

1 Population variation and prognostic potential of gut antibiotic 2 resistome

3 K. Pärnänen^{1*}, M. Ruuskanen^{1,2}, G. Sommeria-Klein¹, V. Laitinen¹, P. Kantanen¹, G.
4 Méric^{3,4,5,6}, C. Gazolla Volpiano^{3,4}, M. Inouye^{3,4,7,8}, R. Knight⁹, V. Salomaa², A. S.
5 Havulinna^{2,10}, T. Niiranen^{2,11}, L. Lahti^{1*}.

6
7 ¹University of Turku, Department of Computing - Turku (Finland), ²Finnish Institute for Health and
8 Welfare, THL - Helsinki (Finland), ³Cambridge Baker Systems Genomics Initiative, Baker Heart and
9 Diabetes Institute, Melbourne, Victoria, (Australia), ⁴Department of Cardiometabolic Health, University of
10 Melbourne, Melbourne, Victoria (Australia), ⁵Central Clinical School, Monash University, Melbourne,
11 Victoria (Australia), ⁶Department of Cardiovascular Research, Translation and Implementation, La Trobe
12 University, Melbourne, Victoria (Australia), ⁷Cambridge Baker Systems Genomics Initiative, Department
13 of Public Health & Primary Care, University of Cambridge, Cambridge (United Kingdom), ⁸BHF
14 Cardiovascular Epidemiology Unit, Department of Public Health & Primary Care, University of Cambridge,
15 Cambridge (United Kingdom), ⁹University of California, San Diego - San Diego (United States), ¹⁰Institute
16 for Molecular Medicine Finland, FIMM-HiLIFE - Helsinki (Finland), ¹¹Department of Internal Medicine,
17 Turku University Hospital and University of Turku, Turku

18

19 * corresponding authors:

20 Katariina Pärnänen, Leo Lahti, katariina.parnanen@utu.fi, leo.lahti@utu.fi

21 **SUMMARY**

22 The spread of antibiotic-resistance genes in bacteria has severely reduced the efficacy
23 of antibiotics, now contributing to 1.3 million deaths annually. Despite the far-reaching
24 epidemiological implications of this trend, the extent to which antimicrobial resistance
25 load varies within human populations and the drivers that contribute most to this
26 variation remain unclear. Here, we demonstrate in a representative cohort of 7,095
27 Finnish adults¹ that socio-demographic factors, lifestyle, and gut microbial community
28 composition shape resistance selection and transmission processes. Antimicrobial
29 resistance gene load was linked not only to prior use of antibiotics, as anticipated, but
30 also to frequent consumption of fresh vegetables and poultry, two food groups
31 previously reported to contain antibiotic-resistant bacteria. Interestingly, ARG load was
32 not associated with high-fat and -sugar foods. Furthermore, antimicrobial resistance
33 gene load was systematically higher in females and the generally healthier high-income
34 demographics in urban and densely populated areas. Data from this prospective cohort
35 with a 17-year follow-up suggests that the prognostic potential of antimicrobial
36 resistome is comparable to blood pressure for mortality and sepsis. These findings
37 highlight population-level risks and socio-demographic dimensions of antimicrobial
38 resistance that are particularly relevant in the current context of global urbanization and
39 middle-class growth.

40

41 INTRODUCTION

42 Antibiotic-resistant bacteria (ARB) pose a rapidly increasing threat to global health due
43 to failures in treating infections. By 2050, antimicrobial resistance (AMR, including
44 antibiotic, antifungal, antiparasitic, and antiviral resistance) is predicted to contribute to
45 10 million deaths annually, surpassing the current most common causes of death, such
46 as cancer and cardiovascular disease². The resistance is encoded by multiple classes
47 of antibiotic resistance genes (ARGs) that can spread across bacterial and human
48 populations. In particular, the human gut has been established as a major reservoir of
49 ARGs and a highway for their lateral transfer^{3,4}. The entire collection of ARGs in a given
50 habitat, such as the gut, is called the *resistome*^{5,6}.

51

52 The emergence and spread of antibiotic resistance in microbial communities^{3,7,8} are
53 driven by the ecological processes of *selection* and *transmission* (Figure 1), which act
54 on both the resistance genes and the bacteria that carry them. *Direct selection* of genes
55 that confer antibiotic resistance can be caused, for instance, by antibiotic use, as it
56 favors the selection of resistant strains and, thus, ARG proliferation in the human gut
57 resistome. *Indirect selection* of resistant bacteria can be caused by environmental
58 factors that influence the growth of bacterial communities with specific types of
59 resistomes⁹. Finally, *transmission* of resistance genes or resistant bacteria between
60 individuals has been suggested to be a key driver of clinical resistance¹⁰. In particular,
61 previous studies have indicated that contaminated food, person-to-person contact, and
62 travel are possible risk factors for acquiring ARGs¹¹⁻¹³.

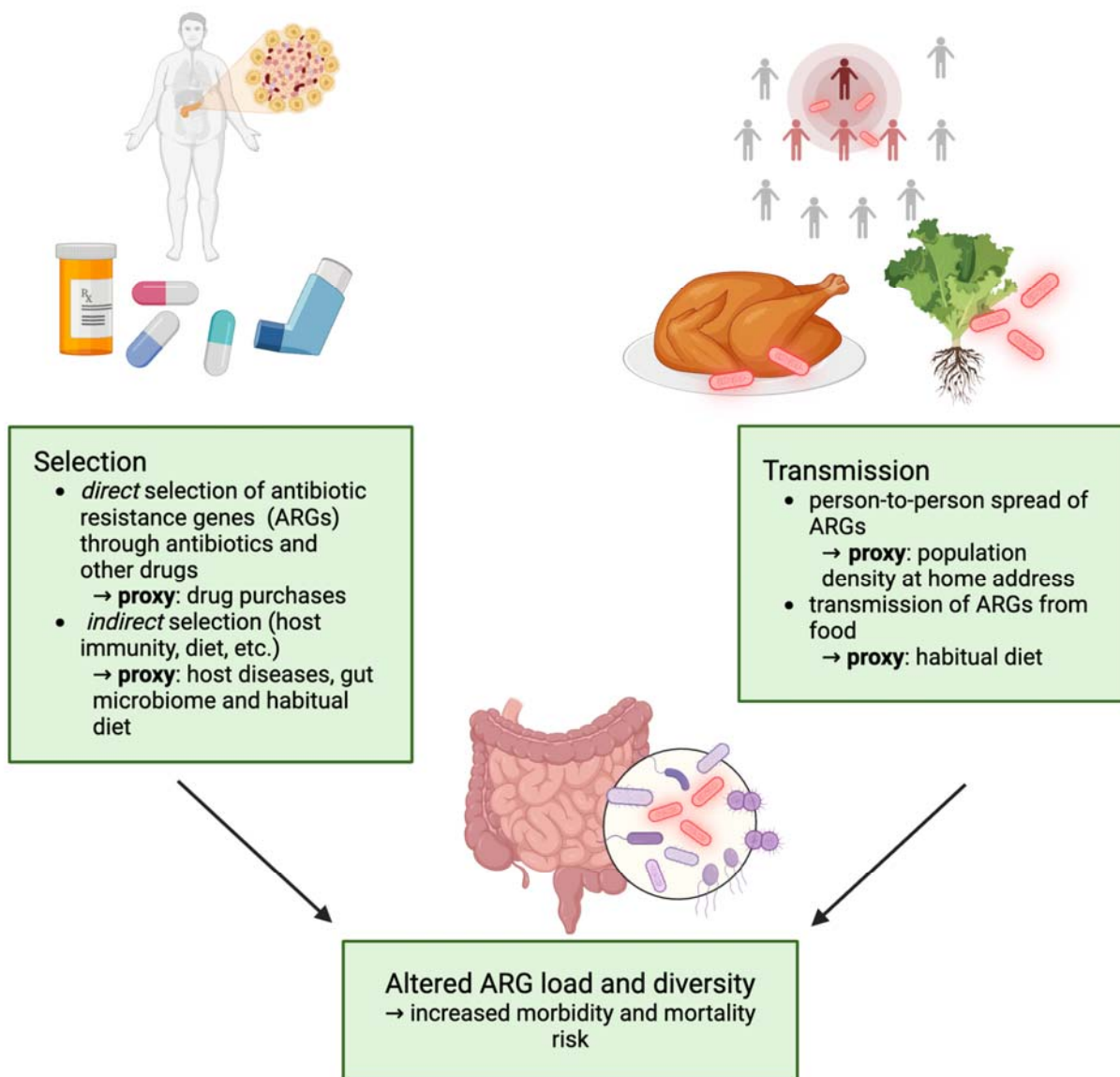
63

64 The World Health Organization has called for research on the demographic parameters
65 underlying the AMR emergence and spread in human populations¹⁴ to tackle the AMR
66 crisis effectively. Previous studies based on country-level statistics have shown that
67 antibiotic resistance varies with socioeconomic markers^{10,15,16} and antibiotic use^{6,17}.
68 However, few studies have analyzed participant-level variation and the role of
69 demographic and lifestyle factors in shaping the resistome within specific populations.
70 Moreover, the associations between population-level resistome variation and long-term
71 mortality and morbidity risk remain largely uncharacterized.

72

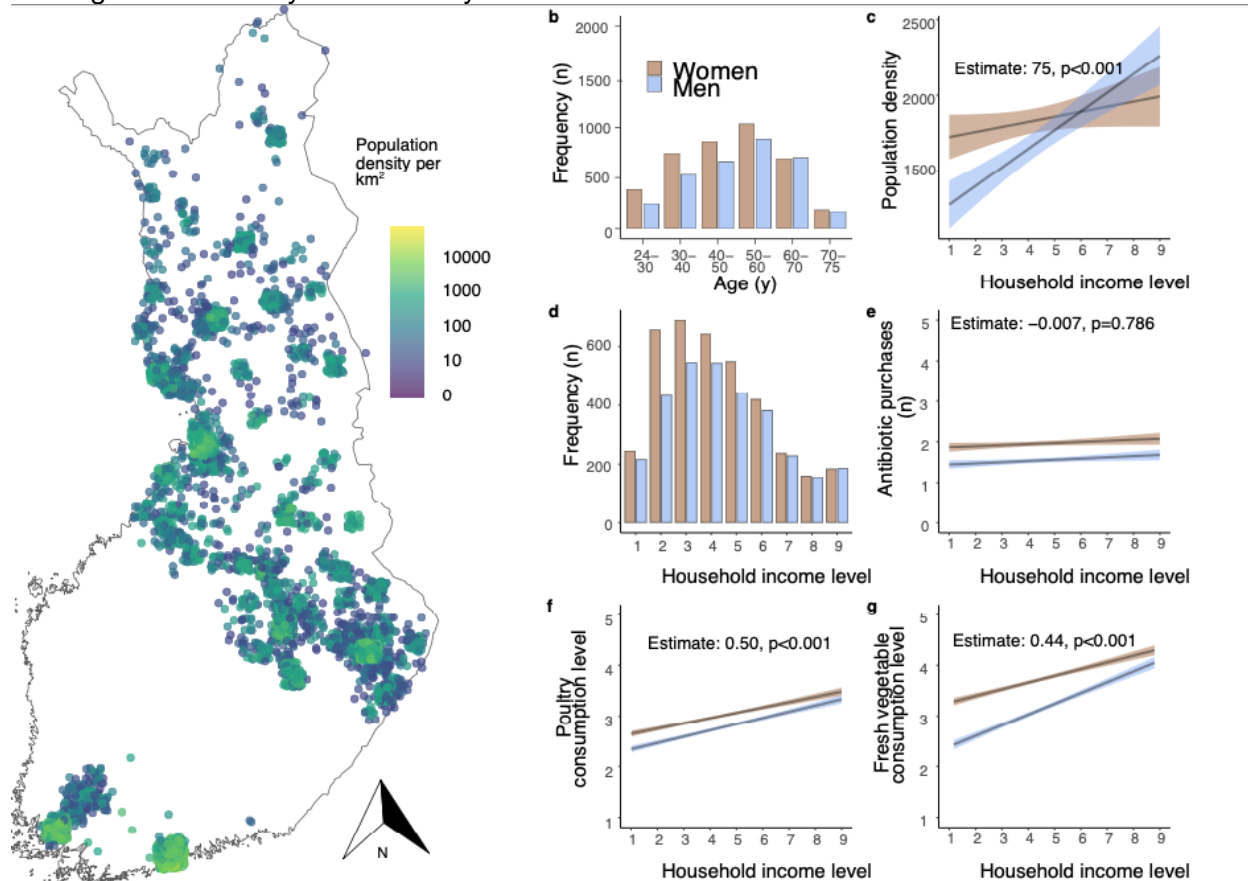
73 We investigated factors associated with gut resistome variation, as well as prospective
74 mortality and morbidity risk associated with ARG load in a well-characterized Finnish
75 population cohort of 7,095 adults across six regions of Finland with varying
76 demographics (mean age 49, 55% women). The cohort represents the general, non-
77 institutionalized population without acute infections (Figure 2; FINRISK¹). Based on
78 metagenomic profiles from fecal samples collected in 2002, we studied gut resistome
79 composition and diversity, the total ARG load, and the factors contributing to their
80 observed variation. We defined the total ARG load as the cumulative relative
81 abundance of all observed ARGs. Data on address-level geographic location, diet,
82 household income level, prescription drug purchases, diseases, and causes of death
83 until 2019 were gathered from electronic population registers, health examinations, and
84 complementary questionnaires (see [Methods](#) for a description of data sources and
85 covariate selection including analysis of collinearity in key covariates). Summary

