

1 **City-wide metagenomic surveillance of food centres reveals**  
2 **location-specific microbial signatures and enrichment of antibiotic**  
3 **resistance genes**

4 Jonathan J.Y. Teo<sup>1</sup>, Eliza Xin Pei Ho<sup>1</sup>, Amanda Hui Qi Ng<sup>1</sup>, Shaun Hong Chuen How<sup>1</sup>,  
5 Kern Rei Chng<sup>1,2</sup>, Yiğit Can Ateş<sup>3</sup>, Muhd Tarmidzi Fau'di<sup>2,4</sup>, Kyaw Thu Aung<sup>2,4,5,6</sup>,  
6 Niranjan Nagarajan<sup>1,7#</sup>

7 *<sup>1</sup>Genome Institute of Singapore, Agency for Science, Technology and Research*  
8 *(A\*STAR), Singapore 138672, Singapore*

9 *<sup>2</sup>National Centre for Food Science, Singapore Food Agency, Singapore 609919,*  
10 *Singapore*

11 *<sup>3</sup>College of Engineering, Koç University, Istanbul 34450, Türkiye*

12 *<sup>4</sup>Environmental Health Institute, National Environment Agency, Singapore 138667,*  
13 *Singapore*

14 *<sup>5</sup>School of Biological Sciences, Nanyang Technological University, Singapore 637551,*  
15 *Singapore*

16 *<sup>6</sup>Department of Food Science & Technology, National University of Singapore,*  
17 *Singapore 117543, Singapore*

18 *<sup>7</sup>Yong Loo Lin School of Medicine, National University of Singapore, Singapore*  
19 *117596, Singapore*

20 *#Corresponding Author*

21 *Lead Contact: [nagarajann@gis.a-star.edu.sg](mailto:nagarajann@gis.a-star.edu.sg)*

## 22 **Abstract**

23 The distribution of microorganisms in built environments with high human traffic, such  
24 as food centres, can potentially have a significant impact on public health, particularly  
25 in the context of increasing worldwide incidence of food and fomite-related outbreaks.  
26 In several major Asian cities, public food centres are the main venue for food  
27 consumption and yet we lack a baseline understanding of their environmental  
28 microbiomes. We conducted city-wide metagenomic surveillance of food-centre  
29 microbiomes in Singapore (16 centres, n=240 samples) to provide a detailed map of  
30 microbial (bacteria, archaea, fungi, viruses) as well as non-microbial DNA abundances  
31 across two timepoints. Food-centre microbiomes were found to be enriched in food-  
32 related DNA signatures compared to other environments such as hospitals and offices,  
33 with specific food-microbe associations (e.g. Enterobacteriaceae and fish) and food  
34 DNA providing a partial explanation for the microbial profiles observed (44% of  
35 variation explained). Machine learning analysis identified a small set of microbial  
36 species (n=22) that serve as highly accurate (>80%) location-specific signatures for  
37 various food centres, some of which persist even after 3 years. Profiling of antibiotic  
38 resistance genes (ARGs) and pathogens identified a surprising enrichment of ARGs  
39 in food centres relative to other non-healthcare environments (>2.5×), and an order of  
40 magnitude enrichment of key pathogenic species (e.g. *Klebsiella pneumoniae*,  
41 *Enterobacter spp*) even compared to hospital environments. These results highlight  
42 the contribution of diverse biotic and abiotic factors in shaping the unique microbiome  
43 profiles of different food-centre environments, and the potential for using metagenomic  
44 surveillance to understand the risk for infections and antibiotic resistance gene  
45 transmission.

## 46 Introduction

47 Globally, the WHO estimates that there are more than 600 million food-borne illnesses  
48 every year leading to >400,000 deaths<sup>1</sup>. The rising incidence of antimicrobial resistant  
49 infections further compounds this challenge with new and emerging pathogens such  
50 as Group B *Streptococcus*<sup>2</sup> being of significant public health concern. In several major  
51 Asian cities (e.g. Singapore, Kuala Lumpur and Bangkok), large public food centres  
52 (also known as hawker centres) are popular venues for daily food consumption for a  
53 significant proportion of the population<sup>3</sup>. Hawkers centers typically represent an  
54 improvement in plumbing infrastructure, water supply and sanitation to reduce the  
55 incidence of foodborne illnesses<sup>4</sup>. Nevertheless as communal eating spaces, they also  
56 have high foot traffic and sharing of tables, plates and utensils, potentially creating an  
57 environment conducive to the spread of pathogens. According to a 2020 report by the  
58 Singapore Ministry of Health, there were 1,894 and 473 cases of Salmonellosis and  
59 Campylobacteriosis in 2019, and >2500 reported food poisoning cases that were  
60 epidemiologically linked to food establishments<sup>5</sup>. Despite improved sanitation, food-  
61 related gastroenteritis persists, highlighting the potential risk for disease transmission  
62 in public eateries. Enhanced surveillance in food centres through microbiological  
63 culture-based and genomic approaches thus has the potential to improve our  
64 understanding of the distribution of pathogens and antibiotic resistance determinants  
65 in the environment, and inform cleaning protocols and other risk mitigation strategies.

66 A growing body of work has highlighted the use of metagenomics as a  
67 surveillance technology for built environments<sup>6</sup>, particularly in areas where there is  
68 high human traffic (e.g. subways<sup>7,8</sup> and offices<sup>9,10</sup>) or where the risk of infection in  
69 vulnerable populations is higher (e.g. hospitals<sup>9,11</sup> and neonatal ICUs<sup>12</sup>). The  
70 International MetaSUB consortium, in particular, has pioneered the global collection  
71 of metagenomic datasets from cities worldwide over many years to understand the  
72 distribution of environmental microbes and the antibiotic resistance genes (ARGs) that  
73 they carry<sup>13</sup>. In addition to 16S rRNA sequencing approaches which provide a  
74 convenient and cost-effective taxonomic survey of environmental microbiomes,  
75 shotgun metagenomics has been particularly popular for such surveys owing to its  
76 ability to provide multi-kingdom characterization of samples as well as information on  
77 ARGs. Despite this, relatively few studies have characterized the microbiomes of food  
78 centres, beyond the use of traditional culture techniques and PCR analysis for select

79 pathogens and ARGs<sup>14–16</sup>. While a few studies have used 16S rRNA sequencing to  
80 characterize bacterial genera in kitchens and slaughterhouses<sup>17,18</sup>, our understanding  
81 of food-centre microbiomes in relation to other built-environment microbiomes is still  
82 limited. In particular, while food centres are not as aggressively cleaned as hospitals<sup>19</sup>,  
83 they still have more frequent cleaning schedules than other public locations such as  
84 offices and subways. In addition, the use of antibiotics in food production<sup>20</sup> has the  
85 potential to impact the resistome of food centres where large quantities of food are  
86 processed and consumed<sup>21</sup>. Finally, food-associated microbes, particularly from  
87 fermented foods that are common in Southeast Asian cuisine<sup>22,23</sup> (e.g. in fish sauce,  
88 fermented bean paste and tempeh), as well as from food spoilage in a hot and humid  
89 environment<sup>24,25</sup>, could further lend unique microbial community profiles in food  
90 centres.

91 To address this knowledge gap, we conducted a city-wide metagenomic  
92 surveillance of food-centre microbiomes in Singapore (16 centres, n=240 samples) to  
93 provide a detailed map of microbial (bacteria, archaea, fungi, viruses) as well non-  
94 microbial DNA abundances across two timepoints. Notably, our analysis identified  
95 enrichment in food-related DNA in food centres compared to other environments,  
96 enabling the identification of specific food-microbe associations (e.g.  
97 Enterobacteriaceae and fish), as well as evidence that food associations could provide  
98 a partial explanation for the microbial profiles observed (44% of variation explained).  
99 Using machine learning methods we identified a small set of microbial species (n=22)  
100 that can also serve as highly accurate (>80%) location-specific signatures for various  
101 food centres, some of which persist even after 3 years. Profiling of antibiotic resistance  
102 genes (ARGs) and pathogens identified a surprising enrichment of ARGs in food  
103 centres relative to other non-healthcare environments (>2.5×), and an order of  
104 magnitude enrichment of key pathogenic species (e.g. *Klebsiella pneumoniae*,  
105 *Enterobacter spp*) even compared to hospital environments. These findings highlight  
106 the diverse biotic and abiotic contributions that could shape the unique microbiome  
107 profiles of food-centre environments, and highlights the utility of food-centre  
108 metagenomic surveillance for future studies aimed at reducing the emergence and  
109 spread of food and fomite-associated infections and human gut colonization by multi-  
110 drug resistant bacteria.

## 111 Results

### 112 City-wide metagenomic surveillance of food centres highlights enrichment of 113 food-related DNA signatures that partially explain microbial profiles

114 To cover a diversity of food-centre environments in Singapore, we selected a  
115 geographically distributed set of sites located in the central, eastern and western parts  
116 of Singapore that span the densely populated urban landscape of the city-state (n=16;  
117 **Figure 1A, Supplementary File 1**). In each site, 10 tables were swabbed based on  
118 established MetaSUB protocols<sup>7</sup> and subjected to deep shotgun metagenomic  
119 sequencing (20 million reads on average/sample, **Methods**), providing data for  
120 relatively comparable high-touch sites across various food centres (**Figure 1B**). A  
121 second set of 80 swabs were collected from these sites 3 years later to assess  
122 temporal stability, providing a complete dataset of 226 metagenomes (<6% of samples  
123 failed library preparation, **Methods**). The shotgun sequencing data was classified  
124 using a custom database of non-redundant genomes from NCBI to enable  
125 identification of microbial (e.g. bacteria, archaea, fungi, viruses) as well as non-  
126 microbial (e.g. human, plant, animal) reads, and compared to control samples to  
127 identify and remove potential signals of laboratory and reagent contamination  
128 (**Methods; Supplementary File 2**). While the majority of identified taxa were microbial  
129 and bacterial as expected (median relative abundance=77%), a substantial proportion  
130 of DNA was also classified to non-microbial origins (median relative abundance=12%,  
131 **Figure 1C**) including diverse animal and plant orders (**Supplementary Figure 1**). We  
132 noted that several of the most abundant orders within the *Animalia* kingdom all had  
133 potential food origins including red meat, chicken, duck, oysters and fish  
134 (**Supplementary Figure 1A**), while plant orders highlighted potential food sources that  
135 are common in local cuisine including peppers, fleshy and leafy vegetables, rice and  
136 wheat (**Supplementary Figure 1B**). Interestingly, manual curation of food-related  
137 taxa highlighted that corresponding animal and plant orders were strongly enriched for  
138 food-related DNA in food-centre metagenomes (>90%, Fisher's exact p-value<0.0001;  
139 **Supplementary Figure 2, Methods**). Correspondingly, the total relative abundance  
140 of food-related taxa was also significantly enriched in food-centre metagenomes  
141 relative to other locations such as hospital sites<sup>9</sup>, office<sup>9</sup> and outdoor environments  
142 surveyed as part of MetaSUB Singapore<sup>13</sup> (Wilcoxon p-value<0.01; **Supplementary**  
143 **Figure 3A**), with vegetables, fish and meat representing the dominant groups in food-

144 centres (**Supplementary Figure 3B**). These observations highlight the ability to  
145 identify potential food-related DNA sources through metagenomic surveillance with  
146 particular relevance to food-centre surveillance.

