools for strain detection is desirable. In this study, we demonstrate that the
250 short read alignment-based programs MetaMLST, StrainPhlAn, and PanPhlAn can accurately
251 identify pathogens in food products.

252 We validated the accuracy of each approach by processing spinach metagenome data from
253 samples that were spiked with the STEC O157:H7 Sakai in a previous study (11). We
254 observed that PanPhlAn was the most sensitive approach. Indeed, PanPhlAn was able to
255 identify STEC in every sample where it was present at >2% relative abundance, whereas the
256 other approaches worked best when STEC was present at high relative abundances. However,
257 none of the tools detected *E. coli* O157:H7 Sakai in every sample tested. The observation of
258 false negatives highlights that the tools are not entirely accurate. It is likely that increased
259 sequencing depth and/or longer sequencing read lengths would reduce the false negative rate.
260 We recommend that these tools be used to supplement data from metagenome sequence
261 classifiers like MetaPhlAn2, which did detect *E. coli* in each sample. Therefore, we
262 subsequently used the strain-level analysis tools in combination with other metagenomic
263 approaches to assess the safety of nunu, a traditional Ghanaian fermented milk product.

264 Nunu is produced through the spontaneous fermentation of raw cow milk in calabashes or
265 other containers for 24-36 hours at ambient temperature (23). The crude nature of the nunu
266 production process has raised food safety concerns (25). Indeed, several potentially
267 pathogenic microorganisms were previously detected in nunu samples by microbial culturing
268 (25). This resulted in some nunu producers receiving hygiene practice training to improve
269 food safety. However, our work suggests that there is little difference in the prevalence of
270 pathogens in nunu samples from trained and untrained producers. One reason for this may be
271 that it is difficult for the nunu producers to adhere to the training recommendations which are
272 not appropriate to the rural production conditions. During training, the producers were
273 advised to pasteurise the milk before cooling and adding a starter culture. After incubating for
274 4-6 hours in a covered container, they were advised to stir the mixture and refrigerate the

275 product. Lack of access to specific heat control and electricity, as well as the variance from
276 the traditional method, which does not use a starter culture, are both reasons why the training
277 is not adhered to.

278 16S rRNA gene sequencing revealed that the samples were dominated by Lactobacillales.
279 However, we also detected high abundances of Enterobacteriales, including *Enterobacter* and
280 *Escherichia*, in both groups. Subsequently, whole metagenome shotgun sequencing showed
281 that most samples were dominated by *Streptococcus infantarius*, a species which was
282 previously identified in other African dairy products (45, 46). Concernedly, *S. infantarius* has
283 been linked to several human diseases, including bacteraemia (47), endocarditis (48) and
284 colon cancer (49). Aside from *S. infantarius*, two other potentially pathogenic species,
285 *Escherichia coli* and *Klebsiella pneumoniae*, were identified in a subset of samples.

286 Overall, our findings indicate that nunu samples from trained producers and untrained
287 producers were contaminated with faecal material. Cattle faeces can be a major source of
288 bacterial contaminants in raw cow milk (29), and thus, our results are not entirely surprising,
289 but the remarkable abundance of such microorganisms in nunu is worrying. It had been
290 hoped that nunu could be used to supplement traditional cereal-based weaning foods to
291 improve infant nutrition. However, qualitative research among mothers and health workers
292 highlighted safety concerns, which, as we have shown here, are valid. In particular, the
293 presence of *E. coli* and *K. pneumoniae* in nunu is a concern, and, thus, we employed strain-
294 level metagenomics for the further characterisation of these bacteria.

295 In terms of *E. coli*, strain-level analysis indicated that the *E. coli* strain in one sample was an
296 enterotoxin producer and it was closely related to *E. coli* O139:H28 E24377A, a strain which
297 was linked to an outbreak of waterborne diarrhoea in India (50). In terms of *K. pneumoniae*,
298 strain-level analysis indicated that the *K. pneumoniae* strains in two samples were antibiotic
299 resistant and they were closely related to *K. pneumoniae* KpQ3, a strain which was linked to
300 nosocomial outbreaks among burn unit patients. Thus, strain-level analysis suggests that there
301 are likely pathogens in some of the samples. Interestingly, PanPhlAn also suggested that
302 there were functionally distinct strains of both species in nunu samples from different
303 producers. Perhaps, this indicates multiple incidences or sources of contamination.
304 Undoubtedly, our work highlights an urgent need to further improve hygiene practices during
305 nunu production, and the pasteurisation of the starting milk and the use of starter-based
306 fermentation systems is an obvious solution.

307 In conclusion, our work suggests that short read alignment-based strain detection tools can be
308 used to detect pathogens in other foods, apart from nut or spinach, and they might also be
309 useful for tracing the sources of foodborne disease outbreaks back to particular foods. Such
310 tools are a significant improvement over 16S rRNA gene sequencing, which is often limited
311 to genus-level identification, or metagenome read classification tools, which are limited to
312 species-level identification (16). In addition, they are faster, and less computationally
313 intensive, than metagenome assembly-based strain detection methods, making them more
314 relevant to real-life scenarios which necessitate the rapid testing of many food samples. With
315 DNA sequencing costs continuing to decrease, the approach outlined here is an affordable
316 option for food safety testing.

317

318 **Acknowledgements**

319 We would like to thank the researchers at the Animal Research Institute, in Accra, Ghana for
320 their help with sample collection, storage and processing. This research was funded by
321 Science Foundation Ireland in the form of a centre grant (APC Microbiome Institute grant
322 number SFI/12/RC/2273). Research in the Cotter laboratory is also funded by Science
323 Foundation Ireland through the PI award “Obesibiotics” (11/PI/1137). Orla O’Sullivan is
324 funded by Science Foundation Ireland through a Starting Investigator Research Grant award
325 (13/SIRG/2160). Kareem Daari is a PhD student funded by Ghana Educational Trust (GET)
326 Fund. The Rowett Institute receives funding from the Scottish Government (RESAS).