86 statistics for all antibiotics purchases during the seven years before sampling are
87 provided in [Extended Data Table 1](#). We combined these data to investigate how
88 geographic and demographic factors, lifestyle, and health shape the antibiotic
89 resistome, and what the taxonomic underpinnings of these associations are. Moreover,
90 we investigated the prognostic value of the resistome for mortality due to infectious or
91 other causes. These analyses shed light on the putative ecological and epidemiological
92 patterns of antibiotic resistance at the population level and their health implications.
93
94
95

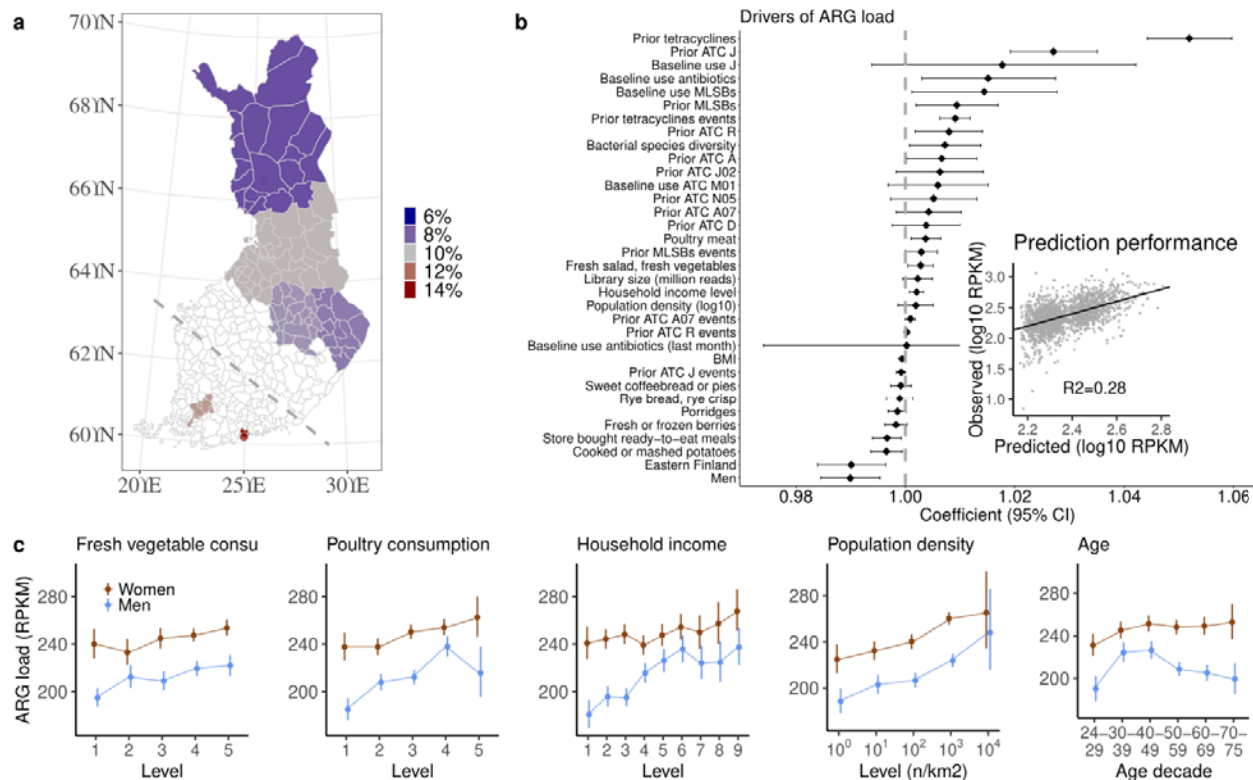


96
97
98 **Figure 1: Ecological and epidemiological framework for population-level variation in ARG**
99 **load and resistome** The ecological processes underlying ARG variation include 1) *selection of*

100 antimicrobial resistance genes (ARGs) and bacteria that carry them, for instance as a *direct*
 101 consequence of antibiotic consumption or *indirectly* due to environmental selection or host
 102 factors (e.g., prevalent disease) that impact microbial community composition; 2) *external*
 103 *transmission* via antimicrobial resistance gene carrying bacteria (ARBs). We associated
 104 resistome diversity, composition, and the overall ARG load with various socio-demographic
 105 parameters - some of which can be used as proxies for the above ecological processes - as well
 106 as long-term morbidity and mortality risk.



107
 108
 109 **Figure 2: Overview of the FINRISK cohort (N=7,095)** a Geographical distribution of the cohort
 110 participants. The jittered data points indicate the population density per km². Some individual
 111 points at remote locations were removed to obscure participants' addresses. b Age distribution
 112 by decade, shown for males (M; blue) and females (F; brown) c Population-density versus
 113 household income level (scale 1-9). d Household income level (scale 1-9) e Household income
 114 level versus antibiotic purchases during the seven years before the sample collection. f
 115 Household income level versus poultry consumption (scale 1-5, from less than once a month to
 116 multiple times per day), and g fresh vegetable and salad consumption. Panels c, e, f, and g
 117 include linear models' estimated slope and p-value. See [Methods](#) for a more detailed
 118 description of the scales.



119

120 **Figure 3: ARG load prediction and variation.** **a** Regional variation in total ARG load across
 121 the six regions in the FINRISK cohort (North Karelia and North Savonia in the east, Oulu region
 122 and Lapland in the north, Turku region in the southwest and Helsinki capital region in the south).
 123 The color indicates the fraction of the population in the high ARG group for each region (>458
 124 RPKM; the sum of all normalized ARGs per kilobase per million reads, i.e., the top-10% quantile
 125 in this cohort; [Extended Data Table 3](#)). The dashed line illustrates the East-West split (the Turku
 126 and Helsinki regions in the West and the other four regions in the East). **b** Drivers of ARG load
 127 (boosted GLM effect size; see [Methods](#)). For drivers of resistome diversity, see [Figure S1a](#). The
 128 line plot shows each predictor's effect sizes and 95% confidence interval. The inset shows the
 129 predicted and observed ARG load in test data ($R^2 = 0.28$). *Prior X*: purchase of drug X in the
 130 seven years before sampling (yes/no), *Prior X events*: number of purchases of drug X in the
 131 seven years before sampling, *Baseline X*: Purchase of drug X in the six months before sampling
 132 (yes/no). ATC class abbreviations: A = Alimentary tract and metabolism, A07 = Antidiarrheals,
 133 intestinal anti-inflammatory/anti-infective agents, D = Dermatologicals, J = Anti-infectives, J02 =
 134 Antifungals, M01 = Anti-inflammatory and antirheumatic products, N05 = Psycholeptics, R =
 135 Respiratory system. **c** Association between ARG load and fresh vegetables and poultry
 136 consumption, household income, population density, and age (N=7,095: see [Methods](#)
 137 for variable descriptions and [Supplementary Table 3](#) for the numerical estimates). Similar trends
 138 can be observed within individual regions ([Figure S2](#)).

139

140 RESULTS

141 Predictors of ARG load and diversity

142 The ability to predict ARG load and other resistome features could help identify
143 individuals at risk and derive population-wide estimates of resistome variation (Figure
144 3). We trained a supervised machine learning model (boosted GLM; [Methods](#)) to predict
145 ARG load from gut microbiome composition and demographic, health, and lifestyle
146 factors. Through cross-validation on left-out (test) data, we identified key predictors of
147 ARG load and quantified their ability to provide generalizable predictions on new
148 individuals (Figure 3b).

149 The prediction model could explain 28% and 21% of the variation in ARG load and
150 resistome diversity, respectively, in the independent test set (Figure 3b; [Figure S1a](#)).
151 Key predictors for both included antibiotic and drug use, diet, gender, household
152 income, East/West geographic division, population density, and bacterial species
153 diversity (see also Figure 3c). Similar, albeit sometimes non-significant trends could be
154 observed independently within each of the six geographical regions ([Figure S2](#),
155 [Extended Data Tables 2-3](#)). Including bacterial families as predictors increased the
156 explained variance to 32% in the test set (boosted GLM, [Figure S1b](#)), suggesting that
157 microbial community composition contributes to the ARG load on top of the other
158 covariates. Antibiotic use explained the largest proportion of variance (GLM with just the
159 respective covariates, 27%) in ARG load in the test data, followed by prevalent bacterial
160 families (3%), demographic variables (household income and gender; 2%), geography
161 (East/West; 1%) and diet (1%). Resistome diversity was primarily explained by species
162 diversity (11%) and antibiotic use (6%), followed by demographics (gender and
163 household income) (3%), geography (1%), population density (1%), and diet (1%)
164 (results on ARG diversity are shown in [Figure S1a](#)). Antibiotic use, diet, population
165 density, geography, and demographic factors represent proxies for the selection and
166 transmission of antibiotic resistance genes (see Figure 1). Our data confirms they have
167 substantial predictive value on individual ARG load and diversity.

168

169 Drug use

170 Antibiotic use promotes the selection of antibiotic resistance. The majority of variation in
171 ARG load was explained by antibiotic use alone. Prior use of several classes of
172 antibiotic and non-antibiotic therapeutics listed in the Anatomical Therapeutic Chemical
173 Classification System (ATC) during the past seven years was positively associated with
174 ARG load ($p < 0.05$, [Supplementary Table 1](#)). Consumption of antimicrobials (ATC class
175 J) was associated with a 55% higher ARG load ($p < 0.001$) compared to no
176 antimicrobial exposure in the past seven years.