147 We next investigated the potential impact of various food sources on the  
148 microbial community using compositionality-aware correlation analysis with SparCC<sup>26</sup>  
149 (**Methods**). As a positive control, common skin commensal species from the  
150 *Cutibacterium* genus, were observed to be strongly correlated with the abundance of  
151 human DNA, in addition to weaker and non-specific associations of *Vibrio* with fish<sup>27</sup>  
152 and *Salmonella* with poultry<sup>28</sup> (**Figure 1D**). Some bacterial (e.g. *Bacillus*) and fungal  
153 (e.g. *Exophiala*) genera were found to have high correlations with most food-related  
154 taxa (Spearman  $\rho > 0.7$ ; Group A, **Figure 1D**; **Supplementary Figure 4**), consistent  
155 with their known association with food environments and fermentation<sup>29,30</sup>. Another  
156 distinct group of species that includes clinically-relevant opportunistic pathogens from  
157 the family Enterobacteriaceae (e.g. *Salmonella*, *Citrobacter*, *Klebsiella*) had strong  
158 correlations with a subset of food-related taxa including fish species (e.g. carp), meat  
159 and rice (Group B, **Figure 1D**) that are commonly featured in Asian cuisine, a notable  
160 association that has not been previously described. Other strong associations were  
161 more specific to a small subset of food sources e.g. *Sphingobacterium* with salmon  
162 and *Aspergillus* with soldierfish and fruits (Group C; **Figure 1D**), which could indicate  
163 a specific source of origin for these organisms. Interestingly, while most of the detected  
164 correlations were positive (72%, 300/418), weaker negative correlations were also  
165 detected (Spearman  $\rho < 0.5$ ) and were often seen consistently in soil, plant-associated  
166 and environmental bacteria (e.g. *Frankia*, *Rhodobacter*, *Azoarcus*). Overall, these  
167 results highlight that while some microbes might be associated with food-related DNA  
168 abundances in general (**Supplementary Figure 4**), other environmental variables are  
169 likely to impact the observed distribution of microbes in food-centre metagenomes as  
170 well (percent variation explained by food-related taxa = 44%; **Supplementary Figure**  
171 **5**).

172 In terms of microbial taxa, we noted that while bacterial and fungal species from  
173 the genus *Klebsiella*, *Enterobacter*, *Acinetobacter*, *Moraxella* and *Candida* were the  
174 most commonly seen in food centres metagenomes in this study, the relative  
175 proportion of these species varied by location and sampling surface (**Supplementary**  
176 **Figure 6**). For example, while food centres C, E, F and N were marked by high

177 abundance of *Klebsiella* species, food centres A, B, F, K, O exhibited profiles with a  
178 higher proportion of *Moraxella* species instead. Overall taxonomic alpha diversities  
179 were observed to be higher in food centres relative to other urban environmental  
180 metagenomes that we could compare to such as hospital, office and MetaSUB  
181 datasets (**Supplementary Figure 7**). Some food centres house market stalls that offer  
182 raw ingredients such as meat, vegetables and dairy (**Supplementary File 1**), and  
183 while the presence of such stalls was associated with higher microbial diversity  
184 (Wilcoxon p-value<0.05; **Supplementary Figure 8A**), the correlation between the  
185 number of stalls and Shannon diversity was weak (Pearson  $\rho=0.17$ ; **Supplementary**  
186 **Figure 8B**). Furthermore, no significant correlation was evident when considering the  
187 number of individual food stalls within the surveyed food centres (**Supplementary**  
188 **Figure 8C**), suggesting that other intrinsic factors, such as geography and  
189 environment, may play an important role in shaping the higher diversities observed in  
190 the metagenomes of these food centres. Intriguingly, microbiomes within the same  
191 food centre exhibited greater similarity to each other than to those from different food  
192 centres ( $R^2=0.873$ , p-value= $10^{-4}$ ; **Figure 1E**). In addition, clustering of microbial  
193 taxonomic profiles resulted in clusters that were more similar to location-based groups  
194 than clusters based on non-microbial taxonomic profiles (Wilcoxon p-value< $10^{-4}$ ;  
195 **Supplementary Figure 9; Methods**). Together, these observations indicate that while  
196 variability in food-centre microbiomes might partially reflect food-source variations,  
197 food-centre locations and related environmental factors are also likely to play a  
198 substantial role in shaping the diversity of microbiota compositions observed.

### 199 **Food-centre metagenomes exhibit location-specific microbial signatures that** 200 **can show long-term persistence**

201 Based on the observation that food-centre metagenomic profiles cluster by location,  
202 we next investigated if a machine learning approach could be used to predict locations  
203 and identify location-specific metagenomic signatures. As metagenomic data is  
204 compositional and high-dimensional in nature, different normalization and feature-  
205 selection approaches were explored to construct models (**Figure 2A; Methods**).  
206 Comparison of performance from different multi-class classification approaches (n=11)  
207 highlighted that while high AUC-ROC scores could be achieved (>0.95, 4-fold cross-  
208 validation), linear classifiers generally performed better for this task relative to non-  
209 linear models (e.g. quadratic discriminant analysis - QDA), and appropriate

210 normalization (ILR: isometric log-ratio transformation, CLR: centered log-ratio  
211 transformation) was essential to achieve good performance (**Figure 2B**). Interestingly,  
212 logistic regression (LogReg) classifiers exhibited the highest performance across  
213 datasets, providing interpretable model weights which are useful for identifying  
214 important features for subsequent biological investigations (**Figure 2B**). In addition,  
215 microbial features were more predictive of location than non-microbial features, with  
216 performance with microbial features alone being similar to the full model, indicating  
217 that location-specific metagenomic signatures are primarily microbial in information  
218 content (**Figure 2B**).

219 To identify the essential subset of features, logistic-regression classifiers were  
220 trained with varying number of features, highlighting that as few as 20 features could  
221 be sufficient to achieve optimal performance (AUC-ROC>0.95, Precision>90%,  
222 Accuracy>80%; **Figure 2C**), though the use of principal components generally led to  
223 lower performance. The most informative features were then selected using recursive  
224 feature elimination with cross-validation (RFECV; **Methods**), yielding 28 species on  
225 average across cross-validation runs. Taking the union of species across 4 cross-  
226 validation runs identified a core set of 22 species that were consistently selected in at  
227 least 3 out of 4 runs (**Table 1**). These 22 species collectively exhibited mean combined  
228 relative abundance of 4.5% across all samples, indicating that the identified microbial  
229 signatures represent non-dominant species in the metagenome that nevertheless  
230 have consistent differences across food centres (**Figure 2D**). Of note, several species  
231 are known to have environmental associations (e.g. *Azorhizobium caulinodans* and  
232 *Dyella japonica* in soil samples<sup>31</sup>), while other species are known to have associations  
233 with foods (e.g. *Clavispora lusitaniae* and *Lactobacillus farciminis* in cheese and  
234 kimchi respectively<sup>32,33</sup>) and food additives (e.g. *Cyberlindnera jadinii* for its high  
235 glutamic acid content<sup>34</sup>), highlighting the diverse potential sources for these species.

236 To assess the temporal persistence of food-centre microbial signatures, we  
237 applied the classifier model to a set of metagenomes (n=79) obtained from the same  
238 food centres more than 3 years later in 2022. Remarkably, the classifier yielded an  
239 AUC-ROC score exceeding 0.7 for food centres C, D, E, H, I, J, L, M and P, which is  
240 significantly higher than a permutation-based null model (9/15; **Figure 2E**). In addition,  
241 ordination analysis revealed that these food centres (stable – S) had less changes to  
242 their microbial profiles compared to other food centres (unstable – U), between the



243 two timepoints of sample collection over a period of three years (**Figure 2F**), further  
244 underscoring the persistent nature of microbial signatures in specific food-centres.  
245 Next, we fine-tuned the classifiers with a single sample from each food-centre from  
246 the second collection (20% of samples) and found a substantial rise in AUC-ROC  
247 scores, surpassing an average value of 0.8 (**Figure 2G**). These results suggest that  
248 location-specific microbial signatures in select food centres remain stable over  
249 extended periods of time, and in cases where the microbiome has shifted over time,  
250 the classifier can be fine-tuned with a small set of new data to improve classification  
251 performance.

252 Generalizing beyond location-specific signatures, we explored if the microbial  
253 profiles carried geographical information. Testing for a distance-decay relationship<sup>35</sup>  
254 between microbial profiles (Bray-Curtis dissimilarity) and different food-centre  
255 locations in Singapore (Euclidean distance), we observed a weak but statistically  
256 significant correlation ( $\rho=0.082$ , Mantel test  $p\text{-value}<0.008$ ). Training a classifier  
257 based on geographical cluster labels (**Supplementary Figure 10**) achieved slightly  
258 lower classification accuracy than for location labels (accuracy= 80%, AUC-ROC=0.9),  
259 while using more features (**Supplementary Figure 11A**). Among the features, 8  
260 species were found in at least 75% of the cross-validation runs (**Supplementary**  
261 **Figure 10B**), identifying 3 species that were not identified by the location-based  
262 classifier, including a newly isolated species from air samples in Singapore  
263 (*Brachybacterium sp. SGAir0954*; **Supplementary Figure 11C**). These results  
264 highlight the potential for environmental microbes in food centres to serve as forensic  
265 signatures of location and geography, as well as providing an important baseline for  
266 identification of potential pathogens and their resistance determinants.

### 267 **Enrichment of antibiotic resistance genes and pathogenic species in food-** 268 **centre microbiomes**

269 To assess the potential role that food-centre microbiomes may play in harbouring  
270 antimicrobial resistance determinants and facilitating their onward transmission, we  
271 profiled the relative abundance of antibiotic resistance genes (ARGs) in food-centre  
272 metagenomes (**Methods**). Food-centre metagenomes harboured relatively higher  
273 ARG abundance compared to other environmental metagenomes including office  
274 environments and outdoor environments profiled as part of a MetaSUB consortium  
275 study ( $>2.5\times$ , Wilcoxon  $p\text{-value}<10^{-4}$ ; **Figure 3A**). Hospital environments were the

276 sites that exhibited highest median relative abundance of antibiotic resistance genes  
277 (**Figure 3A**), and this was true across different resistance gene classes, except for  
278 colistin resistance which was most enriched in food centres ( $>1.5\times$ , Wilcoxon p-  
279 value $<10^{-4}$ ; **Figure 3B**). This is noteworthy, as colistin is often an antibiotic of last  
280 resort and its indiscriminate use in animal husbandry has been seen as a risk factor  
281 for transmission of resistance to human pathogens<sup>36</sup>. Interestingly, for a few other  
282 classes of antibiotic resistance (e.g. aminoglycosides, beta-lactamases and  
283 Fosfomycin), metagenomes of food-centres, a built environment where high human  
284 traffic is typically expected, resembled hospital environments in having higher median  
285 abundances than office and MetaSUB sites (**Figure 3B, Supplementary Figure 12**),  
286 highlighting the importance of food-centre microbiomes as sites for one-health  
287 surveillance. While overall levels of beta-lactamase resistance genes were similar, the  
288 profiles of specific resistance genes detected were found to belong to 3 major clusters,  
289 with cluster 1 being shared between food-centres and hospitals, cluster 2 being unique  
290 to food-centres, and cluster 3 being predominantly seen in hospitals (**Figure 3C**). In  
291 particular, beta-lactamase genes commonly found in pathogenic species such as  
292 *Acinetobacter baumannii*<sup>37</sup>, and *Enterobacter cloacae*<sup>38</sup> (e.g. bla<sub>ADC</sub>, bla<sub>CMH</sub>, bla<sub>ACT</sub>,  
293 bla<sub>MIR</sub>) were relatively more abundant in food centres than in hospitals (Cluster 2;  
294 **Figure 3C**), suggesting that these pathogens may be enriched in food centres.  
295 Additionally, Carbapenemase genes (e.g. bla<sub>OXA</sub>) were also detected in both food  
296 centres and hospitals, but while several bla<sub>OXA</sub> variants from hospitals have known  
297 associations with plasmids, none of the bla<sub>OXA</sub> genes detected in food centres were  
298 linked to plasmids (**Supplementary Figure 13**), indicating that the risk of plasmid-  
299 mediated transmission may be lower in food-centres than in hospitals.