327

328 **Materials and Methods**

329 **Sampling**

330 Five nut samples were collected from producers with hygiene practice training, and another
331 five samples were collected from producers without hygiene practice training. The identity of
332 the samples from trained and untrained individuals was blinded until after sequencing
333 analysis was completed. The samples from the trained group were labelled 1t2am, 1t6am,
334 1t7am, 1t8am, and 2t2am. The samples from the untrained group labelled 1u6am, 2u2am,
335 2u3am, 2u6am, and 2u8am. All samples were collected in the morning and placed on ice for
336 transport to the lab. Sample aliquots (4ml) were then mixed with glycerol to a final
337 concentration of 20% and stored at -20°C prior to DNA extraction. DNA was extracted from

338 the samples at the Animal Research Institute, Accra, Ghana and then sent to Scotland to
339 comply with International laws on the import of animal samples (Import Licence form
340 AB117).

341

342 **Microbiological analysis**

343 Basic microbiology culture analysis was carried out in Ghana. The plate-count technique was
344 used to estimate the total viable bacterial count of the nunu samples on Milk Plate Count
345 Agar (LAB M, UK). Bacterial counts were compared for plates growing aerobically or
346 anaerobically at 30°C for 36-72 h. Anaerobic plates were incubated in airtight canisters
347 containing CO₂Gen sachets (Oxoid, UK), which created an anaerobic atmosphere. Following
348 incubation, colonies were counted using an SC6+ electronic colony counter (Stuart Scientific,
349 UK). The presence of specific pathogens in the nunu samples was determined by streaking
350 nunu directly onto selective agar plates to visually assess bacterial growth. The following
351 selective agars were used: Blood agar (Merck, Germany) for *Staphylococcus*; MacConkey
352 agar (Merck, Germany) for Enterobacteria; de Man Rogosa Sharpe agar (MRS) (Oxoid, UK)
353 for *Lactobacillus* species; and *Salmonella Shigella* agar (Oxoid, UK). Any mixed growth
354 plates were re-purified by streaking onto selected secondary agars. Lactose fermenting
355 colonies identified on MacConkey agar were sub-cultured onto Eosin Methylene Blue Agar
356 (EMBA) (Scharlau Chemie, Spain) to isolate/identify *E. coli*. Additionally, *Staphylococcus*
357 colonies from Blood Agar were sub-cultured onto Mannitol Salt Agar (MSA) (Oxoid, UK) to
358 isolate/identify *Staphylococcus aureus*. The following biochemical tests were used to confirm
359 bacterial identification: the Motility Indole Urea (MIU) test; the catalase test; the Triple
360 Sugar Iron (TSI) test; and the Indole Methyl Red Vorges-Proskeur Citrate (IMViC) tests.
361 Cellular morphology was determined by Gram staining as well as microscopic examination.

362

363 **DNA extraction and next generation sequencing**

364 Briefly, 1 ml of each thawed sample was diluted in 9 ml of sterile PBS, mixed thoroughly
365 using vortex and centrifuged for 10 min (8,000-10,000 g). The bacterial cell pellets were
366 resuspended in 432 µl sterile dH₂O and 48µl 0.5 M EDTA, mixed thoroughly by a
367 combination of vortex and with a sterile pipette tip and the suspension frozen. The frozen
368 samples were thawed on the bench and refrozen and finally thawed (giving a total of two

369 freeze/thaw cycles) before extracting the DNA using the Promega Wizard genomic DNA
370 extraction kit (Promega, Madison, WI, USA) according to the manufacturer's protocol. The
371 freeze/thaw cycles were carried out to maximise bacterial cell lysis. Following extraction, the
372 DNA pellets were air dried for about 60 minutes and stored sealed under airtight conditions
373 and transported from the Animal Research Institute, Accra, Ghana to the Rowett Institute, at
374 University of Aberdeen, for further analysis.

375 DNA extracts were quantified using the Qubit High Sensitivity DNA assay (BioSciences,
376 Dublin, Ireland). 16S rRNA gene sequencing libraries were prepared from extracted DNA
377 using the 16S Metagenomic Sequencing Library Preparation protocol from Illumina, with
378 minor modifications (26). Samples were sequenced on the Illumina MiSeq in the Teagasc
379 sequencing facility, with a 2 x 250 cycle V2 kit, in accordance with standard Illumina
380 sequencing protocols. Whole-metagenome shotgun libraries were prepared in accordance
381 with the Nextera XT DNA Library Preparation Guide from Illumina (26). Samples were
382 sequenced on the Illumina MiSeq in the Teagasc sequencing facility, with a 2 x 300 cycle V3
383 kit, in accordance with standard Illumina sequencing protocols.

384

385 **Bioinformatics**

386 Raw 16S rRNA gene sequencing reads were quality filtered using PRINSEQ (27). Denoising,
387 OTU clustering, and chimera removal were done using USearch (v7-64bit) (28), as described
388 by Doyle *et al.* (29). OTUs were aligned using PyNAST (30). Alpha-diversity and beta-
389 diversity were calculated using Qiime (1.8.0) (31). Taxonomy was assigned using a BLAST
390 search (32) against SILVA SSU 119 database (33).