177 In particular, prior tetracycline (ATC class J01) purchases were associated with a 67%
178 increase in ARG load ($p < 0.001$), and MLSB antibiotics with a 39% increase (ATC class
179 J01F; $p < 0.001$). These associations remained robust after controlling for other
180 participant-level data, such as diet (Figure 3b). Tetracycline was the second most

181 frequently purchased antibiotic class in the cohort ([Extended Data Table 1](#)) and has
182 been widely used since the 1950s. It is also very commonly used in animal production¹⁸.
183 Moreover, purchases of respiratory medication (ATC class R) were associated with a
184 higher ARG load after controlling for past antibiotic use, demographic factors, and other
185 participant-level data (Figure 3b). Other associations between non-antibiotic drugs and
186 ARG load were not robust to adjusting for other covariates or for abundances of
187 bacterial families ([Figure S1](#)). Although some of these drugs might have selective
188 effects favoring resistance, drug use can also be a proxy for additional, unobserved
189 covariates. These observations highlight the need to consider the long-term impact of
190 the use of antibiotics.

191 **Diet**

192
193 Food could directly contribute to antibiotic resistance by transmitting ARGs to the gut,
194 for instance, by consuming fresh produce and meat^{13,19}. Another way food could
195 influence antibiotic resistance is through indirect selection, that is, by promoting or
196 inhibiting the growth of ARB. For instance, it has been hypothesized that high-fat, high-
197 sugar, and processed foods could promote the growth of Proteobacteria²⁰ or other
198 bacterial taxa that tend to carry more ARGs. Conversely, high-fiber foods could select
199 against ARB²¹. Systematic population-level data associating specific foods and
200 antibiotic resistance is limited^{21,22}, and the relative importance of selection and
201 transmission in food-mediated resistance is yet to be characterized.

202 We associated ARG load and diversity with the self-declared habitual consumption
203 frequency of 42 food groups ([Methods](#)). Poultry had the strongest positive association
204 with ARG load (4% average increase per consumption level, adjusted for antibiotics; $p <$
205 0.001 ; [Supplementary Table 1](#)), followed by fresh vegetables and salad (3% increase
206 per level; $p < 0.001$). In contrast to the other dietary components, the associations
207 remained robust for poultry and fresh vegetables even after controlling for other
208 covariates, including socio-demographics and antibiotic use ($p < 0.05$ Figure 3b-c).
209 Moreover, their association with ARG load remained significant also when controlling for
210 the abundance of bacterial families ([Figure S1](#), $p < 0.05$), which suggests that this effect
211 is not due to indirect selection. These results support the hypothesis that these foods
212 contain ARBs that can transfer to the gut.^{23,24} We did not find a similar association for
213 beef or pork, which is in line with the fact that other meat production animals in Finland
214 were reported to have much less antibiotic resistance than poultry at the time of
215 sampling²⁵.

216 In contrast, highly processed, high-sugar, and high-fat foods did not exhibit robust
217 associations with ARG load and diversity. We observed weak associations with ARG
218 load for dietary components such as chocolate and fast food. However, they were not
219 robust to adjusting for participants' covariates (1% increase per level, $p = 0.05$ - 0.15 ;
220 Figure 3b). Nevertheless, the link between these foods and ARG load and diversity may
221 be partially confounded by antibiotic exposure and diseases, which covary with diet and
222 likely influence the abundance of ARG-containing taxa proposed to respond to high-fat

223 diets, such as Proteobacteria²⁰. Interestingly, some of these foods were, in fact,
224 inversely associated with ARG load, as were also total cholesterol and BMI ($p < 0.05$).
225 This might be partly explained by the lack of fresh vegetables and poultry in the typical
226 diets containing these foods. Associations with high-fiber foods were mixed: both fresh
227 and cooked vegetables were positively associated with ARG load, whereas berries and
228 rye bread had an inverse association ($p < 0.05$). Taken together, our findings suggest
229 that the influence of population-level dietary variation on antibiotic resistance is
230 dominated by transmission from food rather than indirect selection by diet.

231

232 **Population density and urban regions**

233

234 Geography is a key epidemiological parameter that co-varies with many genetic,
235 demographic, and lifestyle factors that can contribute to the emergence and spread of
236 antibiotic resistance. In Finland, the Eastern and Western populations have well-
237 characterized genetic and lifestyle differences²⁶. Whereas earlier large-scale studies
238 have consistently reported geographic variation in antibiotic resistance between
239 countries^{5,15,17}, these results have been largely limited to country-level aggregates and
240 thus lack the participant-level resolution that would be necessary for detailed
241 quantification of resistome variation. We used the available participant-level data from
242 national population registers to derive high-resolution estimates of population density
243 based on the participants' home addresses. This enabled a high-resolution analysis of
244 geographic variation.

245

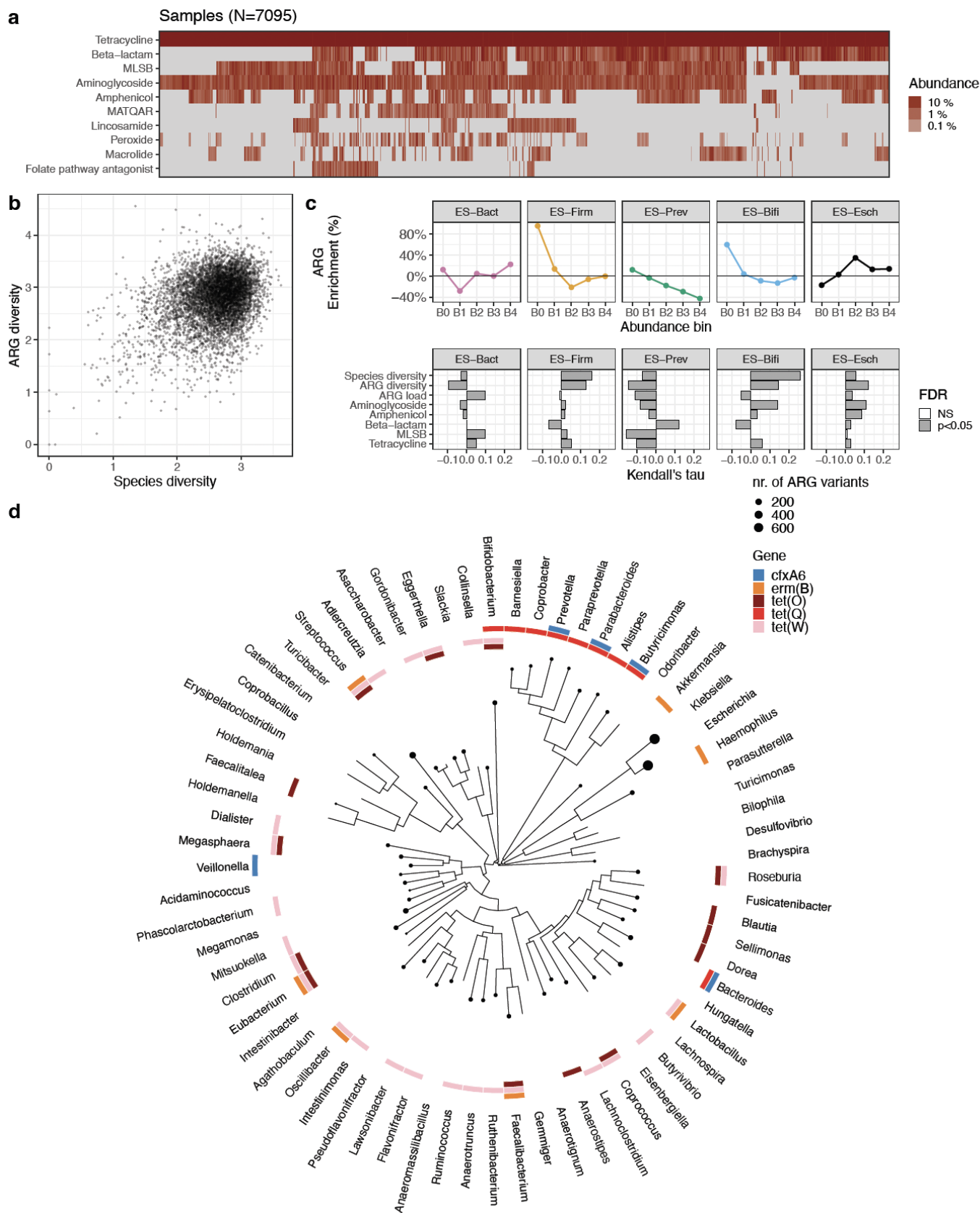
246 Differences in resistome composition between geographic regions were significant in
247 most comparisons ([Extended Data Table 4](#), PERMANOVA, $p=0.0015$), albeit small
248 (explained variance $<0.1\%$; $p<0.002$). Eastern Finns had a generally lower ARG load
249 and diversity even after controlling for (family-level) microbiota composition diet, health,
250 population density, and demographic factors (boosted GLM, $p > 0.05$, [Figure 3b](#); [Figure](#)
251 [S1](#)). Thus, the variation between East and West could be partly due to other
252 differences, such as limited spatial dispersal, lifestyle covariates, and differences in the
253 genetic background. Urban regions were enriched in individuals with a high ARG load
254 (top-10% quantile; >458 RPKM. This trend was particularly notable around the two
255 urban centers: the Helsinki capital area and Turku, the third-largest urban region in
256 Finland ([Figure 3a](#); [Extended Data Table 3](#)). The highest regional ARG load and
257 diversity were observed in these two cities. In contrast, the lowest median ARG load
258 was observed in Lapland, a rural region with a remarkably low population density.
259 Compared to Lapland, the median ARG load was 20% higher in Helsinki and 13%
260 higher in Turku. The increase in the prevalence of high-ARG individuals was even
261 higher (84% and 47%, respectively), suggesting that moderate increases in overall ARG
262 load in urban regions increase the risk of acquiring a high ARG load. This could be
263 explained by increasing ARG transmission in densely populated areas¹², a hypothesis
264 supported by the observation that ARG load and diversity increased also more generally
265 with population density ([Figure 3c](#); [Supplementary Table 1](#)).

266

267 **Demographics**

268

269 Next, we investigated the associations between ARG load and specific socio-
270 demographic factors (Figure 3, [Supplementary Table 1](#)). Men had 92% lower average
271 ARG load than women (95% CI 0.89-0.94, $p < 0.001$; log-linear model); the same trend
272 could be observed across all six regions ([Figure S2](#)). Women tend to purchase more
273 antibiotics (41% more purchases on average in a linear model adjusted for age;
274 $p < 0.001$) and consume raw vegetables more frequently than men (52%; $p < 0.001$),
275 regardless of household income. Nevertheless, the difference between genders
276 remained significant even after controlling for other covariates, including antibiotics, diet,
277 and the relative abundance of bacterial families ($p < 0.05$, Figure 3b, [Figure S1](#)).
278 Potential causes for this result include differences in occupation or caretaking,
279 prevalence of urinary tract infections¹⁴, and propensity to seek medical care²⁷.
280 Household income was positively associated with ARG load, with a 2% average
281 increase in ARG load per income level ($p < 0.001$). This association was again robust to
282 differences in antibiotic use and other participant-level covariates, including abundance
283 variations in bacterial families. This finding contrasts with the generally positive
284 association between higher socio-economic status and health²⁸, and could be partly
285 explained by lifestyle factors associated with a higher income: these participants
286 consumed more raw vegetables, lived in more densely populated areas (Figure 2,
287 [Extended Data Table 1](#)), and are generally more engaged in international travel²⁹.