300 We next assessed the abundance of high-priority ESKAPE pathogens  
301 (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*,  
302 *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, *Enterobacter spp.* and  
303 *Escherichia coli*) in food-centres relative to other environments and found that they  
304 were surprisingly strongly enriched in food-centres ( $>3\times$ , Wilcoxon p-value $<10^{-4}$ ;  
305 **Figure 3D**). In particular, gram-negative Enterobacteriaceae such as *K. pneumoniae*,  
306 *A. baumannii*, *Enterobacter* and *E. coli* (but not *P. aeruginosa*) were found to be  
307 significantly enriched in food-centres than even in hospital environments (Wilcoxon p-  
308 value $<10^{-4}$ ; **Figure 3D**). Overall, these results highlight the utility of metagenomic

309 surveillance in food-centre environments for understanding the distribution of key  
310 pathogens and resistance genes, and for further studies into transmission risk.

## 311 Discussion

312 Despite their social and cultural importance in many major Asian cities, and increasing  
313 concerns over new food and fomite-associated pathogens, large-scale microbial  
314 genomic surveillance of food centres have not been reported before. Our study helps  
315 establish the feasibility of a shotgun metagenomics approach to rapidly conduct city-  
316 wide surveillance (estimated sequencing cost <S\$10,000), and highlights the utility of  
317 this data for tracking microbes and associated genes of concern in the environment.  
318 In addition to microbes, our work shows that metagenomics can reflect specific  
319 enrichment of food-associated non-microbial taxa in food-centres (**Supplementary**  
320 **Figure 1-3**), though further work is needed to understand how much this reflects food  
321 production and consumption patterns. By jointly detecting microbes and food-  
322 associated taxa, metagenomics allows us to discover associations between them that  
323 may be relevant for tracking the source of food-related pathogens (**Figure 1D**). For  
324 e.g., we noted that while some microbes seem to be non-specifically associated to  
325 food-related taxa (such as *Bacillus*, *Exophiala*), potentially reflecting a joint seeding  
326 pattern or their general ability to grow better with food-derived nutrients, others have  
327 more specific associations consistent with prior studies on a food item of interest (e.g.  
328 *Vibrio* species in seafood<sup>27</sup> and *Salmonella* in poultry<sup>28</sup>). Larger-scale analysis may  
329 thus help uncover more such associations, especially those that are for rarer microbial  
330 taxa and foods that are unique to a few food-centres. Further advances in databases  
331 of microbial and non-microbial genomes<sup>39,40</sup>, development of taxonomic classifiers  
332 that are more specific and sensitive<sup>41</sup>, long-read sequencing and genome-resolved  
333 metagenomics<sup>42,43</sup> are expected to further aid such studies.

334 A notable result from this study is that food-associated non-microbial taxa can  
335 potentially explain a substantial fraction, but not all, of the microbial variation observed  
336 in food-centres (44%; **Supplementary Figure 5**), indicating the importance of food  
337 sources in determining food-centre microbiomes. Additionally, it is also clear that there  
338 is a strong location-specific signature in metagenomic profiles, which is primarily  
339 microbially driven (**Figure 2B**). Similar “forensic” signatures have been reported in  
340 other studies as well<sup>44</sup>, but primarily over larger geographic distances (e.g. cities  
341 around the world<sup>8,45</sup>). The training of accurate location classifiers (Accuracy>80%) was

342 facilitated by appropriate pre-processing of relative abundance information from  
343 metagenomics, enabling highly compact signatures with as few as 20 species to be  
344 learnt (**Figure 2C**). Interestingly, these signatures are not defined by abundant taxa  
345 (e.g. *Klebsiella*, *Enterobacter* etc; **Supplementary Figure 6**) but by rarer taxa with  
346 diverse possible origins (**Table 1**). The higher biomass collection enable by our  
347 sampling protocol, as well as the stringency of our 'kitome' removal analysis  
348 (**Methods**), suggests that these signatures are not likely to be artefactual. In addition,  
349 metagenomic data from the second timepoint, more than 3 years later, highlights that  
350 some signatures can be remarkably robust over time (**Figure 2E, 2F**). This is despite  
351 large socio-environmental changes (COVID-19 pandemic related), as well as  
352 unavoidable technical differences (newer DNA extraction protocols; **Methods**)  
353 separating the two timepoints. Nevertheless, it highlights how microbes can serve as  
354 biomarkers and indicator species for food-centre locations, with further study needed  
355 to understand the reasons for a greater shift in these signatures in some sites (**Figure**  
356 **2F**). The ability to rescue classifier performance with a few additional datasets  
357 suggests that these shifts may not lead to a complete replacement of such signatures  
358 (**Figure 2G**). Overall, geographical factors appear to provide only a partial explanation  
359 for the observed location-specific signatures (**Supplementary Figure 11**), and larger  
360 datasets with detailed environmental information could help tease apart the  
361 contributions of diverse local factors in shaping food-centre microbiomes.

362 Food-centre metagenomes were found to be surprisingly enriched for antibiotic  
363 resistance genes relative to other outdoor and indoor environments with high human  
364 contact (>2.5×; **Figure 3A**), with the not so surprising exception of hospital  
365 environments<sup>9</sup>. Of note, food-centres seem to be more similar to hospital  
366 environments in the enrichment of various clinically important antibiotic resistance  
367 gene classes (e.g. aminoglycosides, beta-lactamases, Fosfomycin; **Figure 3B**). This  
368 could in part be due to antibiotic usage in animal husbandry<sup>36</sup> and agriculture<sup>46</sup> globally,  
369 though the use of antibiotics to promote growth of animals is not allowed in Singapore  
370 and so, for example, the enrichment of colistin resistance is unexpected<sup>47</sup>. Other  
371 potential explanations include, higher human traffic than other sites, greater microbial  
372 biomass thriving on food-based nutrients, and specific environments and interventions  
373 (e.g. cleaning protocols) that enrich for bacteria carrying such resistance genes<sup>48</sup>.  
374 While the public health ramifications of this enrichment remain to be determined (e.g.

375 chromosomally integrated carbapenemases in food-centre microbiomes may not be  
376 as concerning as plasmid-borne genes detected in hospital environments<sup>49</sup>), it is clear  
377 that food-centre environments should be an important node for one-health surveillance  
378 efforts. This is particularly the case because of the unexpectedly strong enrichment of  
379 several ESKAPE pathogens of concern, even relative to hospital environments (>10×  
380 for several gram-negative Enterobacteriaceae; **Figure 3D**), and the need to  
381 understand if this has an impact on the risk for transmission and infections. Overall,  
382 this study highlights the valuable information that metagenomic surveillance can  
383 provide about microbes as well as non-microbes in food-centre environments, serving  
384 as the basis for further large-scale cross-sectional and interventional studies to  
385 understand the diverse factors that shape them, their impact on public health, and the  
386 utility of various cleaning and behavioural intervention strategies to reduce infection  
387 risk.

388

## 389 **Materials and Methods**

### 390 **Sample collection**

391 Microbiome sampling was done on food centre table surfaces using Isohelix DNA  
392 Buccal Swabs (SK-4S) based on MetaSUB protocols<sup>7</sup>. Samples were collected from  
393 16 different geographically spread-out locations from around Singapore that represent  
394 popular food centres in densely populated regions (**Figure 1A**). Sampling was done  
395 at two different timepoints (March 2019 and July 2022) within 1 week, where all sites  
396 from the same food-centre were sampled on the same day. The second collection was  
397 delayed due to lockdowns during the COVID-19 pandemic period. During the first  
398 timepoint, 10 different table surfaces were swabbed (total n=160), while 5 tables were  
399 swabbed at the second timepoint (total n=80). Swabbing was done with two swabs  
400 applied over separate halves of a table, each for a duration of two minutes, and the  
401 samples were combined. Swabs were stored in Zymo DNA/RNA Shield™ immediately  
402 after collection and the corresponding tubes were subsequently stored at -80°C prior  
403 to DNA extraction.

### 404 **DNA extraction and sequencing**

405 Extraction of total DNA from swabs was performed via a combination of mechanical  
406 and chemical lysis. Briefly, samples were homogenized on the FastPrep Instrument  
407 (MP Biomedicals) at 6 ms<sup>-1</sup> for 40 s, followed by centrifugation (5mins, 13000 rpm)  
408 and Proteinase K treatment of supernatant (56°C for 20mins). Purified DNA was  
409 obtained using the Maxwell RSC Blood DNA Kit (Promega, AS1400) for the first  
410 collection and QIAamp PowerFecal Pro DNA Kit (Qiagen) for the second collection  
411 according to manufacturer's instructions. DNA concentration was quantified on the  
412 Qubit dsDNA HS Assay Kit (Life Technologies, Q32854). Samples with low DNA  
413 concentration according to the assay were excluded from the study (First collection,  
414 n=13; Second collection, n=1). Metagenomic libraries for all samples were prepared  
415 with the NEBNext® Ultra™ II FS DNA kit (New England Biolabs, E7805L) according  
416 to the manufacturer's instructions. Paired-end sequencing (2×150bp reads) was  
417 performed on the Illumina HiSeq X™ Ten platform. Negative control samples (n=5,  
418 blank swabs in DNA/RNA shield) underwent identical DNA extraction and sequencing  
419 procedures.

## 420 **Taxonomic abundance analysis**

421 Illumina paired-end reads were first trimmed for adapters and filtered for low quality  
422 reads using fastp (v0.20.1 -q 25 -p 20)<sup>50</sup>. The remaining reads were used for  
423 taxonomic profiling with Kraken2<sup>51</sup> (v2.0.8) using a custom index (kraken2build –max-  
424 db-size 60000) built from the NCBI NT (non-redundant nucleotide) genome database.  
425 Taxonomic abundances at the species and genus level were estimated using  
426 Bracken<sup>52</sup> (v2.5).

427 Abundance profiles were filtered such that species with relative abundance less than  
428 0.01% were excluded to reduce the impact of false-positive calls. To assess the  
429 potential impact of reagent and laboratory contamination on taxonomic profiles, DNA  
430 quantification was done for negative control libraries (n=5) and found to be a 100-fold  
431 smaller than libraries for all collected food-centre samples, suggesting that  
432 reagent/laboratory contamination is unlikely to impact the taxonomic profiles for our  
433 relatively high-biomass environments (enabled by 2-swab sampling of the entire table).  
434 As a conservative measure to reduce the impact of contaminants, species that were  
435 present in negative controls were labelled as putative contaminants if their abundance  
436 was inversely correlated (pearson correlation >0.3) with DNA concentration in food-  
437 centre metagenomes, and correspondingly removed from food-centre microbiome  
438 profiles before further analysis. Species from *Animalia* and *Plantae* were manually  
439 curated to be food-related if they were known to be common ingredients in local cuisine  
440 (**Supplementary File 3**).