391 Raw whole-metagenome shotgun sequencing reads were filtered, on the basis of quality and
392 quantity, and trimmed to 200 bp, with a combination of Picard Tools
393 (<https://github.com/broadinstitute/picard>) and SAMtools (34). MetaPhlAn2 was used to
394 characterise the microbial composition of samples at the species-level (35). MetaMLST (20),
395 PanPhlAn (19), and StrainPhlAn (21) were used to characterise the microbial composition of
396 the samples at the strain-level. GraPhlAn (36) was used to construct phylogenetic trees from
397 the StrainPhlAn output. SUPER-FOCUS (37) and HUMAnN2 (38) were used to determine
398 the microbial metabolic potential of samples. IDBA-UD (39) was used for metagenome
399 assembly.

400

401 **Accession numbers**

402 Sequence data have been deposited in the European Nucleotide Archive (ENA) under the
403 project accession number PRJEB20873.

404

405 **Statistical analysis**

406 Statistical analysis was done in R-3.2.2 (40). The Kruskal-Wallis test was done using the
407 compareGroups package, and the resulting p-values were for multiple comparisons. PCoA
408 analysis of 16S rRNA gene sequencing data was done using the phyloseq package (41).
409 Multidimensional scaling (MDS) was done using the vegan package. Data visualisation was
410 done using the ggplot2 package.

411

412 **References**

413

- 414 1. **Walsh AM, Crispie F, Claesson MJ, Cotter PD.** 2017. Translating Omics to Food
415 Microbiology. *Annual Review of Food Science and Technology* **8**.
- 416 2. **Zheng J, Zhao X, Lin XB, Gänzle M.** 2015. Comparative genomics *Lactobacillus*
417 *reuteri* from sourdough reveals adaptation of an intestinal symbiont to food
418 fermentations. *Scientific reports* **5**:18234.
- 419 3. **Sun Z, Harris HM, McCann A, Guo C, Argimón S, Zhang W, Yang X, Jeffery**
420 **IB, Cooney JC, Kagawa TF.** 2015. Expanding the biotechnology potential of
421 lactobacilli through comparative genomics of 213 strains and associated genera.
422 *Nature communications* **6**.
- 423 4. **Mellmann A, Harmsen D, Cummings CA, Zentz EB, Leopold SR, Rico A, Prior**
424 **K, Szczepanowski R, Ji Y, Zhang W.** 2011. Prospective genomic characterization of
425 the German enterohemorrhagic *Escherichia coli* O104: H4 outbreak by rapid next
426 generation sequencing technology. *PloS one* **6**:e22751.
- 427 5. **Dallman TJ, Byrne L, Ashton PM, Cowley LA, Perry NT, Adak G, Petrovska L,**
428 **Ellis RJ, Elson R, Underwood A.** 2015. Whole-Genome Sequencing for National
429 Surveillance of Shiga Toxin–Producing *Escherichia coli* O157. *Clinical Infectious*
430 *Diseases:civ318*.

- 431 6. **Kwong JC, Mercoulia K, Tomita T, Easton M, Li HY, Bulach DM, Stinear TP,**
432 **Seemann T, Howden BP.** 2016. Prospective whole-genome sequencing enhances
433 national surveillance of *Listeria monocytogenes*. *Journal of clinical microbiology*
434 **54**:333-342.
- 435 7. **De Filippis F, Genovese A, Ferranti P, Gilbert JA, Ercolini D.** 2016.
436 Metatranscriptomics reveals temperature-driven functional changes in microbiome
437 impacting cheese maturation rate. *Scientific reports* **6**.
- 438 8. **Quigley L, O'Sullivan DJ, Daly D, O'Sullivan O, Burdikova Z, Vana R,**
439 **Beresford TP, Ross RP, Fitzgerald GF, McSweeney PLH, Giblin L, Sheehan JJ,**
440 **Cotter PD.** 2016. Thermus and the Pink Discoloration Defect in Cheese. *mSystems* **1**.
- 441 9. **Walsh AM, Crispie F, Kilcawley K, O'Sullivan O, O'Sullivan MG, Claesson MJ,**
442 **Cotter PD.** 2016. Microbial Succession and Flavor Production in the Fermented
443 Dairy Beverage Kefir. *mSystems* **1**:e00052-00016.
- 444 10. **Yang X, Noyes NR, Doster E, Martin JN, Linke LM, Magnuson RJ, Yang H,**
445 **Geornaras I, Woerner DR, Jones KL.** 2016. Use of Metagenomic Shotgun
446 Sequencing Technology To Detect Foodborne Pathogens within the Microbiome of
447 the Beef Production Chain. *Applied and environmental microbiology* **82**:2433-2443.
- 448 11. **Leonard SR, Mammel MK, Lacher DW, Elkins CA.** 2015. Application of
449 metagenomic sequencing to food safety: detection of Shiga toxin-producing
450 *Escherichia coli* on fresh bagged spinach. *Applied and environmental microbiology*
451 **81**:8183-8191.
- 452 12. **Wang Q, Garrity GM, Tiedje JM, Cole JR.** 2007. Naive Bayesian classifier for
453 rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and*
454 *environmental microbiology* **73**:5261-5267.
- 455 13. **Allard G, Ryan FJ, Jeffery IB, Claesson MJ.** 2015. SPINGO: a rapid species-
456 classifier for microbial amplicon sequences. *BMC bioinformatics* **16**:1.
- 457 14. **Eren AM, Maignien L, Sul WJ, Murphy LG, Grim SL, Morrison HG, Sogin ML.**
458 2013. Oligotyping: differentiating between closely related microbial taxa using 16S
459 rRNA gene data. *Methods in Ecology and Evolution* **4**:1111-1119.
- 460 15. **Stellato G, Utter DR, Voorhis A, De Angelis M, Eren AM, Ercolini D.** 2017. A
461 few *Pseudomonas* oligotypes dominate in the meat and dairy processing environment.
462 *Frontiers in Microbiology* **8**.
- 463 16. **Lindgreen S, Adair KL, Gardner PP.** 2016. An evaluation of the accuracy and
464 speed of metagenome analysis tools. *Scientific reports* **6**:19233.