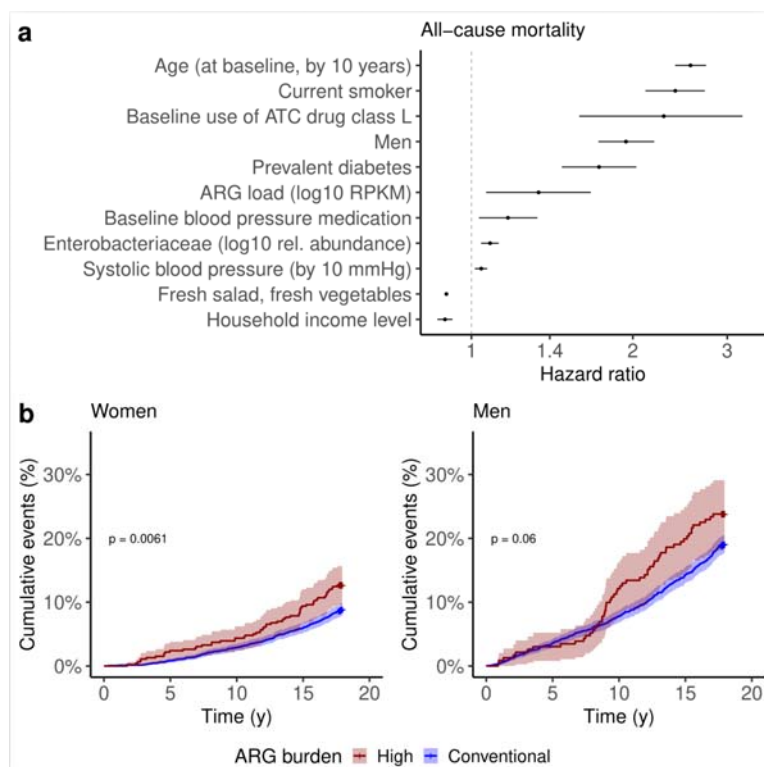


289 **Figure 4. Resistome variation and taxonomic composition.** **a** Resistome composition by the
290 relative abundance for the 10 most abundant ARG classes in the FINRISK cohort. **b** Bacterial
291 species diversity and resistome diversity (Shannon index; Pearson $r=0.32$). **c** *Top panel:* High
292 ARG load (>458 RPKM) enrichment by enterosignature (ES) abundance (B0: not detected; B1-
293 B4: 25% abundance quartiles); each ES represents a sub-community of co-varying genera
294 ([Figure S6](#)). *Lower panel* indicates associations between each ES, bacterial species diversity,
295 resistome diversity, total ARG load, and the five most dominant ARG classes (Kendall's Tau;
296 [Supplementary Table 2](#)). The gray bars indicate significant associations ($p<0.05$).
297 Abbreviations: MLSB: "Macrolide, Lincosamide, Streptogramin B"; MATQAR: "Macrolide,
298 Aminoglycoside, Tetracycline, Quinolone, Amphenicol, Rifamycin". **d** Phylogenetic relatedness
299 among the most prevalent genera in the FINRISK cohort and their association with the most
300 prevalent ARGs. Node size indicates the total number of ARGs found in each genus according
301 to the BLAST *nr* database; the colors indicate a match between each genus and the ARGs.

302
303 **Resistome composition reflects bacterial phylogeny**

304 Tetracycline resistance was found among virtually all participants (prevalence 100%),
305 whereas the prevalence of other ARG classes was more variable ([Figure 4a](#); [Figure S4](#)).
306 Resistome composition and diversity tend to co-vary with bacterial phylogeny⁹ as
307 bacterial taxa differ in their tendency to harbor ARGs^{3,30}. In our data, bacterial species
308 diversity partly explained resistome diversity ($R^2=0.11$; [Figure 4b](#); [Extended Data Fig](#)
309 [1a](#)), although association with ARG load remained weak ($R^2=0.001$, [Figure 3b](#)). We
310 observed phylogenetic clustering among some of the most prevalent ARGs ([Figure 4d](#)).
311 The tetracycline resistance genes covered a broad phylogenetic range, which could
312 partly explain their high prevalence. Macrolide resistance gene *ermB* was observed in a
313 more limited yet phylogenetically disjoint set of taxa (e.g., in *Bifidobacterium*,
314 *Clostridium*, and *Faecalibacterium*). *Escherichia* and *Klebsiella* carried the largest
315 number of unique ARGs. Abundant ARGs were often found in commensal taxa: for
316 instance, beta-lactam resistance gene *cfxA6* was detected in *Bacteroides*,
317 *Parabacteroides*, *Veillonella*, and *Prevotella* (BLAST *nr* database^{31,32}, [Figure 4d](#)). To
318 further examine co-variation between resistome and microbiota composition, we
319 identified five sub-communities, or *enterosignatures* (ES)³³ of co-varying bacterial
320 genera ([Figure S6](#)), which collectively explained 82% of all genus-level variation. They
321 were respectively dominated by members of *Bacteroides* (ES-*Bact*), Firmicutes (ES-
322 *Firm*), *Prevotella* (ES-*Prev*), *Bifidobacterium* (ES-*Bifi*), and *Escherichia* (ES-*Esch*). Each
323 ES was associated with a different characteristic resistome profile ([Figure 4c](#), [Extended](#)
324 [Data Table 5](#)). For instance, ES-*Prev* was associated with high beta-lactam and low
325 tetracycline resistance gene abundances. High ES-*Bact* and ES-*Esch* and low ES-*Prev*
326 and ES-*Bifi* were associated with an increased ARG load. The analysis suggests a non-
327 monotonic relation between ES abundance and ARG enrichment for ES-*Bact*, ES-*Firm*,
328 ES-*Bifi*, and ES-*Esch*, indicating potentially complex ecological relations between
329 bacterial community composition and antibiotic resistance. These observations are
330 supported by the analysis of individual bacterial families ([Figure S5](#); [Supplementary](#)
331 [Table 1](#)). Overall, these findings emphasize the role of microbiota composition as a key
332 confounder in population-level analyses of resistome variation.

333



334

335 **Figure 5: Antibiotic resistance gene (ARG) load predicts long-term mortality risk.** a Total
336 ARG load associates with long-term mortality risk in a 17-year follow-up (proportional hazards;
337 [Extended Data Table 6](#)). The model is adjusted for *Enterobacteriaceae* relative (log10)
338 abundance, age, smoking, gender, diabetes, antineoplastic and immunomodulating agents,
339 body mass index, self-reported antihypertensive medication, systolic blood pressure, recent
340 antibiotics use (six months before baseline), household income, and fresh salad and vegetable
341 consumption. The median hazard ratio (HR) is shown for each variable, along with the 95%
342 credible intervals (CI); variables whose CI overlaps with 1 (no association) are excluded from
343 the graph. b Cumulative incidence of all-cause mortality during the follow-up period for
344 individuals stratified by high (red; >458 RPKM) and conventional (blue) ARG load. High ARG
345 load is associated with significantly higher mortality among women (p=0.006; log-rank test;
346 multivariate Cox). A similar but non-significant trend is observed in men (p=0.06). Associations
347 of ARG load with sepsis and cause-specific mortality are shown in [Figure S7](#).

348 **Antibiotic resistance is associated with long-term mortality and sepsis risk**

349 We gathered follow-up data on all major health events, including deaths from the
350 baseline sample collection in 2002 until 2019. This data allowed us to link the baseline
351 ARG load to mortality rates and sepsis incidence during the 17 years following sample
352 collection. We anticipated that ARG load could predict all-cause mortality risk because
353 antibiotic-resistant infections may occur as comorbidities and thus contribute to mortality
354 risk due to reduced antibiotic efficacy and as a marker for dysbiotic gut microbiota. In
355 total, 863 (12.2%) of the participants died during the 17-year follow-up period, and 197
356 (2.8%) got sepsis. We estimated time-to-event associations for selected variables with a

357 probabilistic Cox model ([Methods](#)). We controlled mortality-associated covariates (age,
358 smoking, gender, diabetes, use of antineoplastic and immunomodulating agents,
359 systolic blood pressure, and self-reported antihypertensive medication) and (log₁₀)
360 relative abundance of *Enterobacteriaceae*. Indeed, this bacterial family is known to
361 harbor many antibiotic-resistant pathogenic members³⁴, and we have previously
362 reported an association between its relative abundance and increased risk of all-cause
363 and cause-specific mortality in the same cohort³⁵. Furthermore, microbiome composition
364 has been shown to be associated with infection mortality and hospitalizations³⁶. In
365 addition, we controlled income and fresh salad and vegetable consumption as potential
366 confounders as they have been inversely associated with mortality^{28,37}.

367 We observed a positive association between ARG load (log₁₀ RPKM) and increased
368 mortality risk (median HR 1.34; Figure 5a). Notably, this association with all-cause
369 mortality was stronger for ARG load than for *Enterobacteriaceae* (median HR 1.08) or
370 systolic blood pressure (median HR 1.04). Moreover, the mortality associated with the
371 high ARG load was more significant in women (Figure 5b). ARG load was specifically
372 associated with increased mortality due to respiratory causes (median HR 2.22; [Figure](#)
373 [S7](#)), although the sample sizes for cause-specific events, including infectious mortality
374 other than respiratory infections (ICD-10 A; N=13), remain relatively low ([Extended Data](#)
375 [Table 7](#)). Moreover, we observed an association between ARG load and an increased
376 sepsis risk (median HR 2.22; [Figure S7](#)). The associations between ARG load, all-
377 cause mortality, mortality by respiratory causes, and sepsis remained significant when
378 controlling for the abundance of *Enterobacteriaceae* ([Extended Data Tables 6-8](#)), other
379 prevalent bacterial families, or the 42 recorded food groups as covariates (Cox model;
380 p>0.05). These observations suggest that ARG load is a robust risk factor for mortality
381 and sepsis (Figure 5; [Figure S7](#)).

382 **DISCUSSION**

383 We have characterized the population variation and prognostic potential of gut antibiotic
384 resistome in a single representative cohort of Finnish adults. Our results highlight
385 ecological and epidemiological processes that can facilitate the emergence and spread
386 of antibiotic resistance in human populations. Moreover, our data indicates that ARG
387 load can predict long-term sepsis and mortality risk, with a prognostic potential
388 comparable to blood pressure in a 17-year follow-up. The strongest predictor of ARG
389 load was prior use of antibiotics. This suggests that the *direct selection* of antibiotic
390 resistance genes induced by antibiotic consumption is a key mechanism underlying
391 individual levels of ARG load, potentially several years ahead.

392 Moreover, further *indirect selection* of the ARGs could be caused by variations in ARG-
393 carrying bacteria in the microbiome, for instance, due to differences in immune function
394 or lifestyle. Finally, we found that known proxies for ARG *transmission* - high population
395 density, income level, and certain food groups - could predict individual ARG load.
396 Thus, these findings highlight population-level patterns that may contribute to antibiotic
397 resistance selection and transmission.

398 Each study must be evaluated in the context of its limitations. Sequence-based
399 identification of ARGs is not a guarantee of phenotypic antibiotic resistance of bacteria
400 but rather a proxy for antibiotic resistance potential in a microbial community³. The
401 patterns observed in a single population may not be generalizable to other countries.
402 Additionally, the lack of longitudinal metagenome data limits our ability to assess causal
403 relations. Our analysis is based on shallow, short-read metagenome profiling. This
404 emphasizes the most abundant and prevalent ARGs and does not allow for the
405 assembly or the creation of draft genomes, which would allow for more detailed
406 information about gene variants as well as the context of the ARGs. The limited
407 sequencing depth is balanced by the sample size, which provides sufficient statistical
408 power for generalizable predictions and distinguishing between the effects of multiple
409 potential confounders.

410 This is the largest currently available population study quantifying the participant-level
411 variation of antibiotic resistance and its socio-demographic determinants. Our results
412 confirm the need to consider such factors and stratify individuals accordingly when
413 studying resistome variation¹⁴. Our findings suggest that antibiotic resistance could be a
414 'disease of affluence', as it is more prevalent among high-income individuals with longer
415 life expectancies. Similarly, a previous report found that higher gross domestic product
416 (GDP) per person is linked to higher clinical isolate resistance across countries when
417 controlling for the level of water-sanitation infrastructure, even though not controlling for
418 infrastructure leads to the opposite trend of lower-income countries harboring higher
419 resistance^{10,15}. The elevated ARG load among women, high-income participants, and
420 the urban population calls for studies on demographic differences in morbidity and
421 mortality associated with antibiotic resistance.