## 441 **Taxonomic co-occurrence analysis**

442 Taxonomic co-occurrence relationships between microbial and non-microbial species  
443 were inferred through the application of the SparCC<sup>26</sup> method, as implemented in  
444 FastSpar v0.0.10<sup>53</sup> with default parameters. Correlations with absolute coefficients  
445 exceeding 0.3 and a statistical significance threshold of p-value<0.05 were retained.  
446 The resulting correlations were visually represented using a heatmap, generated with  
447 the seaborn clustermap function in Python, employing the Euclidean distance metric.

## 448 **Ordination analysis**

449 Ordination analysis was performed using the Uniform Manifold Approximation and  
450 Projection algorithm<sup>54</sup> (UMAP, n\_neighbour=10, min\_dist=0.1, n\_components=2). The  
451 Spearman distance metric, which ranges from 0 to 2, was used to measure similarity

452 between subsets of taxonomic profiles. The selection of Spearman distance metric  
453 over a more conventional Bray-Curtis dissimilarity metric was aimed at capitalizing on  
454 a rank-based methodology to avoid biasing towards highly abundant species. This  
455 step is valuable for measuring similarities between profiles containing only a subset of  
456 taxa in a microbiome, such as those representing signature species.

#### 457 **Canonical correspondence analysis**

458 Canonical correspondence analysis (CCA) was performed using the R package  
459 *vegan*<sup>55</sup> (v2.6.4) to investigate the relationship between microbial composition  
460 variations and food-related taxa abundances. In total, 35 food-related orders were  
461 examined (**Supplementary File 3**). Percentage of microbial variation explained by  
462 food-related taxa was obtained from the ratio of constrained inertia to total inertia. The  
463 biplot and model inertia results are shown in **Supplementary Figure 5**.

#### 464 **Clustering and cluster similarity**

465 Metagenomes were initially clustered based on their microbial abundance profiles or  
466 profiles for food-related taxa. These cluster assignments were subsequently  
467 compared both against the true location labels of the samples and among themselves  
468 using *clusim*<sup>56</sup> (v0.4), with higher scores denoting increased similarity. To assess the  
469 robustness of clustering, we conducted 10-fold cross-validation to generate multiple  
470 similarity scores across the three pairs of comparisons (microbe-food, food-location,  
471 microbe-location; **Supplementary Figure 9**). Additionally, a null model was  
472 established through random permutations of cluster labels to compute similarity scores  
473 and test for statistical significance (one-sided Wilcoxon test).

#### 474 **Training of machine learning models**

475 Multi-class machine learning classifiers were trained to predict the most likely source  
476 of a sample based on its metagenomic profile. To limit the number of features in the  
477 training dataset, we restricted the models to the top 200 species based on median  
478 abundances. Taxonomic profiles were normalized using various approaches including  
479 total sum scaling (TSS), cumulative sum scaling (CSS, *metagenomeSeq*<sup>57</sup> v3.18),  
480 centered log-ratio (CLR) and isometric log-ratio (ILR) as described by Thomas et al..  
481 A pseudo count of 1 was applied across the dataset to allow for log transformation.  
482 Classifiers were trained using 11 different algorithms, including logistic regression,  
483 multi-layer perceptron, support vector classifier (linear basis function), K-nearest



484 neighbor, Gaussian naïve Bayes, random forest, decision tree, Adaboost, quadratic  
485 discriminant analysis, Gaussian process, and support vector classifiers (radial basis  
486 function), as implemented in the python scikit-learn package. Binary classifiers were  
487 adapted for multi-label classification using a one-vs-rest approach. The classifiers  
488 were first trained on a randomly selected training dataset (n=110) and applied to a  
489 withheld test dataset (n=37) to assess classifier performance. This process was  
490 repeated 4 times for different splits of the dataset and the mean accuracy and AUC-  
491 ROC scores (one-vs-rest, average using 'macro') from the test datasets was reported.  
492 Subsequent analysis was limited to the logistic regression classifier as it performed  
493 the best. To test for statistical significance of classification scores at each food-centre  
494 location, a null model was trained from 5 sets of random permutations of the food-  
495 centre labels. A one-sided Wilcoxon rank-sum test was performed subsequently to  
496 identify locations with significantly higher classification performance.

#### 497 **Geographical association analysis**

498 Food-centres were grouped together using hierarchical clustering based on Euclidean  
499 distances of their geographical location. Mantel's test (scikit bio v0.1.3) was performed  
500 between a distance matrix of Bray-Curtis dissimilarity scores between taxonomic  
501 profiles and a second distance matrix of Euclidean distances calculated from  
502 geographical coordinates of the food-centres.

#### 503 **Resistome analysis**

504 Antibiotic resistance gene (ARG) profiles were calculated using SRST2<sup>58</sup> (v0.1.4; --  
505 min\_coverage 100, hits with identity <99% were filtered out) using a curated version  
506 of the CARD database<sup>59</sup> (v3.0.8). Alignment counts to ARGs were normalized by total  
507 reads per sample and gene length to calculate RPKM values. Resistome profiles were  
508 aggregated based on the ontology grouping of resistance genes in the CARD  
509 database, including grouping by gene class and resistance mechanism.

#### 510 **Data Availability**

511 Sequencing reads are available from the European Nucleotide Archive under project  
512 PRJEB73308. Source code and data for reproducing figures are available under MIT  
513 license at: [https://github.com/CSB5/food\\_centre\\_microbiome](https://github.com/CSB5/food_centre_microbiome).

514 **Tables**

515 **Table 1: Details for core species that act as location-specific microbial**  
 516 **signatures.**

Species	Location	Mean RA (%)	Habitat	Type	Ref.
<i>Mycolicibacterium boenickei</i>	M	0.04		Gram+	60
<i>Cyberlindnera jadinii</i>	D, M	0.95	Food	Yeast	34
<i>Klebsiella</i> sp. FDAARGOS_511		0.04	Gut	Gram-	6101/03/2024 10:07:00
<i>Clavispora lusitaniae</i>	N	0.18	Food, Gut	Yeast	62
<i>Aquitalea</i> sp. USM4		0.05	Aquatic	Gram-	63
<i>Lodderomyces elongisporus</i>	L	0.07	Food, Soil, Blood	Yeast	64
<i>Dyella japonica</i>	C, H, M	0.40	Soil	Gram-	65
<i>Kushneria konosiri</i>	F, G	0.05	Food (Salt)	Gram-	66
<i>Mycolicibacterium litorale</i>	C, M	0.05	Soil	Gram+	60
<i>Salinicola tamaricis</i>	F	0.08	Plant	Gram-	67
<i>Rhodobacter</i> sp. LPB0142	A, J, P	0.07	Aquatic	Gram-	
<i>Pseudomonas oleovorans</i>	J, O	0.13	Aquatic, lipid-rich	Gram-	68
<i>Azorhizobium caulinodans</i>	A, C	0.11	Soil, Plant	Gram-	31,69
<i>Kluyvera georgiana</i>	F	0.34	Wastewater, Soil	Gram-	70
<i>Enterobacter cloacae</i>	N	0.16	Plant, Soil, Humans	Gram-	71
<i>Leuconostoc lactis</i>	I	0.04	Food	Gram+	72
<i>Rhizobium</i> sp. S41	G	0.06		Gram-	
<i>Staphylococcus saprophyticus</i>	F	0.09	Skin	Gram+	73

<i>Xanthobacter autotrophicus</i>	C, J, P	0.06	Wastewater	Gram-	74
<i>Dyella jiangningensis</i>		0.09	Soil, Rock	Gram-	75
<i>Enterobacter</i> sp. T2		0.06	Wastewater	Gram+	76
<i>Luteibacter rhizovicinus</i>	M	0.04	Food	Gram-	77

## 518 **Figure Legends**

519 **Figure 1: Distribution of microbial and non-microbial taxa in food-centre**  
520 **metagenomes.** (A) Map of Singapore showing population density and locations of the  
521 16 food centres that were sampled. (B) Schematic showing a typical food centre  
522 setting and sample collection approach for the project. (C) Boxplots showing kingdom-  
523 level taxonomic abundances (microbial and non-microbial) based on shotgun  
524 metagenomics data (first timepoint). (D) Heatmap of correlation coefficients between  
525 microbial genera and non-microbial orders as determined by SparCC (first timepoint;  
526 \*=key known associations). Only correlations with p-value<0.05 are shown. In addition,  
527 only taxa which have at least one correlation value>0.5 are shown. (E) Uniform  
528 Manifold Approximation and Projection (UMAP) analysis showing the degree of  
529 similarity of taxonomic profiles across food centres based on a Spearman distance  
530 metric (first timepoint). PERMANOVA analysis was done based on food-centre labels  
531 to obtain the R<sup>2</sup> and p-value.

532 **Figure 2: Predicting food-centre locations based on microbial profiles.** (A)  
533 Schematic of workflow used to assess various normalization techniques and do  
534 feature selection. (B) Area-under-curve of the receiver operating characteristic curve  
535 (AUC-ROC) for predicting the source of a taxonomic profile using different classifiers  
536 (linear models in red text) and normalization techniques. The scores were calculated  
537 based on classification of held-out test datasets and 4-fold cross-validation runs. (C)  
538 Performance of classifiers with increasing number of features evaluated using 3  
539 scoring metrics - classification accuracy, AUC-ROC score and average precision from  
540 the precision-recall curve. Scores are averaged over all food centres. (D) 3D  
541 Histogram showing the normalized abundances of 22 microbes that were identified as  
542 microbial signatures for various food centres. (E) Boxplots showing the AUC-ROC  
543 classification performance across food-centres using a logistic-regression classifier  
544 trained with samples from the first collection to classify samples collected 3 years later.  
545 Scores are averaged across the classifiers from 4 cross-validation runs. Star notation  
546 indicates significant p-values based on the one-sided Wilcoxon rank-sum test (\*\*\*\*: p-  
547 value<10<sup>-4</sup>, \*\*\*: p-value<10<sup>-3</sup>, \*\*: p-value<10<sup>-2</sup>, \*: p-value<0.05). (F) UMAP plot of  
548 microbial signature profiles from both the first and second collection. Only the  
549 centroids from each food-centre cluster are shown. Lines connect the centroids of

550 each food-centre at both timepoints and their opacities increase with shorter distances  
551 to indicate higher similarity based on the Spearman distance metric. Inset shows  
552 UMAP distances between timepoints for all pairs of tables in food-centres which have  
553 stable (S) and unstable (U) classification models (**Figure 2E**). Star notation indicates  
554 significant p-values based on the one-sided Wilcoxon rank-sum test (\*\*\*\*:  $p\text{-value} < 10^{-4}$ ).  
555 4). (G) Comparison of classification performance between classifiers with and without  
556 fine-tuning. Fine-tuning describes the process where the model is trained with the  
557 original dataset plus a small subset of additional data from the second collection (1 out  
558 of 5 samples per food-centre). The mean AUC-ROC scores averaged across food-  
559 centres and cross-validation runs are shown (line) along with the 95% confidence  
560 interval (shading). Wilcoxon rank-sum test was performed on AUC-ROC scores from  
561 models utilizing 20 features.