- 465 17. **Stasiewicz MJ, den Bakker HC, Wiedmann M.** 2015. Genomics tools in microbial
466 food safety. *Current Opinion in Food Science* **4**:105-110.
- 467 18. **Leonard SR, Mammel MK, Lacher DW, Elkins CA.** 2016. Strain-Level
468 Discrimination of Shiga Toxin-Producing *Escherichia coli* in Spinach Using
469 Metagenomic Sequencing. *PloS one* **11**:e0167870.
- 470 19. **Scholz M, Ward DV, Pasolli E, Tolio T, Zolfo M, Asnicar F, Truong DT, Tett A,**
471 **Morrow AL, Segata N.** 2016. Strain-level microbial epidemiology and population
472 genomics from shotgun metagenomics. *Nature methods*.
- 473 20. **Zolfo M, Tett A, Jousson O, Donati C, Segata N.** 2016. MetaMLST: multi-locus
474 strain-level bacterial typing from metagenomic samples. *Nucleic Acids*
475 *Research*:gkw837.
- 476 21. **Asnicar F, Manara S, Zolfo M, Truong DT, Scholz M, Armanini F, Ferretti P,**
477 **Gorfer V, Pedrotti A, Tett A, Segata N.** 2017. Studying Vertical Microbiome
478 Transmission from Mothers to Infants by Strain-Level Metagenomic Profiling.
479 *mSystems* **2**.
- 480 22. **Loman NJ, Constantinidou C, Christner M, Rohde H, Chan JZ-M, Quick J,**
481 **Weir JC, Quince C, Smith GP, Betley JR.** 2013. A culture-independent sequence-
482 based metagenomics approach to the investigation of an outbreak of Shiga-toxigenic
483 *Escherichia coli* O104: H4. *Jama* **309**:1502-1510.
- 484 23. **Akabanda F, Owusu-Kwarteng J, Tano-Debrah K, Glover RL, Nielsen DS,**
485 **Jespersen L.** 2013. Taxonomic and molecular characterization of lactic acid bacteria
486 and yeasts in nunu, a Ghanaian fermented milk product. *Food microbiology* **34**:277-
487 283.
- 488 24. **Marsh AJ, Hill C, Ross RP, Cotter PD.** 2014. Fermented beverages with health-
489 promoting potential: past and future perspectives. *Trends in Food Science &*
490 *Technology* **38**:113-124.
- 491 25. **Akabanda F, Owusu-Kwarteng J, Glover R, Tano-Debrah K.** 2010.
492 Microbiological characteristics of Ghanaian traditional fermented milk product, Nunu.
493 *Nature and Science* **8**:178-187.
- 494 26. **Clooney AG, Fouhy F, Sleator RD, O'Driscoll A, Stanton C, Cotter PD, Claesson**
495 **MJ.** 2016. Comparing Apples and Oranges?: Next Generation Sequencing and Its
496 Impact on Microbiome Analysis. *PloS one* **11**:e0148028.
- 497 27. **Schmieder R, Edwards R.** 2011. Quality control and preprocessing of metagenomic
498 datasets. *Bioinformatics* **27**:863-864.

- 499 28. **Edgar RC.** 2010. Search and clustering orders of magnitude faster than BLAST.
500 *Bioinformatics* **26**:2460-2461.
- 501 29. **Doyle CJ, Gleeson D, O'Toole PW, Cotter PD.** 2017. Impacts of Seasonal Housing
502 and Teat Preparation on Raw Milk Microbiota: a High-Throughput Sequencing Study.
503 *Applied and Environmental Microbiology* **83**:e02694-02616.
- 504 30. **Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R.**
505 2010. PyNAST: a flexible tool for aligning sequences to a template alignment.
506 *Bioinformatics* **26**:266-267.
- 507 31. **Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello**
508 **EK, Fierer N, Pena AG, Goodrich JK, Gordon JI.** 2010. QIIME allows analysis of
509 high-throughput community sequencing data. *Nature methods* **7**:335-336.
- 510 32. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local
511 alignment search tool. *Journal of molecular biology* **215**:403-410.
- 512 33. **Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J,**
513 **Glöckner FO.** 2013. The SILVA ribosomal RNA gene database project: improved
514 data processing and web-based tools. *Nucleic acids research* **41**:D590-D596.
- 515 34. **Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis**
516 **G, Durbin R.** 2009. The sequence alignment/map format and SAMtools.
517 *Bioinformatics* **25**:2078-2079.
- 518 35. **Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasoli E, Tett A,**
519 **Huttenhower C, Segata N.** 2015. MetaPhlan2 for enhanced metagenomic taxonomic
520 profiling. *Nature methods* **12**:902-903.
- 521 36. **Asnicar F, Weingart G, Tickle TL, Huttenhower C, Segata N.** 2015. Compact
522 graphical representation of phylogenetic data and metadata with GraPhlAn. *PeerJ*
523 **3**:e1029.
- 524 37. **Silva GGZ, Green KT, Dutilh BE, Edwards RA.** 2016. SUPER-FOCUS: a tool for
525 agile functional analysis of shotgun metagenomic data. *Bioinformatics* **32**:354-361.
- 526 38. **Abubucker S, Segata N, Goll J, Schubert AM, IZard J, Cantarel BL, Rodriguez-**
527 **Mueller B, Zucker J, Thiagarajan M, Henrissat B.** 2012. Metabolic reconstruction
528 for metagenomic data and its application to the human microbiome. *PLoS*
529 *computational biology* **8**:e1002358.
- 530 39. **Peng Y, Leung HC, Yiu S-M, Chin FY.** 2012. IDBA-UD: a de novo assembler for
531 single-cell and metagenomic sequencing data with highly uneven depth.
532 *Bioinformatics* **28**:1420-1428.