422 Interestingly, despite their generally higher ARG load, the overall mortality rates are
423 generally lower in women²⁷ and urban high-income groups²⁸. These groups also had a
424 lower risk of acquiring sepsis and dying of respiratory causes. The risks associated with
425 a higher ARG load might be thus mitigated by other factors in these subpopulations.
426 However, the observed prognostic potential of high ARG load and the increasing
427 number of reported AMR-related deaths^{38,39} implies that this might be changing.
428 Resistome composition could complement the ongoing efforts to define a healthy
429 microbiome⁴⁰.

430 Antimicrobial resistance and infection mortality³⁸ are rising as leading causes of
431 mortality globally. This trend may be exacerbated by shifts in lifestyle towards conditions
432 that favor ARG selection and transmission through increased consumption of antibiotics
433 and animal protein, urbanization, and international travel. In Finland, poultry
434 consumption has increased by 57% during the past two decades⁴¹ since our original
435 sample collection, and the population has become more centered in urban areas.
436 Although in Finland, antibiotic use in humans declined over that period, and resistance
437 levels mostly remained stable⁴², global antibiotic consumption and mortality attributed to
438 antibiotic resistance steadily increased over time^{43,44}. Moreover, the demand for animal

439 protein and the use of antibiotics in meat production^{18,45} has increased globally. It has
440 been predicted that AMR infections will be the most common cause of death by 2050,
441 marking a shift from cancer and cardiovascular diseases². Our findings represent early
442 warning signals of these shifts.

443

444 **STAR Methods**

445

446 **RESOURCE AVAILABILITY**

447

448 **Lead contact**

449 Further information and requests for resources should be directed to and will be fulfilled by the lead
450 contacts: Katariina Pärnänen and Leo Lahti, katariina.parnanen@utu.fi, leo.lahti@utu.fi

451

452 **Materials availability**

453 This study did not generate new unique reagents

454

455 **Data and code availability**

456 The metagenomic data are available from the European Genome-Phenome Archive
457 (accession number [EGAD00001007035](https://www.ebi.ac.uk/ena/browser/view/EGAD00001007035)). The phenotype data contain sensitive
458 information from healthcare registers and are available through the THL biobank upon
459 submission of a research plan and signing a data transfer agreement
460 (<https://thl.fi/en/web/thl-biobank/for-researchers/application-process>).

461

462 The analysis source code is available for review as a tar archive. The code will be made
463 public in a GitHub repository with a permanent DOI upon acceptance.

464

465 **EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**

466 The FINRISK population surveys were conducted every five years from 1972 to 2012
467 with the primary objective of tracking trends in cardiovascular disease risk factors in the
468 Finnish adult population. The FINRISK 2002 study utilized a stratified random sampling
469 approach of individuals between the ages of 25 and 74 from specific regions of Finland
470 (Figure 2). These areas included North Karelia in the east, Northern Savonia in the east,
471 Oulu in the northwest, the province of Lapland in the north, Turku and Loimaa regions in
472 the southwest, and the cities of Helsinki and Vantaa capital region in the south. In
473 addition, we used the West-East split of the regions based on the broad demographic
474 and genetic characteristics of the Finnish population; the Western subset covers the
475 regions of Turku/Loimaa and Helsinki/Vantaa, and the Eastern subset covers the rest of
476 the regions (North Karelia, Northern Savonia, Oulu, Lapland). The sampling procedure
477 was stratified by sex, region, and 10-year age group, resulting in 250 participants in
478 each stratum. For Northern Karelia, Lapland, and the cities of Helsinki and Vantaa, the

479 strata of 65-74-year-old men and women were also sampled, each with 250
480 participants. The initial population sample comprised 13,500 individuals (excluding 64
481 who had died or moved away between sample selection and the survey), with an overall
482 participation rate of 65.5% (n = 8,798). Of the participants, n = 7,231 individuals
483 successfully underwent stool shotgun sequencing. Of those, 129 participants withdrew
484 their consent from the THL Biobank at the time of the study. We excluded four
485 individuals with zero reads mapping to the ARG database from the analysis.
486 Subsequently, n = 7,095 participants (mean age 49 years, 55 % women) remained for
487 unsupervised analysis. Due to a lack of external cohorts with sufficient microbiome
488 profiling and long-term health data, we used two internal subsamples to achieve a 70/30
489 train test split (n = 5,000 and n = 2,098). We used cross-validation to examine the
490 robustness of the results within the cohort.

491 *Population density* The address-level coordinates of the participants were mapped with
492 the *sf* R package⁴⁶ v. 1.0.9. to a 1 km² population grid from 2005, obtained through the
493 *geofi* R package⁴⁷ v. 1.0.7. of the participants' home addresses ranged from 1 to 19,175
494 inhabitants/km² (mean 1,753 inhabitants/km²). The most densely populated regions are
495 in Southern and South-Western Finland (in the cities of Helsinki and Turku,
496 respectively). We classified the population density into five levels: (<10) 0-9
497 inhabitants/km²; (<100) 10-99; (<1e3) 100-999; (<1e4) 1000-9999; (<2e4) 10,000-
498 20,000. The data points were randomly displaced within a 5 km x 5 km grid to obscure
499 identifiable addresses in the figures. The figures do not show addresses with a
500 population density of less than 10/km².

501 *Cumulative number of total antibiotic drug purchases* during the past seven years
502 before baseline varied from 0 to 85 (mean 3.3; Figure 2).

503 *Household income* data was collected based on a questionnaire and was used as the
504 primary demographic descriptor variable alongside gender and age. We also used
505 education level (educational years adjusted for birth year, with the levels low, medium,
506 and high) in the models.

507 **Ethical approval**

508 The study protocol of FINRISK 2002 was approved by the Coordinating Ethical
509 Committee of the Helsinki and Uusimaa Hospital District (Ref. 558/E3/20 1 All
510 participants signed informed consent. The study was conducted according to the World
511 Medical Association's Declaration of Helsinki on ethical principles.

512

513 **METHOD DETAILS**

514

515 **Baseline examination**

516 The FINRISK 2002 survey included a self-administered questionnaire, physical
517 measurements, and blood and stool sample collection. The questionnaire and an

518 invitation to the health examination were mailed to all subjects. Trained nurses
519 conducted physical examinations and blood sampling in local health centers or other
520 survey sites. The participants were advised to fast for ≥ 4 hours and avoid heavy meals
521 earlier during the day. The venous blood samples were centrifuged at the field survey
522 sites, stored at -70°C , and transferred daily to the Finnish Institute for Health and
523 Welfare laboratory. Data was collected for physiological measures, biomarkers, and
524 dietary, demographic, and lifestyle factors.

525 **Stool sample collection**

526 All willing participants were given a stool sampling kit at the baseline examination with
527 detailed instructions. The participants mailed their samples overnight between Monday
528 and Thursday under Finnish winter conditions to the Finnish Institute for Health and
529 Welfare laboratory, where they were stored at -20°C . The stool samples were
530 transferred frozen in 2017 to the University of California San Diego for microbiome
531 sequencing.

532 **Stool DNA extraction and library preparation**

533 A miniaturized version of the Kapa HyperPlus Illumina-compatible library prep kit (Kapa
534 Biosystems) was used for library generation. DNA extracts were normalized to $5\text{ }\mu\text{g}$
535 total input per sample in an Echo 550 acoustic liquid-handling robot (Labcyte Inc. A
536 Mosquito HV liquid-handling robot (TTP Labtech Inc. was used for 1/10 scale enzymatic
537 fragmentation, end-repair, and adapter-ligation reactions. Sequencing adapters were
538 based on the iTru protocol⁴⁸, in which short universal adapter stubs are ligated first, and
539 then sample-specific barcoded sequences are added in a subsequent PCR step.
540 Amplified and barcoded libraries were then quantified by the PicoGreen assay and
541 pooled in approximately equimolar ratios before being sequenced on an Illumina HiSeq
542 4000 instrument to an average read count of $\sim 900,000$ reads per sample.

543 **Taxonomic and ARG profiling from sequencing data**

544 We analyzed shotgun metagenomic sequences using a pipeline built with the
545 Snakemake⁴⁹ bioinformatics workflow library. We trimmed the sequences for quality and
546 adapter sequences using Atropos⁵⁰. We removed host reads by genome mapping
547 against the human genome assembly GRCh38 with Bowtie2⁵¹.

548 We performed taxonomic profiling using MetaPhlan3⁵² for R1 and R2 reads using the
549 default settings. We mapped the R1 and R2 reads with Bowtie2 v 2.4.4⁵¹ against the
550 ResFinder database version 2.1.1⁵³ with the following options: “-D 20 -R 3 -N 1 -L 20 -i
551 S,1,0 5”. SAMtools v1.10⁵⁴ was used to filter and count reads and if both reads mapped
552 to the same gene the read was counted as one match and if the reads mapped to

553 different genes, both were counted as hits to the respective gene. ARG counts were
554 normalized by library size (number of reads per sample), ARG length, and the sum of all
555 normalized ARGs per kilobase per million reads (RPKM).

556 **Phylogenetic tree visualization of bacterial taxa and ARGs**

557 We explored the phylogenetic distribution of ARGs in the ResFinder4 database in public
558 sequence data, as our shallow shotgun sequencing did not allow for assigning ARGs to
559 their host genomes using our data. We ran blastn using the ResFinder4 database as
560 the query and the nucleotide collection “nt” database³¹ as a reference, filtering for e-
561 value <10⁻⁶ with custom ‘outfmt 6’ including ‘taxid’ for the taxonomic identifier. The
562 blastn results were processed using TaxonKit⁵⁵ to add genus information based on the
563 identifier to match the genera found in our cohort using genus names. Genera found in
564 MetaPhlan3 mapping filtered using *mergeFeaturesByPrevalence* function in mia for at
565 least 0.1% abundance in 1% of the samples was used to build a phylogenetic tree of the
566 prevalent genera using ggtree v.3.8.2⁵⁶. The most abundant ARGs in the cohort and
567 their presence in the genera were visualized on the tree using the *heatmap* function
568 from ggtree.

569 **Register linkage for pre-existing diseases and medication use at baseline**

570 In Finland, each permanent resident is assigned a unique personal identity number at
571 birth or after immigration, which ensures reliable linkage to the electronic health
572 registers. The Finnish health registers cover nearly 100% of all major health events
573 (Hospital Discharge Register, since 1969) and all prescription drug purchases (Drug
574 Purchase Register, since 1995). The quality of the diagnoses in the Finnish national
575 registers has been previously validated^{5,6}. Antibiotic drug usage was based on
576 prescription drug purchases (Drug Purchase Register with the Anatomical Therapeutic
577 Chemical (ATC) class J01, which we used as a proxy for actual antibiotic use. Baseline
578 antibiotics use (n=1,246) was defined as a purchase with an ATC code of J01* up to
579 6 months before baseline. The participants were followed through Dec 31, 2019.

580 Penicillin and other beta-lactam-antibiotics were purchased most often (5390 and 5529
581 unique purchases in the cohort during 7 years of recording before sampling).
582 Tetracyclines were purchased 5179 times, macrolides, lincosamides, and
583 streptogramins 4620 times. There were no purchases of aminoglycosides
584 ([Supplementary Table 1](#)).