562 **Figure 3: Distribution of antibiotic resistance genes and pathogens across food**  
563 **centres.** Boxplots showing (A) antibiotic resistance gene (ARG) abundance (reads  
564 per kilobase per million, RPKM) across 4 location classes: food-centres, hospital sites,  
565 office and outdoor environments surveyed as part of MetaSUB Singapore ( $n=221, 429,$   
566  $23$  and  $99$  respectively; two-sided Wilcoxon rank-sum test, \*\*\*\*:  $p\text{-value} < 10^{-4}$ ). (B)  
567 ARG abundance (RPKM) for various classes of antibiotics across the 4 different  
568 locations. Statistically significant differences between food-centres and other locations  
569 are shown (two-sided Wilcoxon rank-sum test, \*\*\*\*:  $p\text{-value} < 10^{-4}$ , \*\*\*:  $p\text{-value} < 10^{-3}$ , \*\*:  $p\text{-value} < 10^{-2}$ , \*:  $p\text{-value} < 0.05$ ). (C) Hierarchically-clustered heatmap showing  
570 carbapenemase gene abundances (RPKM) in various locations. (D) Strip plots  
571 showing the median relative abundance of each ESKAPE pathogen across 4 different  
572 location classes (two-sided Wilcoxon rank-sum test, \*\*\*\*:  $p\text{-value} < 10^{-4}$ ).  
573

574 **References**

- 575 1. WHO Estimates of the Global Burden of Foodborne Diseases: Foodborne  
576 Disease Burden Epidemiology - Reference Group 2007–2015. (2015).
- 577 2. Seale, A. C. *et al.* Estimates of the Burden of Group B Streptococcal Disease  
578 Worldwide for Pregnant Women, Stillbirths, and Children. *Clinical Infectious*  
579 *Diseases* **65**, S200–S219 (2017).
- 580 3. Duruz, J. & Khoo, G. C. *Eating Together: Food, Space, and Identity in Malaysia*  
581 *and Singapore*. (Rowman & Littlefield, 2014).
- 582 4. Kong, L. *Singapore Hawker Centres: People, Places, Food*. vol. 1 (2007).
- 583 5. Ministry of Health, S. Communicable Diseases Surveillance in Singapore 2019-  
584 2020. (2020).
- 585 6. Ko, K. K. K., Chng, K. R. & Nagarajan, N. Metagenomics-enabled microbial  
586 surveillance. *Nat Microbiol* **7**, 486–496 (2022).
- 587 7. The MetaSUB International Consortium. The Metagenomics and Metadesign of  
588 the Subways and Urban Biomes (MetaSUB) International Consortium inaugural  
589 meeting report. *Microbiome* **4**, 24 (2016).
- 590 8. Zhang, R., Walker, A. R. & Datta, S. Unraveling city-specific signature and  
591 identifying sample origin locations for the data from CAMDA MetaSUB challenge.  
592 *Biol Direct* **16**, 1 (2021).
- 593 9. Chng, K. R. *et al.* Cartography of opportunistic pathogens and antibiotic  
594 resistance genes in a tertiary hospital environment. *Nat Med* **26**, 941–951 (2020).
- 595 10. Gilbert, J. A. & Stephens, B. Microbiology of the built environment. *Nat Rev*  
596 *Microbiol* **16**, 661–670 (2018).
- 597 11. Lax, S. *et al.* Bacterial colonization and succession in a newly opened hospital.  
598 *Sci. Transl. Med.* **9**, eaah6500 (2017).

- 599 12. Hourigan, S. K. *et al.* Comparison of Infant Gut and Skin Microbiota, Resistome  
600 and Virulome Between Neonatal Intensive Care Unit (NICU) Environments.  
601 *Frontiers in Microbiology* **9**, (2018).
- 602 13. Danko, D. *et al.* A global metagenomic map of urban microbiomes and  
603 antimicrobial resistance. *Cell* **184**, 3376-3393.e17 (2021).
- 604 14. Borrusso, P. A. & Quinlan, J. J. Prevalence of Pathogens and Indicator  
605 Organisms in Home Kitchens and Correlation with Unsafe Food Handling  
606 Practices and Conditions. *Journal of Food Protection* **80**, 590–597 (2017).
- 607 15. Møretrø, T. *et al.* Consumer practices and prevalence of *Campylobacter*,  
608 *Salmonella* and norovirus in kitchens from six European countries. *International*  
609 *Journal of Food Microbiology* **347**, 109172 (2021).
- 610 16. Osaili, T. M. *et al.* Microbiological quality of kitchens sponges used in university  
611 student dormitories. *BMC Public Health* **20**, 1322 (2020).
- 612 17. Carstens, C. K., Salazar, J. K., Sharma, S. V., Chan, W. & Darkoh, C. Evaluation  
613 of the kitchen microbiome and food safety behaviors of predominantly low-  
614 income families. *Frontiers in Microbiology* **13**, (2022).
- 615 18. Shedleur-Bourguignon, F. *et al.* Distinct Microbiotas Are Associated with Different  
616 Production Lines in the Cutting Room of a Swine Slaughterhouse.  
617 *Microorganisms* **11**, 133 (2023).
- 618 19. Saito, R. *et al.* Characterization of cleaning and disinfecting tasks and product  
619 use among hospital occupations. *American Journal of Industrial Medicine* **58**,  
620 101–111 (2015).
- 621 20. Van Boeckel, T. P. *et al.* Global trends in antimicrobial use in food animals. *Proc.*  
622 *Natl. Acad. Sci. U.S.A.* **112**, 5649–5654 (2015).

- 623 21. Wu, S. *et al.* Staphylococcus aureus Isolated From Retail Meat and Meat  
624 Products in China: Incidence, Antibiotic Resistance and Genetic Diversity. *Front.*  
625 *Microbiol.* **9**, 2767 (2018).
- 626 22. Rodzi, N. A. R. M. & Lee, L. K. Traditional fermented foods as vehicle of non-  
627 dairy probiotics: Perspectives in South East Asia countries. *Food Research*  
628 *International* **150**, 110814 (2021).
- 629 23. Campbell-Platt, G. Fermented foods — a world perspective. *Food Research*  
630 *International* **27**, 253–257 (1994).
- 631 24. Lax, S. *et al.* Microbial and metabolic succession on common building materials  
632 under high humidity conditions. *Nat Commun* **10**, 1767 (2019).
- 633 25. Smith, T. P. *et al.* Community-level respiration of prokaryotic microbes may rise  
634 with global warming. *Nat Commun* **10**, 5124 (2019).
- 635 26. Friedman, J. & Alm, E. J. Inferring Correlation Networks from Genomic Survey  
636 Data. *PLoS Comput Biol* **8**, e1002687 (2012).
- 637 27. Messelhäusser, U. *et al.* Detection and differentiation of *Vibrio* spp. in seafood  
638 and fish samples with cultural and molecular methods. *International Journal of*  
639 *Food Microbiology* **142**, 360–364 (2010).
- 640 28. Yildirim, Y., Gonulalan, Z., Pamuk, S. & Ertas, N. Incidence and antibiotic  
641 resistance of *Salmonella* spp. on raw chicken carcasses. *Food Research*  
642 *International* **44**, 725–728 (2011).
- 643 29. Moktan, B., Saha, J. & Sarkar, P. K. Antioxidant activities of soybean as affected  
644 by *Bacillus*-fermentation to kinema. *Food Research International* **41**, 586–593  
645 (2008).



- 646 30. Chancharonpong, C., Hsieh, P.-C. & Sheu, S.-C. Enzyme Production and  
647 Growth of *Aspergillus oryzae* S. on Soybean Koji Fermentation. *APCBEE*  
648 *Procedia* **2**, 57–61 (2012).
- 649 31. Ladha, J., Garcia, M., Miyan, S., Padre, A. T. & Watanabe, I. Survival of  
650 *Azorhizobium caulinodans* in the soil and rhizosphere of wetland rice under  
651 *Sesbania rostrata*-rice rotation. *Applied and environmental microbiology* **55**, 454–  
652 460 (1989).
- 653 32. Kaminarides, S. E. & Anifantakis, E. M. Evolution of the microflora of Kopanisti  
654 cheese during ripening. Study of the yeast flora. *Lait* **69**, 537–546 (1989).
- 655 33. Yoon, J. H. *et al.* *Lactobacillus kimchii* sp. nov., a new species from kimchi.  
656 *International Journal of Systematic and Evolutionary Microbiology* **50**, 1789–1795  
657 (2000).
- 658 34. Sousa-Silva, M., Vieira, D., Soares, P., Casal, M. & Soares-Silva, I. Expanding  
659 the Knowledge on the Skillful Yeast *Cyberlindnera jadinii*. *J Fungi (Basel)* **7**, 36  
660 (2021).
- 661 35. Bell, T. Experimental tests of the bacterial distance–decay relationship. *ISME J* **4**,  
662 1357–1365 (2010).
- 663 36. Kempf, I., Jouy, E. & Chauvin, C. Colistin use and colistin resistance in bacteria  
664 from animals. *International Journal of Antimicrobial Agents* **48**, 598–606 (2016).
- 665 37. Bhattacharya, M., Toth, M., Antunes, N. T., Smith, C. A. & Vakulenko, S. B.  
666 Structure of the extended-spectrum class C  $\beta$ -lactamase ADC-1 from  
667 *Acinetobacter baumannii*. *Acta Cryst D* **70**, 760–771 (2014).
- 668 38. Ku, Y.-H., Lee, M.-F., Chuang, Y.-C. & Yu, W.-L. Detection of Plasmid-Mediated  
669  $\beta$ -Lactamase Genes and Emergence of a Novel AmpC (CMH-1) in *Enterobacter*

- 670 cloacae at a Medical Center in Southern Taiwan. *Journal of Clinical Medicine* **8**, 8  
671 (2019).
- 672 39. Parks, D. H. *et al.* GTDB: an ongoing census of bacterial and archaeal diversity  
673 through a phylogenetically consistent, rank normalized and complete genome-  
674 based taxonomy. *Nucleic Acids Research* **50**, D785–D794 (2022).
- 675 40. Nasko, D. J., Koren, S., Phillippy, A. M. & Treangen, T. J. RefSeq database  
676 growth influences the accuracy of k-mer-based lowest common ancestor species  
677 identification. *Genome Biology* **19**, 165 (2018).
- 678 41. Kobus, R. *et al.* A big data approach to metagenomics for all-food-sequencing.  
679 *BMC Bioinformatics* **21**, 102 (2020).
- 680 42. Bertrand, D. *et al.* Hybrid metagenomic assembly enables high-resolution  
681 analysis of resistance determinants and mobile elements in human microbiomes.  
682 *Nat Biotechnol* **37**, 937–944 (2019).
- 683 43. Kolmogorov, M. *et al.* metaFlye: scalable long-read metagenome assembly using  
684 repeat graphs. *Nat Methods* **17**, 1103–1110 (2020).
- 685 44. Kawulok, J., Kawulok, M. & Deorowicz, S. Environmental metagenome  
686 classification for constructing a microbiome fingerprint. *Biology Direct* **14**, 20  
687 (2019).
- 688 45. Leung, M. H. Y. *et al.* Characterization of the public transit air microbiome and  
689 resistome reveals geographical specificity. *Microbiome* **9**, 112 (2021).
- 690 46. Umair, M. *et al.* International manufacturing and trade in colistin, its implications  
691 in colistin resistance and One Health global policies: a microbiological, economic,  
692 and anthropological study. *The Lancet Microbe* **4**, e264–e276 (2023).