- 533 40. **Team RC.** 2014. R: A language and environment for statistical computing. R
534 Foundation for Statistical Computing, Vienna, Austria. 2013. ISBN 3-900051-07-0.
- 535 41. **McMurdie PJ, Holmes S.** 2013. phyloseq: an R package for reproducible interactive
536 analysis and graphics of microbiome census data. *PloS one* **8**:e61217.
- 537 42. **Fukushima M, Kakinuma K, Kawaguchi R.** 2002. Phylogenetic analysis of
538 Salmonella, Shigella, and Escherichia coli strains on the basis of the gyrB gene
539 sequence. *Journal of clinical microbiology* **40**:2779-2785.
- 540 43. **Scallan E, Hoekstra R, Mahon B, Jones T, Griffin P.** 2015. An assessment of the
541 human health impact of seven leading foodborne pathogens in the United States using
542 disability adjusted life years. *Epidemiology and infection* **143**:2795-2804.
- 543 44. **Nurk S, Meleshko D, Korobeynikov A, Pevzner PA.** 2017. metaSPAdes: a new
544 versatile metagenomic assembler. *Genome Research*:gr. 213959.213116.
- 545 45. **Abdelgadir W, Nielsen DS, Hamad S, Jakobsen M.** 2008. A traditional Sudanese
546 fermented camel's milk product, Gariss, as a habitat of *Streptococcus infantarius*
547 subsp. *infantarius*. *International journal of food microbiology* **127**:215-219.
- 548 46. **Jans C, Kaindi DWM, Böck D, Njage PMK, Kouamé-Sina SM, Bonfoh B,**
549 **Lacroix C, Meile L.** 2013. Prevalence and comparison of *Streptococcus infantarius*
550 subsp. *infantarius* and *Streptococcus gallolyticus* subsp. *macedonicus* in raw and
551 fermented dairy products from East and West Africa. *International journal of food*
552 *microbiology* **167**:186-195.
- 553 47. **Beck M, Frodl R, Funke G.** 2008. Comprehensive study of strains previously
554 designated *Streptococcus bovis* consecutively isolated from human blood cultures and
555 emended description of *Streptococcus gallolyticus* and *Streptococcus infantarius*
556 subsp. *coli*. *Journal of clinical microbiology* **46**:2966-2972.
- 557 48. **Herrero IA, Rouse MS, Piper KE, Alyaseen SA, Steckelberg JM, Patel R.** 2002.
558 Reevaluation of *Streptococcus bovis* endocarditis cases from 1975 to 1985 by 16S
559 ribosomal DNA sequence analysis. *Journal of clinical microbiology* **40**:3848-3850.
- 560 49. **Biarç J, Nguyen IS, Pini A, Gossé F, Richert S, Thiersé D, Van Dorsselaer A,**
561 **Leize-Wagner E, Raul F, Klein J-P.** 2004. Carcinogenic properties of proteins with
562 pro-inflammatory activity from *Streptococcus infantarius* (formerly *S. bovis*).
563 *Carcinogenesis* **25**:1477-1484.
- 564 50. **Tamhankar AJ, Nerkar SS, Khadake PP, Akolkar DB, Apurwa SR, Deshpande**
565 **U, Khedkar SU, Stålsby-Lundborg C.** 2015. Draft genome sequence of

566 enterotoxigenic *Escherichia coli* strain E24377A, obtained from a tribal drinking
567 water source in India. Genome announcements **3**:e00225-00215.
568

569 **Table 1. The results of MetaMLST and PanPhlAn analysis of spinach metagenomes**
 570 **spiked with *E. coli* O157:H7 Sakai**

Sequence accession number	Reads	<i>E. coli</i> abundance			Sequence type (ST)	Confidence (%)
		(%)	stx2A	stx2B		
SRR2177250	9,365,812	5.28412	1	1	Unknown	NA
SRR2177251	17,562,542	4.31712	1	1	11	99.97
SRR2177280	11,707,292	21.16364	1	1	100001	99.97
SRR2177281	10,580,532	2.84187	1	1	Unknown	NA
SRR2177282	6,155,636	60.51406	1	1	11	100
SRR2177283	13,120,244	10.11327	1	1	11	100
SRR2177284	7,500,056	2.05064	NA	NA	Unknown	NA
SRR2177285	14,482,370	66.69813	1	1	11	100
SRR2177286	14,035,970	69.17834	1	1	11	100
SRR2177287	12,242,348	5.62746	1	1	Unknown	NA
SRR2177288	8,303,788	10.75005	1	1	11	100
SRR2177357	14,621,672	8.02047	1	1	11	100
SRR2177358	10,684,052	3.18652	1	1	Unknown	NA
SRR2177359	4,964,436	1.17146	1	1	Unknown	NA
SRR2177360	12,729,834	1.81229	1	0	Unknown	NA
SRR2177361	11,946,092	0.70921	0	1	Unknown	NA

571

572 **Table 2. The results of MetaMLST analysis of the nunu metagenomic samples**

Species	Sequence type (ST)	Confidence (%)	Sample
<i>Klebsiella pneumoniae</i>	100001	98.7	1t2am
<i>Klebsiella pneumoniae</i>	100002	100	1t6am
<i>Esherichia coli</i>	100001	100	1t7am
<i>Klebsiella pneumoniae</i>	100003	99.9	1t7am
<i>Klebsiella pneumoniae</i>	100004	100	1t8am
<i>Klebsiella pneumoniae</i>	39	100	2u3am
<i>Klebsiella pneumoniae</i>	100005	99.91	2u6am
<i>Klebsiella pneumoniae</i>	100006	99.91	2u8am

573

574

575 **Figure legends**

576 **Figure 1. 16S rRNA gene sequencing based analysis of nunu samples.** (A) Heat map
577 showing the 25 most abundant bacterial genera across the nunu samples. (B) Bar plot shoing
578 genera which were differentially abundant in either group.