585 **Food questionnaire**

586 Habitual diet was assessed using a food propensity questionnaire (FPQ), which
587 contained 42 food items with choices ranging from 1 – 6 for consumption frequency.
588 Answers denote the following descriptions: An answer 1 (“Less than once a month”) 2
589 (“Once or twice a month”) 3 (“Once a week”) 4 (“Couple of times a week”) 5 (“Almost
590 every day”), and 6 (“Once a day or more often”). For fresh vegetable and salad
591 consumption, the answers 1 and 2 were combined, resulting in new levels 1 (Less than
592 twice a month), 2 (“Once a week”), 3 (“Couple of times a week”), 4 (“Almost every day”),
593 5 (“Once a day or more often”). For poultry, levels 5 and 6 were combined, resulting in a
594 new level 5 (Almost every day or more often), but the other levels were kept the same

595 as in the original questionnaire. Additionally, the healthy food score⁵⁷, HFC was used
596 as a proxy for the general healthiness of the diet.

597 **Regional analysis**

598 For regional analysis, we used the six geographical regions defined above (North
599 Karelia, Northern Savonia, Turku, and Loimaa, Helsinki, and Vantaa, Oulu, and
600 Lapland) or East-West split of the regions.

601

602 **QUANTIFICATION AND STATISTICAL ANALYSIS**

603 **Participant data and variable preprocessing**

604 We excluded all variables with near zero variance (*caret* R package⁵⁸ v. 6.0-94) or more
605 than 500 missing values. The ARG load was \log_{10} transformed for all the statistical
606 analyses. The dichotomous variables and variables with less than ten levels were
607 unscaled, and other variables, excluding ARG load, were scaled.

608 **Statistical analysis**

609 All statistical analyses were done in R⁵⁹ version 4.3.1. We corrected for multiple testing
610 using FDR correction (Benjamini–Hochberg) (*R stats* package). We report the adjusted
611 p values. We considered an FDR-corrected $P < 0.05$ significant. All figures were
612 created with *ggplot2*⁶⁰ v. 3.4.4 unless otherwise indicated. For all analyses, including
613 microbial taxa, the taxa abundances were centered log-ratio (CLR) transformed to
614 account for compositionality in sequencing data unless otherwise indicated.

615 **General cohort statistics**

616 ARG load was measured using the total sum of all ARGs' reads per kilobase per million
617 mapped reads (RPKM). The RPKM values varied considerably among the participants
618 (mean 268 RPKM). The total number of ARG reads mapping to the ARG database had
619 a mean of 468 per sample. There was no association between library size and ARG
620 relative abundances ($p = 0.4$, linear model). The diversity of ARGs, as measured by the
621 Shannon diversity index, ranged from 0 to 5, with a mean of 3 (Supplementary Figure
622 1). The number of unique ARGs detected ranged from 1 to 194, with a mean of 42.
623 ARG diversity and load were higher in participants with prior antibiotic use, but ARG
624 richness did not vary significantly (linear model, log-transformed, $\alpha = 0.05$).

625 **Alpha diversity**

626 We characterized the alpha diversity of the microbiome with the Shannon index using
627 the complete species-level abundance data for the taxonomic profiles, and using the
628 complete ARG abundance data for the resistome profiles.

629 **Beta diversity**

630 We used the standard combination of (non-linear) principal coordinate analysis (PCoA)
631 based on the Bray–Curtis dissimilarity index (estimated with the R packages *scater*⁶¹
632 v1.29.4 and *vegan*⁶² v2.6-4 to visualize the overall population variation of the
633 microbiome and resistome composition. The beta-diversity analysis for taxonomic
634 composition was based on species-level relative abundance data from MetaPhlan3⁵².
635 The beta-diversity analysis for resistome composition was based on the ARG profiles.

636 **Selection of covariates for modeling**

637 The covariates included in linear models were chosen based on 1148 available
638 covariates. We removed covariates that defined events after sampling and had at least
639 500 missing values or near zero variance (R package *caret* v6.0-94, *nzv* function with
640 default settings), yielding 134 covariates ([Supplementary Table 1](#)) that included prior
641 disease diagnoses for major non-communicable diseases and drug purchase events, as
642 well as geographic region, gender, age, and food frequency. The diseases that passed
643 the filtering criteria included high blood pressure, asthma, diabetes, skeletal fractures,
644 ischemic heart disease, and major cardiovascular events. The majority of diseases did
645 not pass the near-zero variance criterion described above. Two categories for drug
646 purchases were used to investigate if the association differed between recent (6
647 months) and prior use (7 years).

648 **Linear models**

649 We performed linear models pairwise with \log_{10} (ARG load in RPKM) and ARG diversity
650 and all the explanatory variables that fit the selection criteria. Antibiotic use was
651 controlled using the following parameters: number of events treated with all antibiotics
652 and tetracyclines, prevalent MLSB and tetracycline use, and use of any antibiotic during
653 the past month before baseline. The exponent of the coefficients for ARG load is
654 reported in the main text for ease of interpretation. The pairwise Pearson correlations
655 between key variables (for estimating collinearity) are reported in [Extended Data Table](#)
656 [9](#).

657 **Boosted GLM**

658 Boosted Generalized linear models (GLMs) with Gaussian distribution were fitted to
659 associate \log_{10} (ARG load in RPKM) (and, separately, ARG diversity) with covariates
660 using the R packages *mboost*⁶³ v2.9-8, and *caret*⁵⁸ v6.0-94. We followed the same
661 selection for boosted GLM as for pairwise linear models, except excluding cholesterol
662 and BMI since they were collinear with income level and diet ([Extended Data Table 9](#)),
663 which were of interest (Pearson correlation, p-value <0.05). We further excluded the
664 general diagnosis for mental diseases as it overlapped with drug purchases. Drug
665 purchases were included as both prior (since 1995) and as baseline (past six months) to
666 investigate both short and long-term associations. In boosted GLMs, the generalized
667 linear model is fitted using a boosting algorithm based on component-wise univariate
668 (generalized) linear models. During fitting, the variable selection is performed. The
669 regression coefficients can be interpreted as regular GLM covariates. We reported the
670 exponent of the coefficients for models with \log_{10} (ARG load) in the main text for ease
671 of interpretation.

672 The generalizability of the fitted models was assessed with cross-validation. Five
673 thousand randomly selected participants from the cohort were used for model training,
674 while the remaining 2,095 individuals were used for testing to obtain a 70/30 train-test
675 split. We excluded anthropometric variables such as height and blood markers for
676 cholesterol and triglycerides since those markers are collinear with diet, lifestyle, and
677 gender, which were our main research interests.

678 Additionally, we fit the boosted GLM by including the most prevalent bacterial families
679 (an abundance of > 0.01% in 1% of samples) and the above-mentioned covariates.

680 Our analysis revealed collinearity between several covariates in the data used to build
681 the boosted GLM (Pearson correlation, p-value <0.05), including income class, sex, and
682 food frequencies for foods such as ready-to-eat meals, raw vegetables, and sausages.
683 Further examination of collinearity is presented in [Extended Data Table 9](#). Despite the
684 collinearity of the covariates, our results show that each model component explains
685 additional variation not captured by the other covariates, and the resulting model does
686 not have variable collinearity.

687

688 **GLMs**

689 To validate that the association of fresh vegetables and poultry with ARG load (log₁₀) is
690 not because of underlying taxonomic composition shifts due to the consumption of
691 these foods, we ran pairwise GLMs with fresh salad or poultry consumption and
692 prevalent bacterial families using Gamma distribution and log link. We corrected the p-
693 values with the FDR method.

694

695 To validate gender differences, we ran GLMs with fresh vegetable consumption and
696 antibiotic purchases as responses and adjusted with age using Gaussian distribution.

697

698 **Probabilistic analyses**

699 To detect potentially non-linear trends in ARG load concerning key covariates (fresh
700 vegetable consumption, poultry consumption, household income, population density,
701 age) and to quantify uncertainties (Figure 3c), we implemented a probabilistic model to
702 predict mean ARG load. We modeled the relation between ARG load and each factor
703 level in the given covariate based on the lognormal distribution using the default values
704 in the R *brm* function from the *brms* package (version 2.21.0). The model can be
705 summarized in the following pseudocode: `brm(ARGload ~ factor(variable)-1, family =`
706 `lognormal())`. We ran this model separately for each gender. Posterior simulations were
707 used to estimate the mean and credible intervals at the top 10% quantiles for the
708 lognormal distribution at each factor level.

709 **Survival analysis**

710 We used the Cox-proportional hazards model to predict all-cause mortality during the
711 17-year follow-up after baseline. We inferred the model parameters with a probabilistic
712 multivariate Cox model using the *brms* v 2.20.4⁶⁴ and *tidybayes* v3.0.6 R packages.
713 We used the *ggfortify*^{65,66} v0.4.16 and *survminer*⁶⁷ v0.4.9 packages to generate the
714 Kaplan-Meier curve. We verified the probabilistic analyses with frequentist analyses
715 based on the *survival*^{68,69} v3.5-7 R package. Survival analysis was controlled for
716 *Enterobacteriaceae* abundance, which we previously reported to associate with
717 increasing mortality risk in this cohort (Salosensaari et al. 2021), other mortality-
718 associated covariates used in that publication (age, smoking, gender, diabetes, use of
719 antineoplastic and immunomodulating agents, body-mass index, self-reported
720 antihypertensive medication), and fresh vegetable consumption and income class were
721 included as controls. Income and fresh salad and vegetable consumption were adjusted
722 for in the model since they have been negatively associated with mortality^{28,37} but
723 positively with ARG load. *Enterobacteriaceae* abundance and total ARG load were
724 log₁₀-transformed before the analysis. The median hazard ratio (HR) in Fig. 5a shows a

725 relative change in mortality risk following a unit change in each covariate based on the
726 probabilistic multivariate Cox regression model. The covariates that exhibited strong
727 independent association with mortality are shown (>95% Bayesian credible intervals do
728 not include zero); the credible intervals can be considered a probabilistic version of
729 classical confidence intervals. The Kaplan-Meier curves (Fig. 5b) compare survival
730 between the individuals with high vs. conventional ARG load (the top-10% quantile vs.
731 others; >458 RPKM), controlled for the same covariates as in the Cox regression model.
732 The classical multivariate Cox model further confirms the significant association
733 ($p < 0.02$).

734

735 **Enterosignatures**

736 We adapted recently proposed *enterosignatures* to summarize the community
737 composition in a few coherent subcommunities³³. The enterosignature approach was
738 proposed to complement the earlier attempts to stratify each individual into one of the
739 few distinct community types driven by major groups of gut bacteria. In summary, we
740 applied non-negative matrix factorization (NMF) on the genus-level relative abundances
741 after combining the rare genera (<1% prevalence above 0.1% relative abundance) into
742 a single group ("Other"). We run NMF with 2-10 components with 10 runs using the
743 default parameters in the function *nmf* from the R package NMF (v. 0.26). The
744 Silhouette consensus measure in the function output indicated an optimal solution of 5
745 NMF components. Frioux et al. (2023) originally reported the same optimal number of
746 components. We observed a notable correspondence of these signatures between
747 FINRISK (Figure S6) and Frioux et al. (2023; Figure 1). Three of the components had
748 the same most abundant genus in both cases (*Bacteroides*, *Prevotella*,
749 *Bifidobacterium*). In the *Firmicutes* component, the most abundant ten genera were
750 *Firmicutes*, which accounted for 94% of this component in FINRISK. The *Escherichia*
751 component was dominated by *Butyrivibrio* in FINRISK, with fewer Proteobacteria in the
752 FINRISK data. Yet, it was the only signature observed in FINRISK associated with
753 *Escherichia* (a scaled component loading 100% for this signature). Notably, despite the
754 differences, including independent data sets, metagenomic preprocessing pipelines,
755 and implementation details, the five ES identified in FINRISK had a direct qualitative
756 correspondence with the initially reported enterosignatures. We also checked that the
757 enterosignature abundances were robust to variations in library sizes (Kendall's tau;
758 $p > 0.05$ for all ES).