- 693 47. Teo, J. W., Chew, K. L. & Lin, R. T. Transmissible colistin resistance encoded by  
694 mcr-1 detected in clinical Enterobacteriaceae isolates in Singapore. *Emerging*  
695 *Microbes & Infections* **5**, 1–12 (2016).
- 696 48. Poole, K. Mechanisms of bacterial biocide and antibiotic resistance. *Journal of*  
697 *Applied Microbiology* **92**, 55S-64S (2002).
- 698 49. Sakamoto, N. *et al.* Role of Chromosome- and/or Plasmid-Located bla<sub>NDM</sub> on  
699 the Carbapenem Resistance and the Gene Stability in Escherichia coli.  
700 *Microbiology Spectrum* **10**, e00587-22 (2022).
- 701 50. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ  
702 preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
- 703 51. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken  
704 2. *Genome Biol* **20**, 257 (2019).
- 705 52. Lu, J., Breitwieser, F. P., Thielen, P. & Salzberg, S. L. Bracken: estimating  
706 species abundance in metagenomics data. *PeerJ Computer Science* **3**, e104  
707 (2017).
- 708 53. Watts, S. C., Ritchie, S. C., Inouye, M. & Holt, K. E. FastSpar: rapid and scalable  
709 correlation estimation for compositional data. *Bioinformatics* **35**, 1064–1066  
710 (2019).
- 711 54. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and  
712 Projection for Dimension Reduction. Preprint at  
713 <https://doi.org/10.48550/arXiv.1802.03426> (2020).
- 714 55. Dixon, P. VEGAN, a package of R functions for community ecology. *Journal of*  
715 *Vegetation Science* **14**, 927–930 (2003).
- 716 56. Gates, A. J. & Ahn, Y.-Y. CluSim: a python package for calculating clustering  
717 similarity. *Journal of Open Source Software* **4**, 1264 (2019).

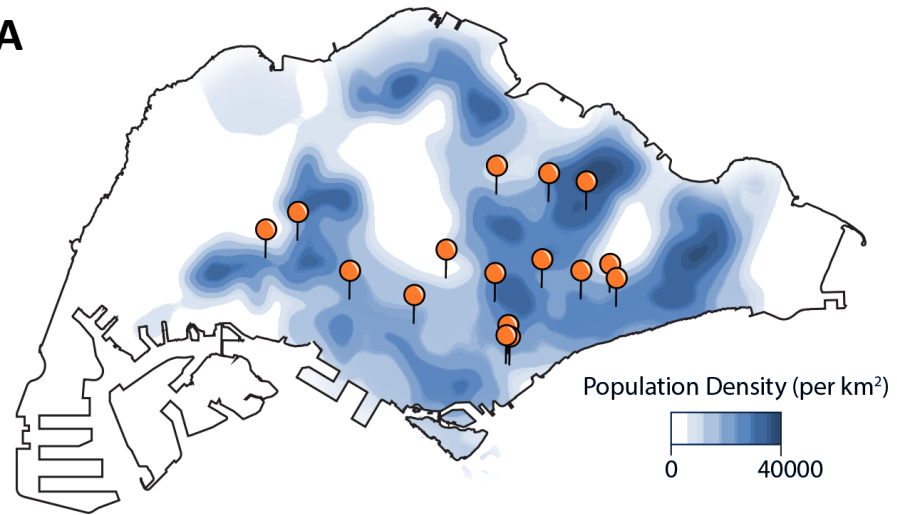
- 718 57. Paulson, J. N. *et al.* metagenomeSeq: Statistical analysis for sparse high-  
719 throughput sequencing. *Bioconductor package* (2013).
- 720 58. Inouye, M. *et al.* SRST2: Rapid genomic surveillance for public health and  
721 hospital microbiology labs. *Genome Medicine* **6**, 90 (2014).
- 722 59. Lam, M. M. C. *et al.* A genomic surveillance framework and genotyping tool for  
723 *Klebsiella pneumoniae* and its related species complex. *Nat Commun* **12**, 4188  
724 (2021).
- 725 60. Gupta, R. S., Lo, B. & Son, J. Phylogenomics and Comparative Genomic Studies  
726 Robustly Support Division of the Genus *Mycobacterium* into an Emended Genus  
727 *Mycobacterium* and Four Novel Genera. *Front Microbiol* **9**, 67 (2018).
- 728 61. Sichtig, H. *et al.* FDA-ARGOS is a database with public quality-controlled  
729 reference genomes for diagnostic use and regulatory science. *Nat Commun* **10**,  
730 3313 (2019).
- 731 62. de Melo Pereira, G. V. *et al.* What Is *Candida* Doing in My Food? A Review and  
732 Safety Alert on Its Use as Starter Cultures in Fermented Foods. *Microorganisms*  
733 **10**, 1855 (2022).
- 734 63. Sedláček, I. *et al.* *Aquitalea pelogenes* sp. nov., isolated from mineral peloid.  
735 *International Journal of Systematic and Evolutionary Microbiology* **66**, 962–967  
736 (2016).
- 737 64. Al-Obaid, K., Ahmad, S., Joseph, L. & Khan, Z. *Lodderomyces elongisporus*: a  
738 bloodstream pathogen of greater clinical significance. *New Microbes New Infect*  
739 **26**, 20–24 (2018).
- 740 65. Xie, C.-H. & Yokota, A. *Dyella japonica* gen. nov., sp. nov., a  $\gamma$ -proteobacterium  
741 isolated from soil. *International Journal of Systematic and Evolutionary*  
742 *Microbiology* **55**, 753–756 (2005).

- 743 66. Yun, J.-H., Park, S.-K., Lee, J.-Y., Jung, M.-J. & Bae, J.-W. *Kushneria konosiri*  
744 sp. nov., isolated from the Korean salt-fermented seafood Daemi-jeot.  
745 *International Journal of Systematic and Evolutionary Microbiology* **67**, 3576–3582  
746 (2017).
- 747 67. Zhao, G.-Y. *et al.* *Salinicola tamaricis* sp. nov., a heavy-metal-tolerant,  
748 endophytic bacterium isolated from the halophyte *Tamarix chinensis* Lour.  
749 *International Journal of Systematic and Evolutionary Microbiology* **67**, 1813–1819  
750 (2017).
- 751 68. Brandl, H., Gross, R. A., Lenz, R. W. & Fuller, R. C. *Pseudomonas oleovorans* as  
752 a Source of Poly( $\beta$ -Hydroxyalkanoates) for Potential Applications as  
753 Biodegradable Polyesters. *Applied and Environmental Microbiology* **54**, 1977–  
754 1982 (1988).
- 755 69. DREYFUS, B., GARCIA, J. L. & GILLIS, M. Characterization of *Azorhizobium*  
756 *caulinodans* gen. nov., sp. nov., a Stem-Nodulating Nitrogen-Fixing Bacterium  
757 Isolated from *Sesbania rostrata*. *International Journal of Systematic and*  
758 *Evolutionary Microbiology* **38**, 89–98 (1988).
- 759 70. Thanyacharoen, U., Tani, A. & Charoenpanich, J. Isolation and characterization  
760 of *Kluyvera georgiana* strain with the potential for acrylamide biodegradation.  
761 *Journal of Environmental Science and Health, Part A* **47**, 1491–1499 (2012).
- 762 71. Liu, W.-Y., Wong, C.-F., Chung, K. M.-K., Jiang, J.-W. & Leung, F. C.-C.  
763 Comparative Genome Analysis of *Enterobacter cloacae*. *PLOS ONE* **8**, e74487  
764 (2013).
- 765 72. Saravanan, C. & Shetty, P. K. H. Isolation and characterization of  
766 exopolysaccharide from *Leuconostoc lactis* KC117496 isolated from idli batter.  
767 *International Journal of Biological Macromolecules* **90**, 100–106 (2016).

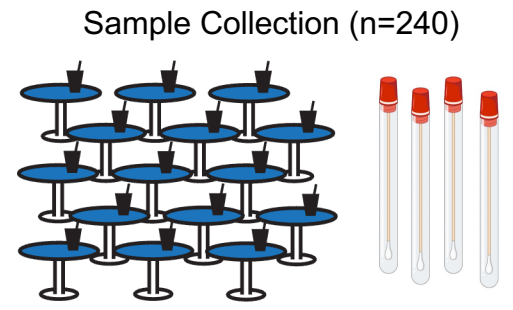
- 768 73. Widerström, M., Wiström, J., Sjöstedt, A. & Monsen, T. Coagulase-negative  
769 staphylococci: update on the molecular epidemiology and clinical presentation,  
770 with a focus on *Staphylococcus epidermidis* and *Staphylococcus saprophyticus*.  
771 *Eur J Clin Microbiol Infect Dis* **31**, 7–20 (2012).
- 772 74. Larsen, R. A., Wilson, M. M., Guss, A. M. & Metcalf, W. W. Genetic analysis of  
773 pigment biosynthesis in *Xanthobacter autotrophicus* Py2 using a new, highly  
774 efficient transposon mutagenesis system that is functional in a wide variety of  
775 bacteria. *Arch Microbiol* **178**, 193–201 (2002).
- 776 75. Zhao, F. *et al.* *Dyellajiangningensis* sp. nov., a  $\gamma$ -proteobacterium isolated from  
777 the surface of potassium-bearing rock. *International Journal of Systematic and*  
778 *Evolutionary Microbiology* **63**, 3154–3157 (2013).
- 779 76. Peng, X. *et al.* Efficient biodegradation of tetrabromobisphenol A by the novel  
780 strain *Enterobacter* sp. T2 with good environmental adaptation: Kinetics,  
781 pathways and genomic characteristics. *Journal of Hazardous Materials* **429**,  
782 128335 (2022).
- 783 77. Johansen, J. E., Binnerup, S. J., Kroer, N. & Mølbak, L. *Luteibacter rhizovicinus*  
784 gen. nov., sp. nov., a yellow-pigmented gammaproteobacterium isolated from the  
785 rhizosphere of barley (*Hordeum vulgare* L.). *International Journal of Systematic*  
786 *and Evolutionary Microbiology* **55**, 2285–2291 (2005).
- 787