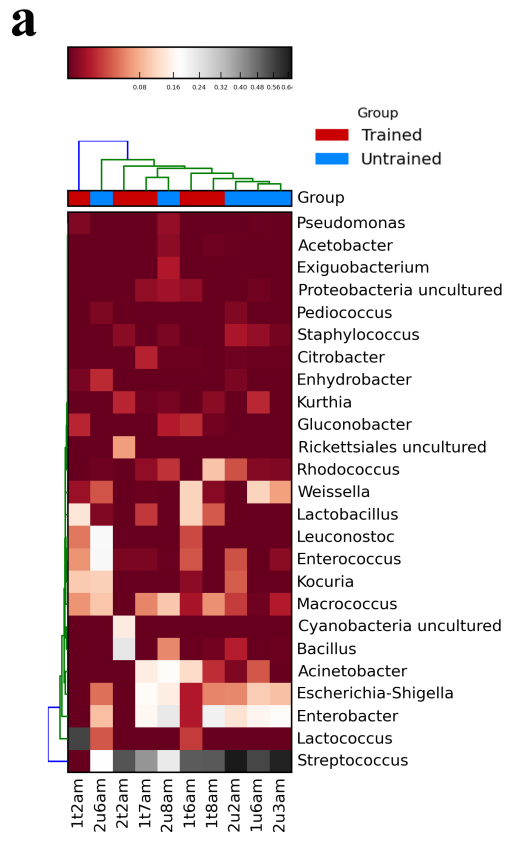
579 **Figure 2. The species-level microbial composition of nunu samples, as determined by**
580 **MetaPhlAn2.**

581 **Figure 3. The average abundances of the SUPER-FOCUS Level 1 functions that were**
582 **detected in nunu samples.**

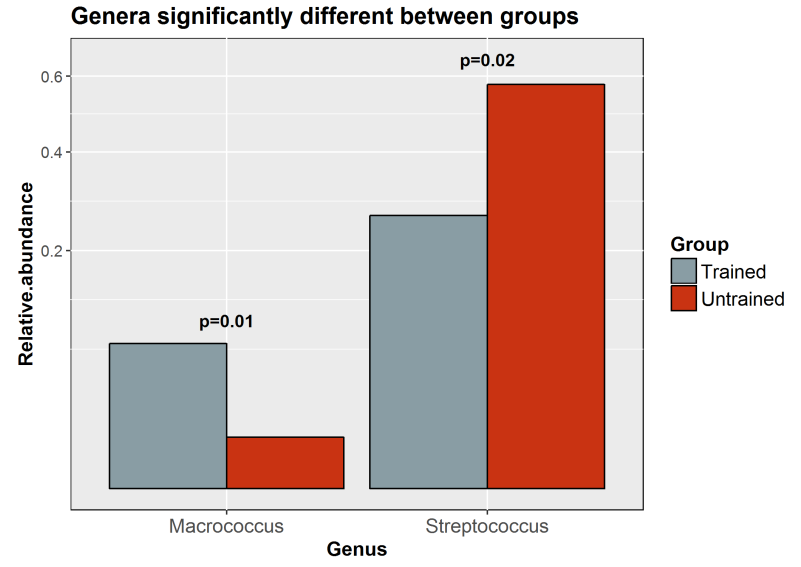
583 **Figure 4. HUMAnN2 analysis.** (A) Heat map showing the 25 most abundant MetaCyc
584 pathways detected across the ten nunu metagenomic samples. (B) Bar plot showing
585 differences in histidine metabolic potential between nunu samples from trained producers and
586 nunu samples from untrained producers. (C) Bar plots showing the relative contributions of
587 *E. cloacae*, *E. coli* and *K. pneumoniae* to the MetaCyc pathways PWY-6305 (putrescine
588 biosynthesis) and PWY0-1338 (polymyxin resistance).