759

760 **Kendall's Rank Correlation**

761 Associations between diversities, ARG load, and bacterial families or Enterosignatures were
762 calculated using Kendall's Rank Correlation (also known as Kendall's *Tau*), followed by FDR
763 correction using the *mia*⁷⁰ package v. 1.9.19 function *getExperimentCrossAssociation*.

764

765 **ACKNOWLEDGEMENTS**

766 KP, LL, VL, and GSK were supported by grants from the Research Council of Finland
767 (decisions 348439 to KP; 330887 to LL and VL; 340314 to GSK) and the Alhopuro
768 Foundation (decisions 20220114 to KP and 20210172 to GSK). CGV and GM are
769 supported by an Australian National Health and Medical Research Council (NHMRC)

770 grant GNT2013468. CSC IT Centre for Science computational resources were used for
771 bioinformatic analysis of the samples. BioRender was used to create Figure 1. Shivang
772 Bhanushali is acknowledged for participating in proofreading the manuscript.

773 **AUTHOR CONTRIBUTIONS**

774
775 KP wrote the first draft of the manuscript. LL supervised the work. KP, MR, VL and LL
776 analyzed the data. RK, VS, ASH, and TN acquired the data. KP, VS, TN, ASH and LL
777 designed the study. KP, GSK, and LL wrote the manuscript. All authors contributed to
778 and accepted the manuscript.

779 **DECLARATION OF INTERESTS**

780 R.K. is a scientific advisory board member and consultant for BiomeSense, Inc., has
781 equity, and receives income. He is a scientific advisory board member and has equity in
782 GenCirq. He is a consultant and scientific advisory board member for DayTwo and
783 receives income. He has equity in and acts as a consultant for Cybele. He is a co-
784 founder of Biota, Inc., and has equity. He is a cofounder of Micronoma and has equity
785 and is a scientific advisory board member. The terms of these arrangements have been
786 reviewed and approved by the University of California, San Diego, in accordance with
787 its conflict of interest policies. Other authors declare no competing interests.

788

789 **REFERENCES**

- 790 1. Borodulin, K., Tolonen, H., Jousilahti, P., Jula, A., Juolevi, A., Koskinen, S.,
791 Kuulasmaa, K., Laatikainen, T., Männistö, S., Peltonen, M., et al. (2018). Cohort
792 Profile: The National FINRISK Study. *International Journal of Epidemiology* 47, 696–
793 696i. <https://doi.org/10.1093/ije/dyx239>.
- 794 2. O'Neill, J. (2014). Antimicrobial resistance: tackling a crisis for the health and wealth
795 of nations. *Rev. Antimicrob. Resist.*
- 796 3. van Schaik, W. (2015). The human gut resistome. *Phil. Trans. R. Soc. B* 370,
797 20140087. <https://doi.org/10.1098/rstb.2014.0087>.
- 798 4. Smillie, C.S., Smith, M.B., Friedman, J., Cordero, O.X., David, L.A., and Alm, E.J.
799 (2011). Ecology drives a global network of gene exchange connecting the human
800 microbiome. *Nature* 480, 241–244. <https://doi.org/10.1038/nature10571>.
- 801 5. Forslund, K., Sunagawa, S., Kultima, J.R., Mende, D.R., Arumugam, M., Typas, A.,
802 and Bork, P. (2013). Country-specific antibiotic use practices impact the human gut
803 resistome. *Genome Research* 23, 1163–1169.
804 <https://doi.org/10.1101/gr.155465.113>.
- 805 6. Forslund, K., Sunagawa, S., Coelho, L.P., and Bork, P. (2014). Metagenomic
806 insights into the human gut resistome and the forces that shape it: Prospects &
807 Overviews. *BioEssays* 36, 316–329. <https://doi.org/10.1002/bies.201300143>.
- 808 7. Berendonk, T.U., Manaia, C.M., Merlin, C., Fatta-Kassinos, D., Cytryn, E., Walsh, F.,
809 Burgmann, H., Sorum, H., Norstrom, M., Pons, M.-N., et al. (2015). Tackling

- 810 antibiotic resistance: The environmental framework. *Nat Rev Micro* 13, 310–317.
- 811 8. Martinez, J.L. (2008). Antibiotics and antibiotic resistance genes in natural
812 environments. *Science* 321, 365–367. <https://doi.org/10.1126/science.1159483>.
- 813 9. Forsberg, K.J., Patel, S., Gibson, M.K., Lauber, C.L., Knight, R., Fierer, N., and
814 Dantas, G. (2014). Bacterial phylogeny structures soil resistomes across habitats.
815 *Nature* 509, 612–616. <https://doi.org/10.1038/nature13377>.
- 816 10. Collignon, P., Beggs, J.J., Walsh, T.R., Gandra, S., and Laxminarayan, R. (2018).
817 Anthropological and socioeconomic factors contributing to global antimicrobial
818 resistance: a univariate and multivariable analysis. *The Lancet Planetary Health* 2,
819 e398–e405. [https://doi.org/10.1016/S2542-5196\(18\)30186-4](https://doi.org/10.1016/S2542-5196(18)30186-4).
- 820 11. Bengtsson-Palme, J., Angelin, M., Huss, M., Kjellqvist, S., Kristiansson, E.,
821 Palmgren, H., Larsson, D.G.J., and Johansson, A. (2015). The human gut
822 microbiome as a transporter of antibiotic resistance genes between continents.
823 *Antimicrob Agents Chemother* 59, 6551–6560. [https://doi.org/10.1128/AAC.00933-](https://doi.org/10.1128/AAC.00933-15)
824 15.
- 825 12. Bruinsma, N. (2003). Influence of population density on antibiotic resistance.
826 *Journal of Antimicrobial Chemotherapy* 51, 385–390.
827 <https://doi.org/10.1093/jac/dkg072>.
- 828 13. Johnson, J.R., Kuskowski, M.A., Smith, K., O'Bryan, T.T., and Tatini, S. (2005).
829 Antimicrobial-resistant and extraintestinal pathogenic *Escherichia coli* in retail
830 foods. *J INFECT DIS* 191, 1040–1049. <https://doi.org/10.1086/428451>.
- 831 14. WHO World Health Organization (2018). Tackling antimicrobial resistance (AMR)
832 together: working paper 5.0: enhancing the focus on gender and equity.
833 <https://apps.who.int/iris/handle/10665/336977>.
- 834 15. Hendriksen, R.S., Munk, P., Njage, P., Van Bunnik, B., McNally, L., Lukjancenko,
835 O., Röder, T., Nieuwenhuijse, D., Pedersen, S.K., Kjeldgaard, J., et al. (2019).
836 Global monitoring of antimicrobial resistance based on metagenomics analyses of
837 urban sewage. *Nat Commun* 10, 1124. [https://doi.org/10.1038/s41467-019-08853-](https://doi.org/10.1038/s41467-019-08853-3)
838 3.
- 839 16. Collignon, P., and Beggs, J.J. (2019). Socioeconomic enablers for contagion:
840 Factors impelling the antimicrobial resistance epidemic. *Antibiotics* 8, 86.
841 <https://doi.org/10.3390/antibiotics8030086>.
- 842 17. Lee, K., Raguideau, S., Sirén, K., Asnicar, F., Cumbo, F., Hildebrand, F., Segata, N.,
843 Cha, C.-J., and Quince, C. (2023). Population-level impacts of antibiotic usage on
844 the human gut microbiome. *Nat Commun* 14, 1191.
845 <https://doi.org/10.1038/s41467-023-36633-7>.
- 846 18. Van Boeckel, T.P., Brower, C., Gilbert, M., Grenfell, B.T., Levin, S.A., Robinson,
847 T.P., Teillant, A., and Laxminarayan, R. (2015). Global trends in antimicrobial use in
848 food animals. *Proc. Natl. Acad. Sci. U.S.A.* 112, 5649–5654.
849 <https://doi.org/10.1073/pnas.1503141112>.
- 850 19. Blau, K., Bettermann, A., Jechalke, S., Fornefeld, E., Vanrobaeys, Y., Stalder, T.,
851 Top, E.M., and Smalla, K. (2018). The transferable resistome of produce. *mBio* 9,

- 852 e01300-18. <https://doi.org/10.1128/mBio.01300-18>.
- 853 20. Beam, A., Clinger, E., and Hao, L. (2021). Effect of diet and dietary components on
854 the composition of the gut microbiota. *Nutrients* *13*, 2795.
855 <https://doi.org/10.3390/nu13082795>.
- 856 21. Oliver, A., Xue, Z., Villanueva, Y.T., Durbin-Johnson, B., Alkan, Z., Taft, D.H., Liu, J.,
857 Korf, I., Laugero, K.D., Stephensen, C.B., et al. (2022). Association of diet and
858 antimicrobial resistance in healthy U.S. adults. *mBio*, e00101-22.
859 <https://doi.org/10.1128/mbio.00101-22>.
- 860 22. Aarestrup, F.M., Wegener, H.C., and Collignon, P. (2008). Resistance in bacteria of
861 the food chain: epidemiology and control strategies. *Expert Review of Anti-infective*
862 *Therapy* *6*, 733–750. <https://doi.org/10.1586/14787210.6.5.733>.
- 863 23. Kurittu, P., Khakipoor, B., Aarnio, M., Nykäsenoja, S., Brouwer, M., Myllyniemi, A.-
864 L., Vatunen, E., and Heikinheimo, A. (2021). Plasmid-borne and chromosomal
865 ESBL/AmpC genes in *Escherichia coli* and *Klebsiella pneumoniae* in global food
866 products. *Front. Microbiol.* *12*, 592291. <https://doi.org/10.3389/fmicb.2021.592291>.
- 867 24. Johnson, J.R., Sannes, M.R., Croy, C., Johnston, B., Clabots, C., Kuskowski, M.A.,
868 Bender, J., Smith, K.E., Winokur, P.L., and Belongia, E.A. (2007). Antimicrobial
869 drug-resistant *Escherichia coli* from humans and poultry products, Minnesota and
870 Wisconsin, 2002–2004. *Emerg. Infect. Dis.* *13*, 838–846.
871 <https://doi.org/10.3201/eid1306.061576>.
- 872 25. Myllyniemi, A.-L., Gindonis, V., Nykäsenoja, S., and Koppinen, J. (2004). FINRES-
873 Vet 2002-2003. Finnish Veterinary Antimicrobial Resistance Monitoring and
874 Consumption of Antimicrobial Agents.
- 875 26. Kerminen, S., Havulinna, A.S., Hellenthal, G., Martin, A.R., Sarin, A.-P., Perola, M.,
876 Palotie, A., Salomaa, V., Daly, M.J., Ripatti, S., et al. (2017). Fine-Scale Genetic
877 Structure in Finland. *G3 Genes|Genomes|Genetics* *7*, 3459–3468.
878 <https://doi.org/10.1534/g3.117.300217>.
- 879 27. Oksuzyan, A., Juel, K., Vaupel, J.W., and Christensen, K. (2008). Men: good health
880 and high mortality. Sex differences in health and aging. *Aging clinical and*
881 *experimental research* *20*, 91–102.
- 882 28. Lynch, J.W., Kaplan, G.A., Cohen, R.D., Wilson, T.W., Smith, N.L., Kauhanen, J.,
883 and Salonen, J.T. (1994). Childhood and adult socioeconomic status as predictors
884 of mortality in Finland. *The Lancet* *343*, 524–527. [https://doi.org/10.1016/S0140-](https://doi.org/10.1016/S0140-6736(94)91468-0)
885 [6736\(94\)91468-0](https://doi.org/10.1016/S0140-6736(94)91468-0).
- 886 29. Gallet, C.A., and Doucouliagos, H. (2014). The income elasticity of air travel: A
887 meta-analysis. *Annals of Tourism Research* *49*, 141–155.
888 <https://doi.org/10.1016/j.annals.2014.09.006>.
- 889 30. Hu, Y., Yang, X., Li, J., Lv, N., Liu, F., Wu, J., Lin, I.Y.C., Wu, N., Weimer, B.C., Gao,
890 G.F., et al. (2016). The bacterial mobile resistome transfer network connecting the
891 animal and human microbiomes. *Applied and Environmental Microbiology* *82*,
892 6672–6681. <https://doi.org/10.1128/AEM.01802-16>.
- 893 31. NCBI Resource Coordinators (2017). Database Resources of the National Center