**Figure 1**

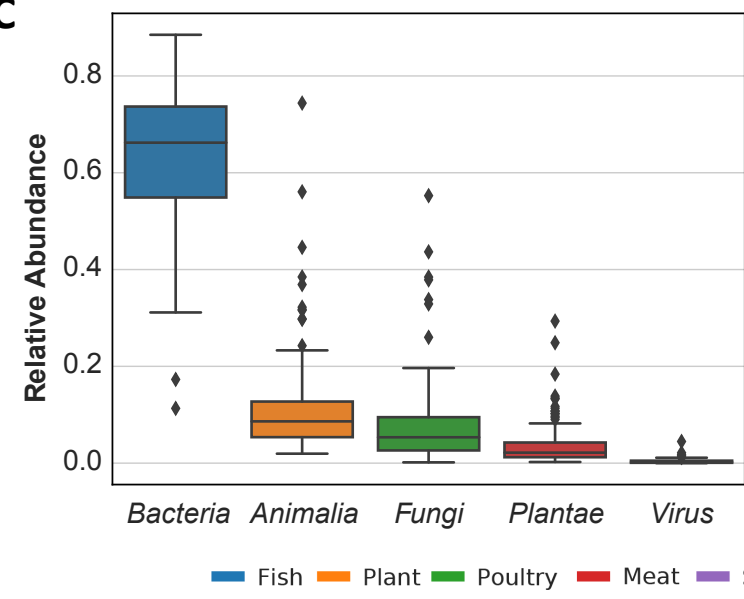
**A**



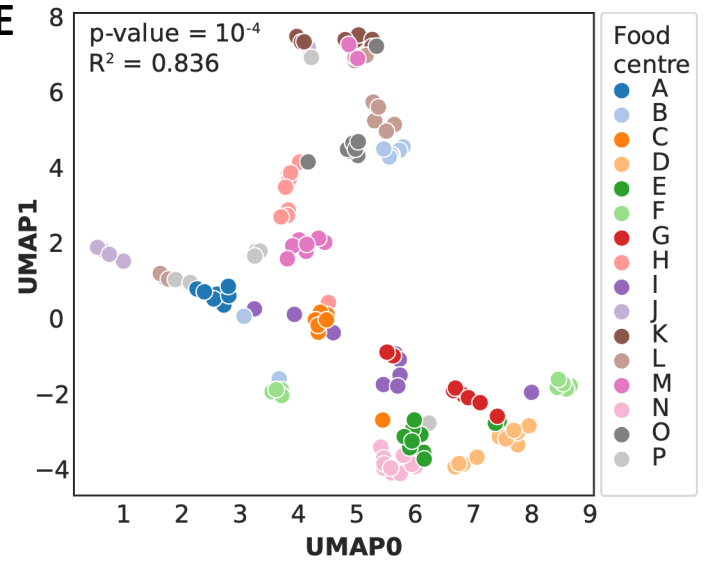
**B**



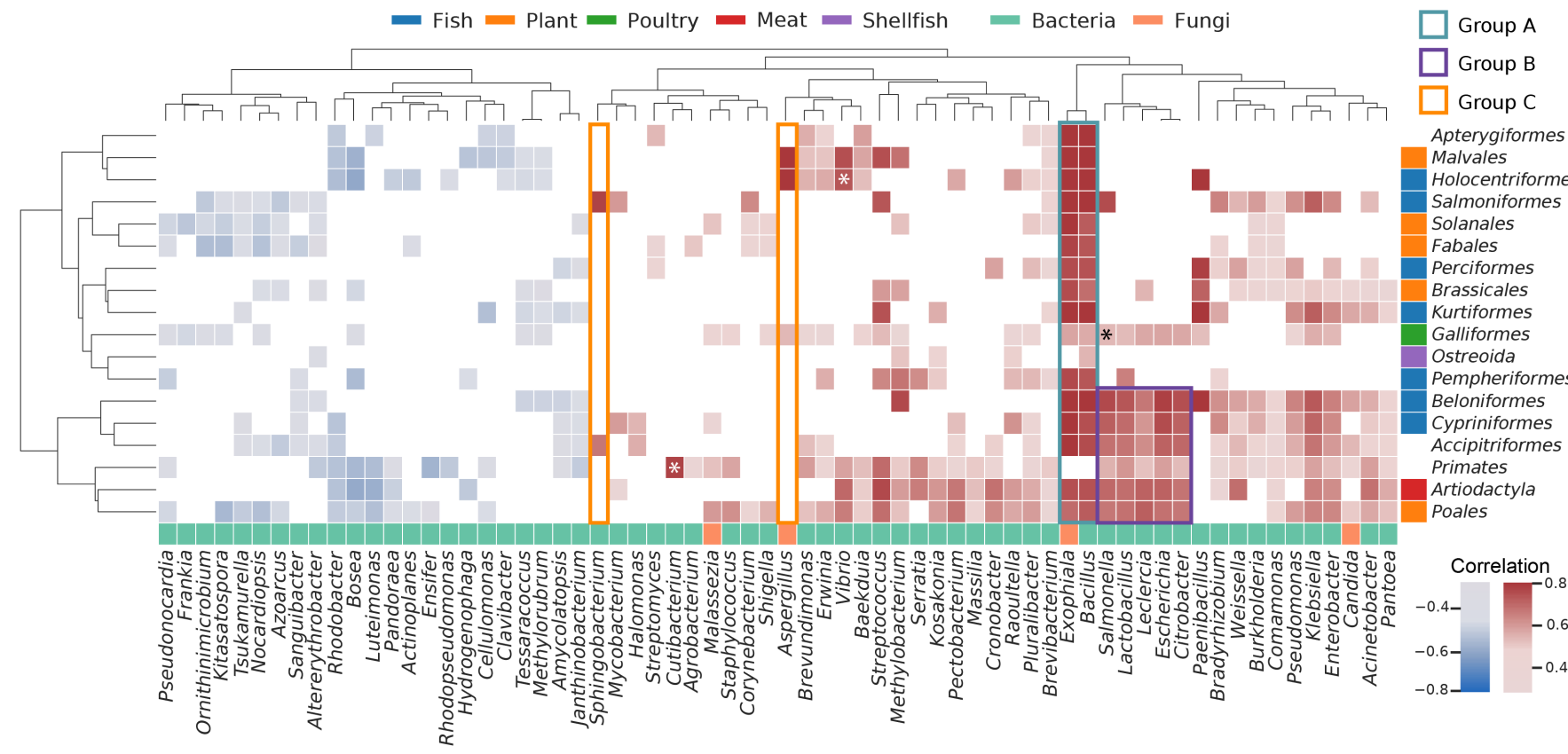
**C**

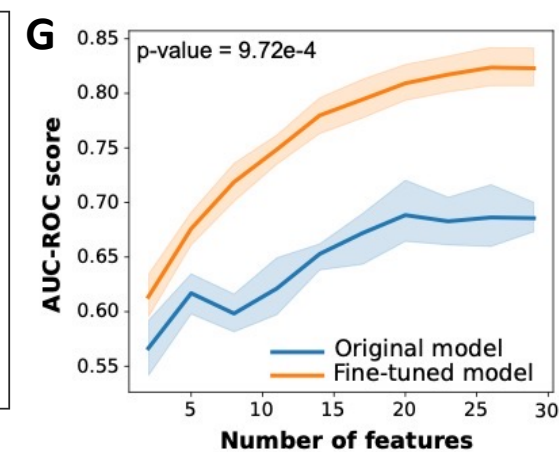
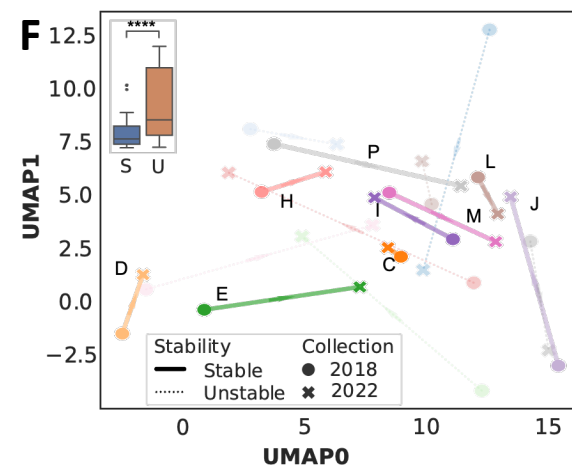
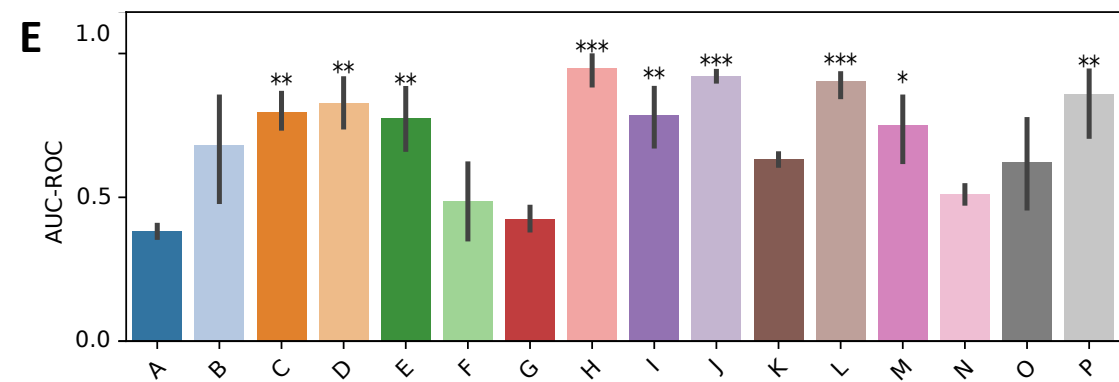
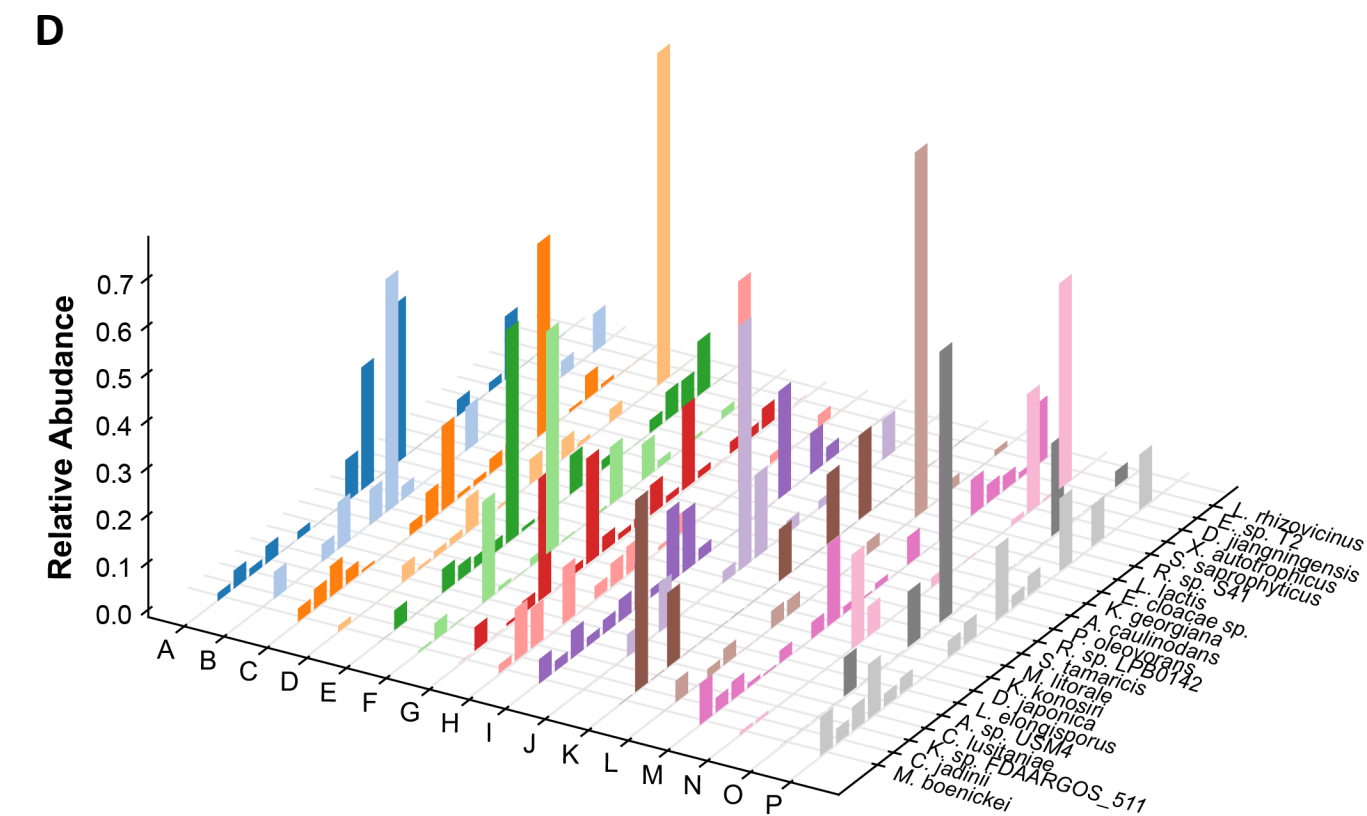
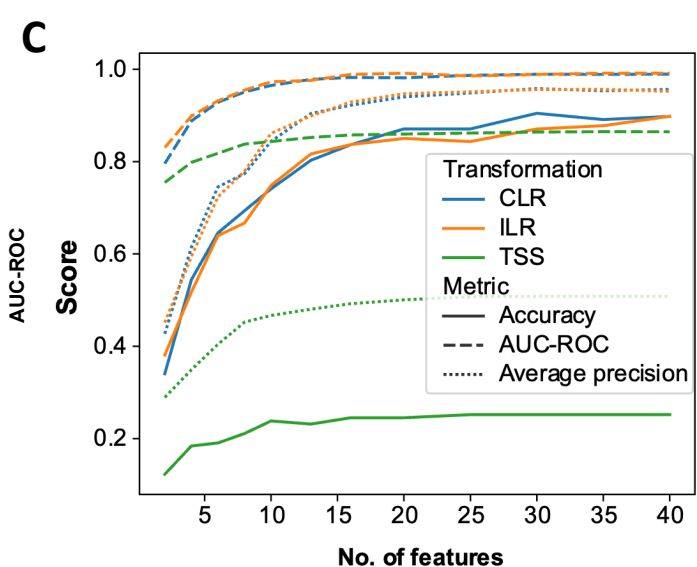
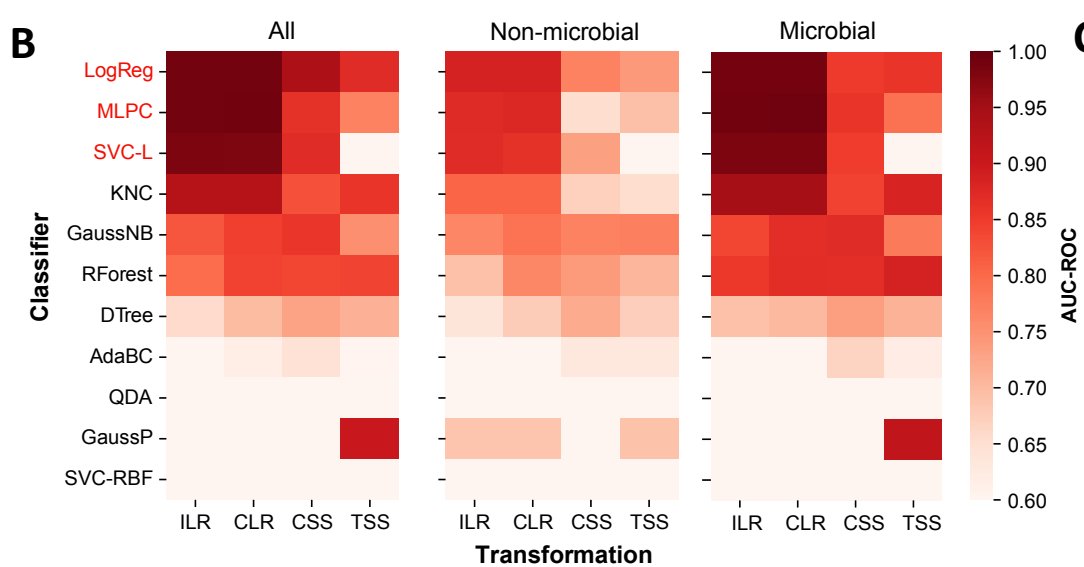
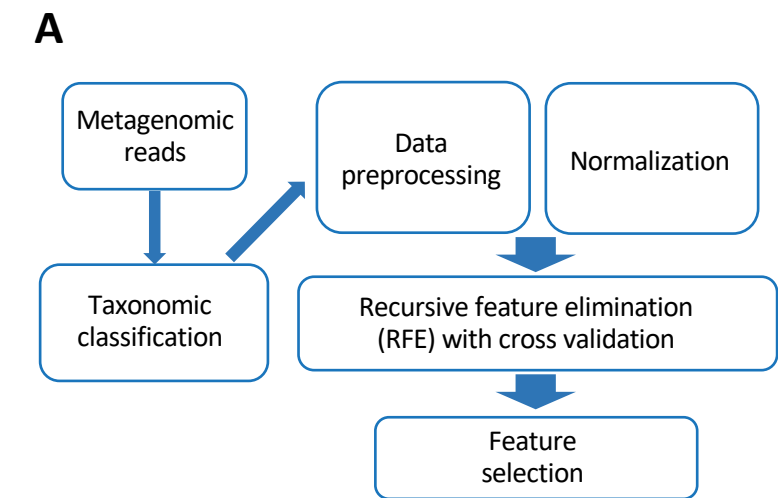