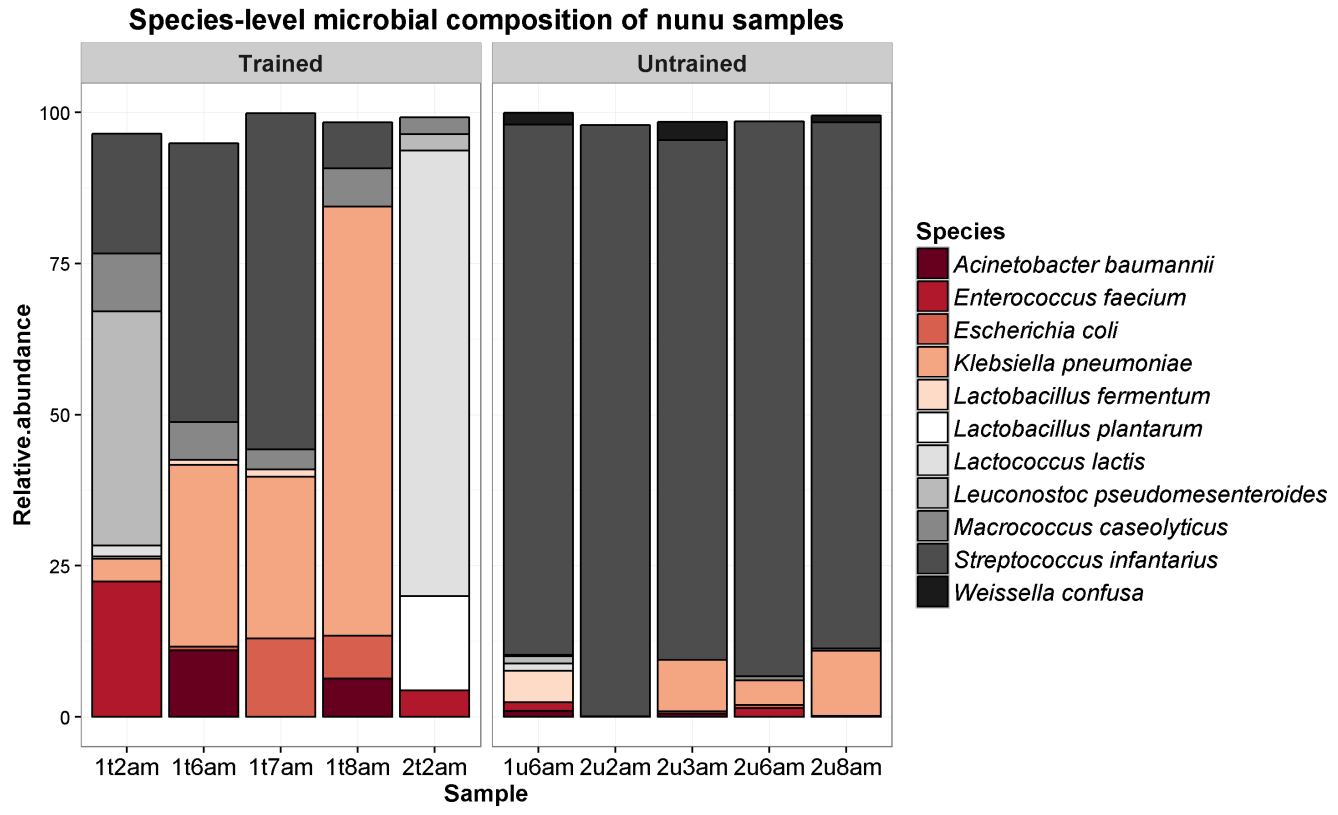
589 **Figure 5. StrainPhlAn analysis of the spinach metagenome.**

590 **Figure 6. Strain-level analysis.** Phylogenetic trees showing the relationships between (A) *E.*
591 *coli* strains and (B) *K. pneumoniae* strains detected in the nunu metagenomic samples and
592 their respective reference genomes, as predicted by StrainPhlAn. (C) MDS showing the
593 functional similarities between strains detected in the nunu metagenomic samples, as
594 predicted by PanPhlAn; reference genomes are shown in faded grey.