- 894 for Biotechnology Information. *Nucleic Acids Research* 45, D12–D17.
895 <https://doi.org/10.1093/nar/gkw1071>.
- 896 32. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic
897 local alignment search tool. *Journal of Molecular Biology* 215, 403–410.
898 [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
- 899 33. Frioux, C., Ansorge, R., Özkurt, E., Ghassemi Nedjad, C., Fritscher, J., Quince, C.,
900 Waszak, S.M., and Hildebrand, F. (2023). Enterosignatures define common
901 bacterial guilds in the human gut microbiome. *Cell Host & Microbe* 31, 1111-
902 1125.e6. <https://doi.org/10.1016/j.chom.2023.05.024>.
- 903 34. Boucher, H.W., Talbot, G.H., Bradley, J.S., Edwards, J.E., Gilbert, D., Rice, L.B.,
904 Scheld, M., Spellberg, B., and Bartlett, J. (2009). Bad Bugs, No Drugs: No ESKAPE!
905 An Update from the Infectious Diseases Society of America. *Clinical Infectious*
906 *Diseases* 48, 1–12. <https://doi.org/10.1086/595011>.
- 907 35. Salosensaari, A., Laitinen, V., Havulinna, A.S., Meric, G., Cheng, S., Perola, M.,
908 Valsta, L., Alfthan, G., Inouye, M., Watrous, J.D., et al. (2021). Taxonomic
909 signatures of cause-specific mortality risk in human gut microbiome. *Nat Commun*
910 12, 2671. <https://doi.org/10.1038/s41467-021-22962-y>.
- 911 36. Kullberg, R.F.J., Wikki, I., Haak, B.W., Kauko, A., Galenkamp, H., Peters-Sengers,
912 H., Butler, J.M., Havulinna, A.S., Palmu, J., McDonald, D., et al. (2024). Association
913 between butyrate-producing gut bacteria and the risk of infectious disease
914 hospitalisation: results from two observational, population-based microbiome
915 studies. *The Lancet Microbe*. [https://doi.org/10.1016/S2666-5247\(24\)00079-X](https://doi.org/10.1016/S2666-5247(24)00079-X).
- 916 37. Oyeboode, O., Gordon-Dseagu, V., Walker, A., and Mindell, J.S. (2014). Fruit and
917 vegetable consumption and all-cause, cancer and CVD mortality: analysis of Health
918 Survey for England data. *J Epidemiol Community Health* 68, 856–862.
919 <https://doi.org/10.1136/jech-2013-203500>.
- 920 38. Murray, C.J., Ikuta, K.S., Sharara, F., Swetschinski, L., Robles Aguilar, G., Gray, A.,
921 Han, C., Bisignano, C., Rao, P., Wool, E., et al. (2022). Global burden of bacterial
922 antimicrobial resistance in 2019: a systematic analysis. *The Lancet* 399, 629–655.
923 [https://doi.org/10.1016/S0140-6736\(21\)02724-0](https://doi.org/10.1016/S0140-6736(21)02724-0).
- 924 39. European Centre for Disease Prevention and Control. (2022). Assessing the health
925 burden of infections with antibiotic-resistant bacteria in the EU/EEA, 2016-2020.
926 (Publications Office).
- 927 40. Shanahan, F., Ghosh, T.S., and O’Toole, P.W. (2021). The Healthy Microbiome—
928 What Is the Definition of a Healthy Gut Microbiome? *Gastroenterology* 160, 483–
929 494. <https://doi.org/10.1053/j.gastro.2020.09.057>.
- 930 41. Natural resources institute (LUKE) (2023). Balance Sheet for Food Commodities.
931 <https://www.luke.fi/en/statistics/balance-sheet-for-food-commodities/releases>
932 <https://www.luke.fi/en/statistics/balance-sheet-for-food-commodities/releases>.
- 933 42. European Centre for Disease Prevention and Control (2023). Antimicrobial
934 consumption in the EU/EEA (ESAC-Net) - Annual Epidemiological Report 2022.
- 935 43. Browne, A.J., Chipeta, M.G., Haines-Woodhouse, G., Kumaran, E.P.A., Hamadani,

- 936 B.H.K., Zaraa, S., Henry, N.J., Deshpande, A., Reiner, R.C., Day, N.P.J., et al.
937 (2021). Global antibiotic consumption and usage in humans, 2000–18: a spatial
938 modelling study. *The Lancet Planetary Health* 5, e893–e904.
939 [https://doi.org/10.1016/S2542-5196\(21\)00280-1](https://doi.org/10.1016/S2542-5196(21)00280-1).
- 940 44. Cassini, A., Högberg, L.D., Plachouras, D., Quattrocchi, A., Hoxha, A., Simonsen,
941 G.S., Colomb-Cotinat, M., Kretzschmar, M.E., Devleeschauwer, B., Cecchini, M.,
942 et al. (2019). Attributable deaths and disability-adjusted life-years caused by
943 infections with antibiotic-resistant bacteria in the EU and the European Economic
944 Area in 2015: a population-level modelling analysis. *The Lancet Infectious Diseases*
945 19, 56–66. [https://doi.org/10.1016/S1473-3099\(18\)30605-4](https://doi.org/10.1016/S1473-3099(18)30605-4).
- 946 45. Browne, A.J., Chipeta, M.G., Haines-Woodhouse, G., Kumaran, E.P.A., Hamadani,
947 B.H.K., Zaraa, S., Henry, N.J., Deshpande, A., Reiner, R.C., Day, N.P.J., et al.
948 (2021). Global antibiotic consumption and usage in humans, 2000–18: a spatial
949 modelling study. *The Lancet Planetary Health* 5, e893–e904.
950 [https://doi.org/10.1016/S2542-5196\(21\)00280-1](https://doi.org/10.1016/S2542-5196(21)00280-1).
- 951 46. Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector
952 Data. *The R Journal* 10, 439. <https://doi.org/10.32614/RJ-2018-009>.
- 953 47. Kainu, M., Lehtomaki, J., Parkkinen, J., Miettinen, J., Kantanen, P., Vesanen, S.,
954 and Lahti, L. (2023). geofi: Access Finnish Geospatial Data.
- 955 48. Glenn, T.C., Nilsen, R.A., Kieran, T.J., Sanders, J.G., Bayona-Vásquez, N.J., Finger,
956 J.W., Pierson, T.W., Bentley, K.E., Hoffberg, S.L., Louha, S., et al. (2016).
957 Adapterama I: Universal stubs and primers for 384 unique dual-indexed or 147,456
958 combinatorially-indexed Illumina libraries (iTru & iNext) (Genomics)
959 <https://doi.org/10.1101/049114>.
- 960 49. Koster, J., and Rahmann, S. (2012). Snakemake--a scalable bioinformatics
961 workflow engine. *Bioinformatics* 28, 2520–2522.
962 <https://doi.org/10.1093/bioinformatics/bts480>.
- 963 50. Didion, J.P., Martin, M., and Collins, F.S. (2017). Atropos: specific, sensitive, and
964 speedy trimming of sequencing reads. *PeerJ* 5, e3720.
965 <https://doi.org/10.7717/peerj.3720>.
- 966 51. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie
967 2. *Nat Meth* 9, 357–359. <https://doi.org/10.1038/nmeth.1923>.
- 968 52. Franzosa, E.A., McIver, L.J., Rahnnavard, G., Thompson, L.R., Schirmer, M.,
969 Weingart, G., Lipson, K.S., Knight, R., Caporaso, J.G., Segata, N., et al. (2018).
970 Species-level functional profiling of metagenomes and metatranscriptomes. *Nature*
971 *Methods* 15, 962–968. <https://doi.org/10.1038/s41592-018-0176-y>.
- 972 53. Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O.,
973 Aarestrup, F.M., and Larsen, M.V. (2012). Identification of acquired antimicrobial
974 resistance genes. *Journal of Antimicrobial Chemotherapy* 67, 2640–2644.
975 <https://doi.org/10.1093/jac/dks261>.
- 976 54. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G.,
977 Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup

- 978 (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25,
979 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
- 980 55. Shen, W., and Ren, H. (2021). TaxonKit: A practical and efficient NCBI taxonomy
981 toolkit. *Journal of Genetics and Genomics* 48, 844–850.
982 <https://doi.org/10.1016/j.jgg.2021.03.006>.
- 983 56. Yu, G., Smith, D., Zhu, H., Guan, Y., and Lam, T.T.-Y. (2017). ggtree: an R package
984 for visualization and annotation of phylogenetic trees with their covariates and other
985 associated data. *Methods in Ecology and Evolution* 8, 28–36.
986 <https://doi.org/10.1111/2041-210X.12628>.
- 987 57. Koponen, K.K., Salosensaari, A., Ruuskanen, M.O., Havulinna, A.S., Männistö, S.,
988 Jousilahti, P., Palmu, J., Salido, R., Sanders, K., Brennan, C., et al. (2021).
989 Associations of healthy food choices with gut microbiota profiles. *The American*
990 *Journal of Clinical Nutrition* 114, 605–616. <https://doi.org/10.1093/ajcn/nqab077>.
- 991 58. Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal*
992 *of Statistical Software* 28. <https://doi.org/10.18637/jss.v028.i05>.
- 993 59. R Core Team (2017). R: A Language and Environment for Statistical Computing. (R
994 Foundation for Statistical Computing).
- 995 60. Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis. (Springer-Verlag
996 New York).
- 997 61. McCarthy, D.J., Campbell, K.R., Lun, A.T.L., and Wills, Q.F. (2017). Scater: pre-
998 processing, quality control, normalization and visualization of single-cell RNA-seq
999 data in R. *Bioinformatics* 33, 1179–1186.
1000 <https://doi.org/10.1093/bioinformatics/btw777>.
- 1001 62. Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D.,
1002 Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., et al. (2017). vegan:
1003 Community Ecology Package.
- 1004 63. Hothorn, T., Buehlmann, P., Kneib, T., Schmid, M., and Hofner, B. (2022). mboost:
1005 Model-Based Boosting.
- 1006 64. Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using
1007 Stan. *J. Stat. Soft.* 80. <https://doi.org/10.18637/jss.v080.i01>.
- 1008 65. Tang, Y., Horikoshi, M., and Li, W. (2016). ggfortify: Unified Interface to Visualize
1009 Statistical Result of Popular R Packages. *The R Journal* 8, 474–485.
1010 <https://doi.org/10.32614/RJ-2016-060>.
- 1011 66. Horikoshi, M., and Tang, Y. (2018). ggfortify: Data Visualization Tools for Statistical
1012 Analysis Results.
- 1013 67. Kassambara, A., Kosinski, M., and Biecek, P. (2021). survminer: Drawing Survival
1014 Curves using “ggplot2.”
- 1015 68. Terry M. Therneau and Patricia M. Grambsch (2000). *Modeling Survival Data:*
1016 *Extending the Cox Model* (Springer).
- 1017 69. Therneau, T.M. (2023). A Package for Survival Analysis in R.
- 1018 70. Ernst, F.G.M., Shetty, S.A., Borman, T., and Lahti, L. (2023). mia: Microbiome
1019 analysis.