**E**



**D**

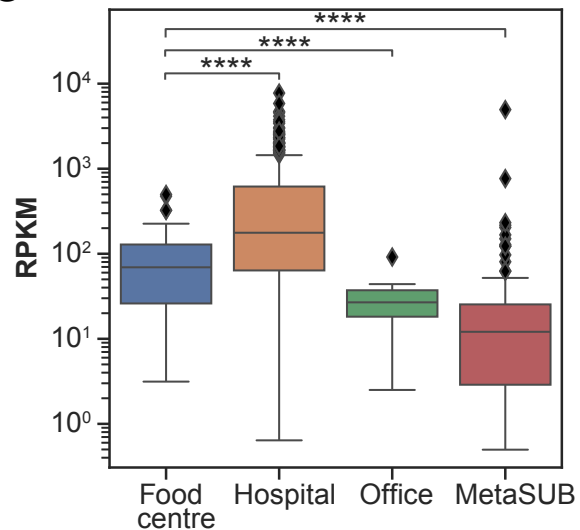


**Figure 2**

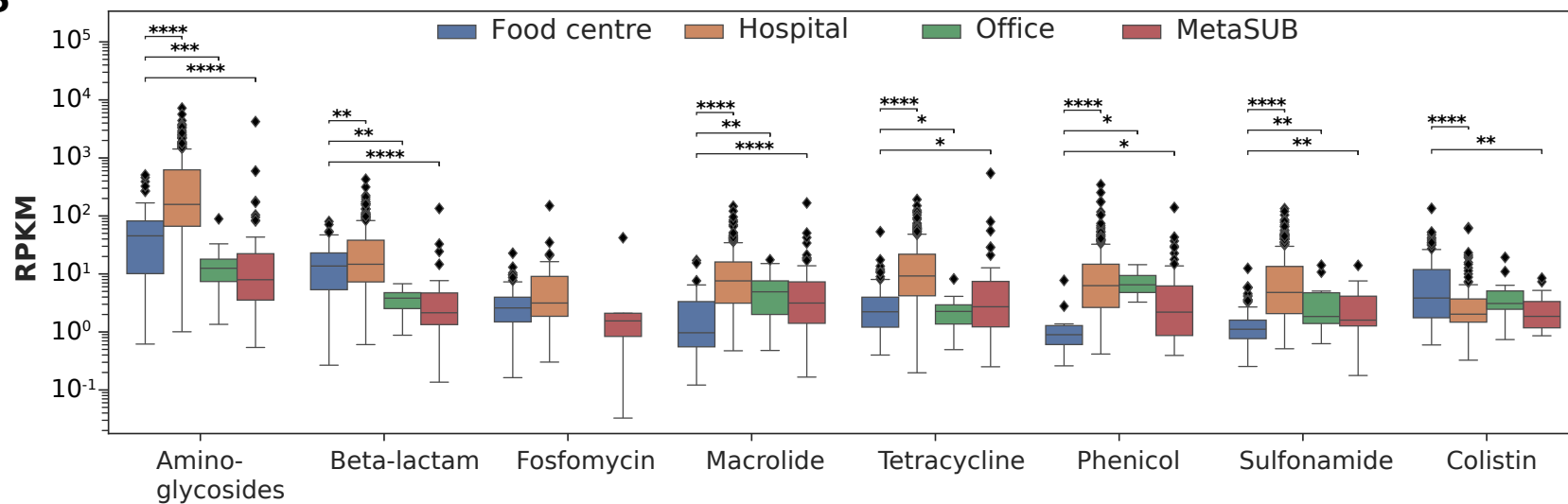


**Figure 3**

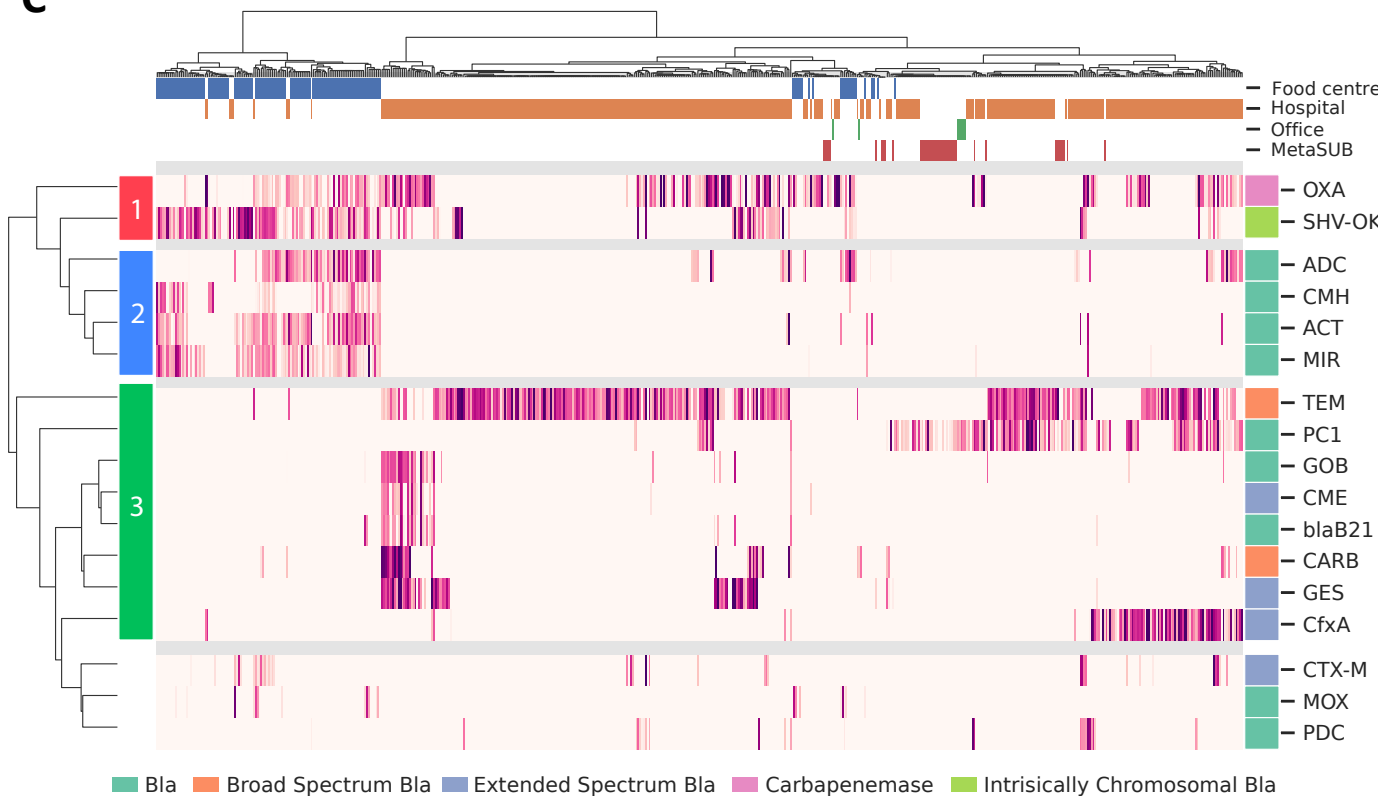
**A**



**B**



**C**



**D**