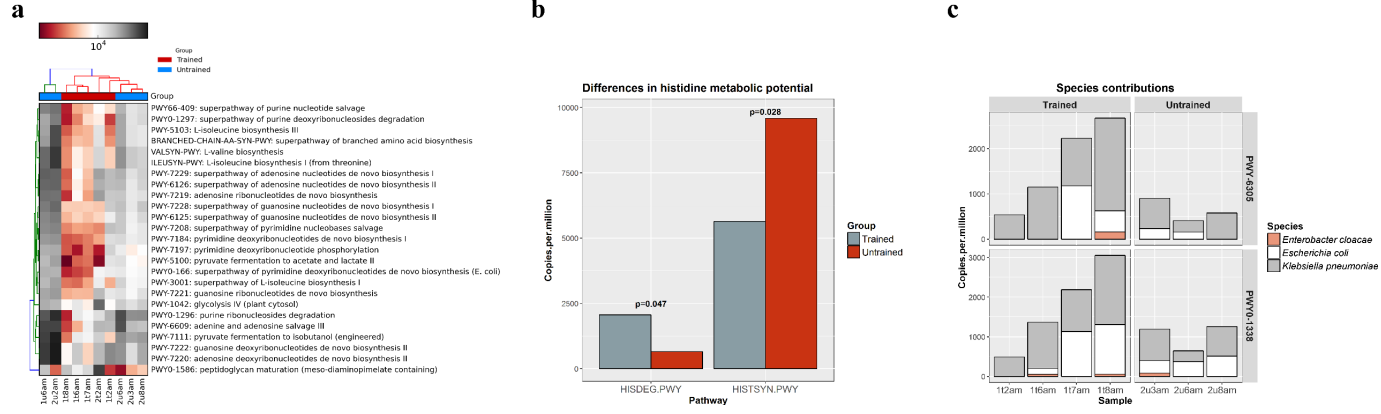
b





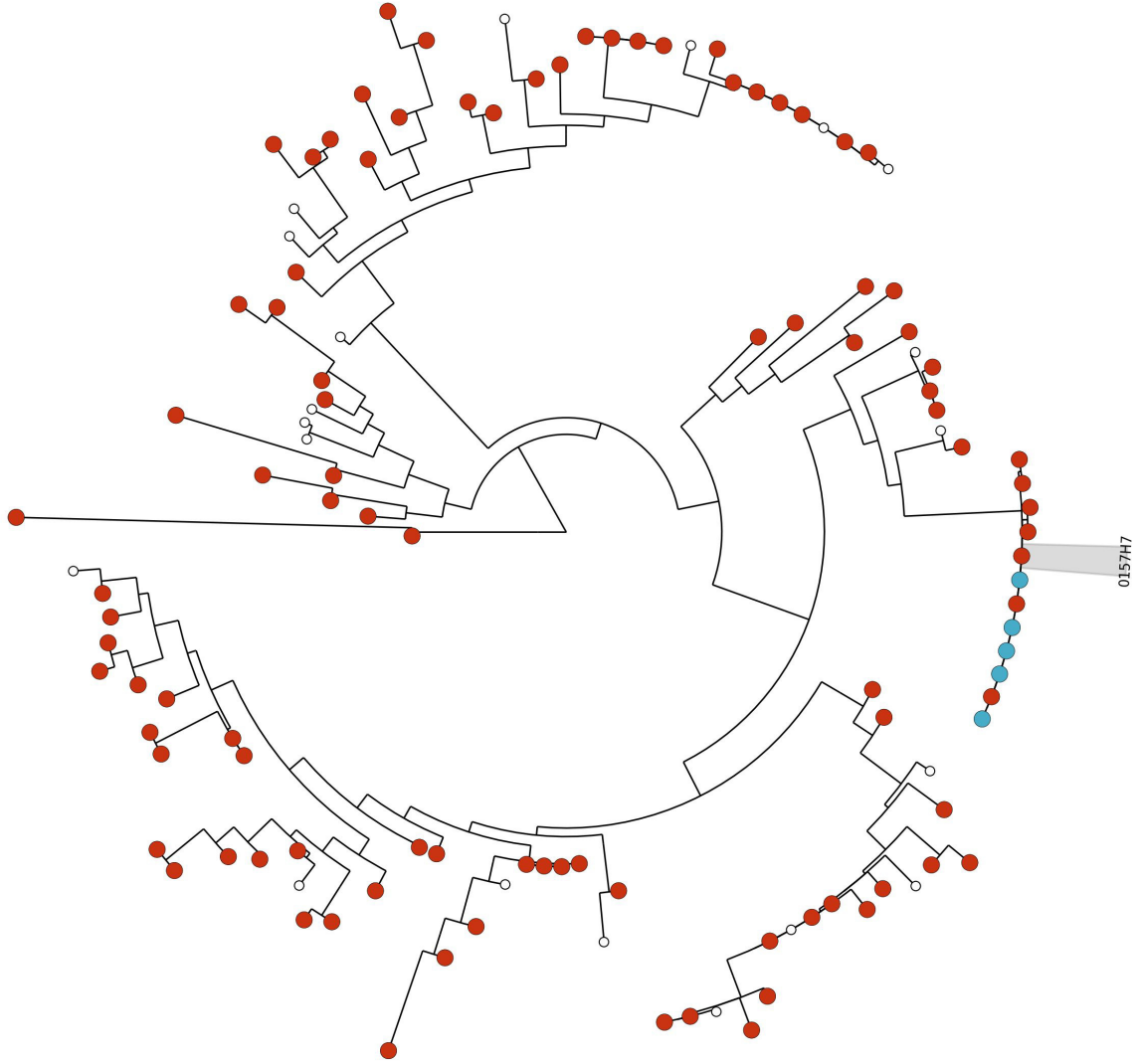
Overview of the metabolic potential of the nunu metagenome



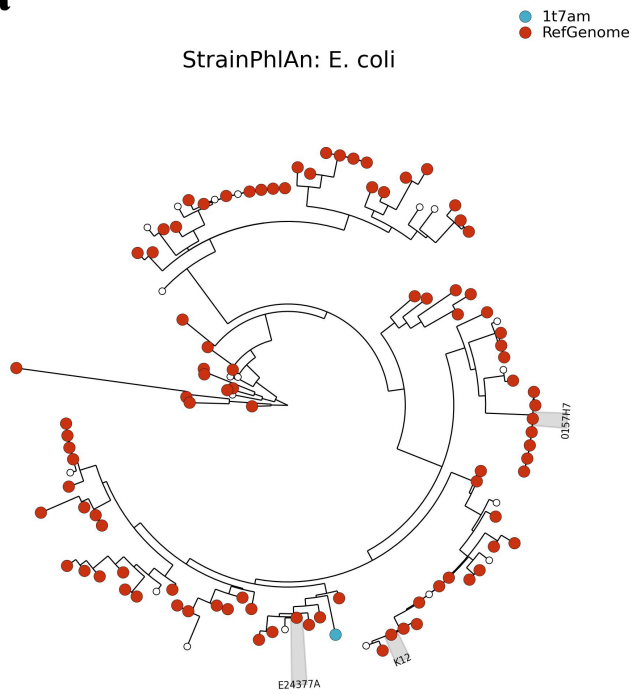


StrainPhlAn: E. coli (spinach metagenome)

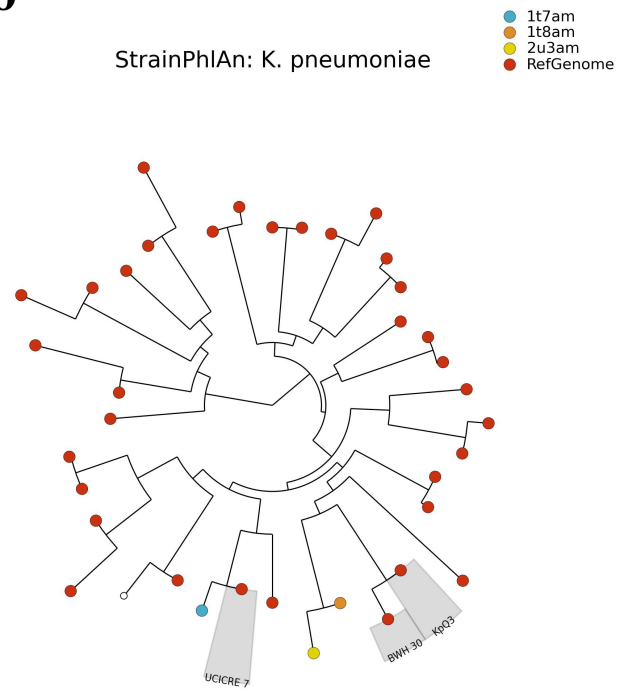
- RefGenome
- Spinach



a



b



c

