

1 Supporting Information 1

2
3 Metagenomic insights into the abundance and composition of antibiotic resistance genes in hospital
4 wastewater in Benin, Burkina Faso, and Finland

5
6 Melina A. Markkanen^{a, b #}, Kaisa Haukka^{a, g}, Katariina M. M. Pärnänen^a, Victorien Tamegnon
7 Dougnon^c, Isidore Juste O. Bonkougou^d, Zakaria Garba^e, Halidou Tinto^e, Anniina Sarekoski^a, Antti
8 Karkman^{a, b}, Anu Kantele^{b, f, g *}, Marko P. J. Virta^{a, b #*}

9
10 ^a Department of Microbiology, University of Helsinki, Finland

11 ^b Multidisciplinary Center of Excellence in Antimicrobial Resistance Research, Finland

12 ^c Research Unit in Applied Microbiology and Pharmacology of Natural Substances, Polytechnic School of Abomey-
13 Calavi, University of Abomey-Calavi, Benin

14 ^d Department of Biochemistry and Microbiology, University Joseph KI-ZERBO, Ouagadougou, Burkina Faso

15 ^e Clinical Research Unit Nanoro, Institute for Research in Health Sciences, National Center for Scientific and Technological
16 Research, Burkina Faso

17 ^f Meilahti Infectious Diseases and Vaccine Research Center, MeiVac, Department of Infectious Diseases, University of
18 Helsinki and Helsinki University Hospital, Helsinki, Finland

19 ^g Human Microbiome Research Program, Faculty of Medicine, University of Helsinki, Helsinki, Finland

20
21 Corresponding authors:

22 # Marko P. J. Virta, marko.virta@helsinki.fi

23 # Melina A. Markkanen, melina.markkanen@helsinki.fi

24 * Marko P. J. Virta and Anu Kantele contributed equally to this work.

25
26 Table of Contents:

27 1. Sample Collection

28 **Figure S1.** Sample collection in Benin and Burkina Faso (p. 2)

29 **Figure S2.** Examples of sampling sites (p. 3)

30 2. Supporting Results

31 **Figure S3.** Diversity of ARGs, MGEs, and taxa (p. 4)

32 **Figure S4.** Correlation between ARGs and MGEs (p. 6)

33 **Figure S5.** Examples of putative class 1 integron elements with ARGs (p. 7)

34 **Figure S6.** PCA plots for ARGs, MGEs, and taxa with all collected samples (p. 8)

35 **Figure S7.** Examples of genetic environments of *mcr-5* genes (p. 10)

36 **Figure S8.** Significantly differentially abundant taxa in HWW from Benin, Burkina Faso, and Finland (p. 11)

37 **Figure S9.** Top 11 most abundant taxa in all samples (p. 12)

38 **Figure S10.** Analysis of replicate samples (p. 13)

39 **Figure S11.** Correlations between ALDEx2 and DESeq2 results for significantly differentially abundant ARGs (p. 14)

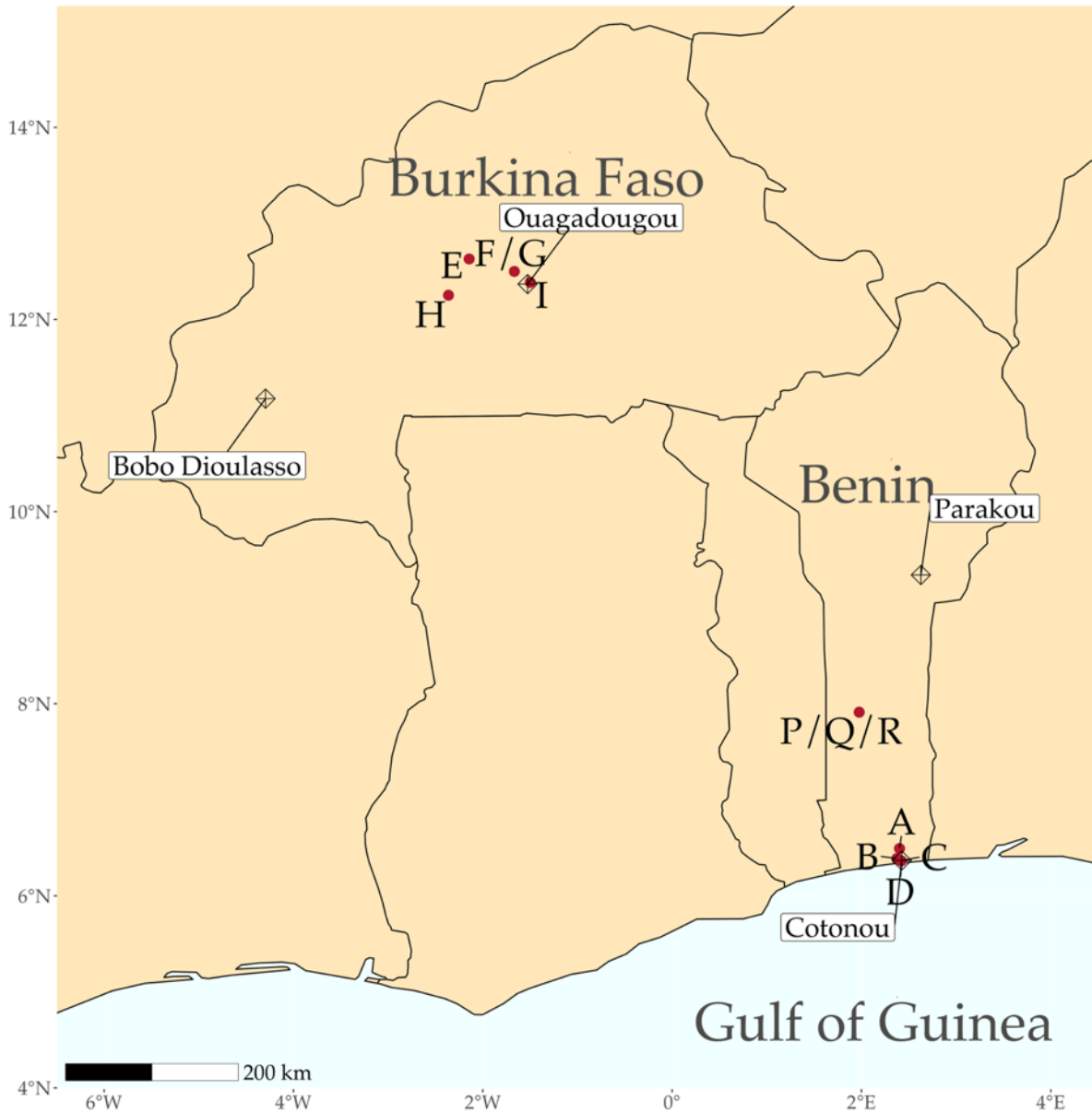
40 3. References (p. 14)

41
42 Pages: 15

43 Figures: 11

44

45 1. Sample collection



46

47 **Figure S1.** Sample collection in Benin and Burkina Faso. Sample collection sites are denoted with red dots on the
48 map. The climate in the coastal region of Benin is hot and humid. The average daytime temperature in Cotonou
49 varies between 28–32 °C and humidity between 77–84 %, with an annual rainfall of 1,308 mm. Burkina Faso is
50 characterized by a dry tropical climate. The average daytime temperature in Ouagadougou varies between 31–
51 39 °C, and the annual rainfall is 745 mm. During the sampling period in November/December, it typically rains
52 only a little in Cotonou and not at all in Ouagadougou.

53



54 A)

B)



55 C)

D)



56 E)

F)

G)

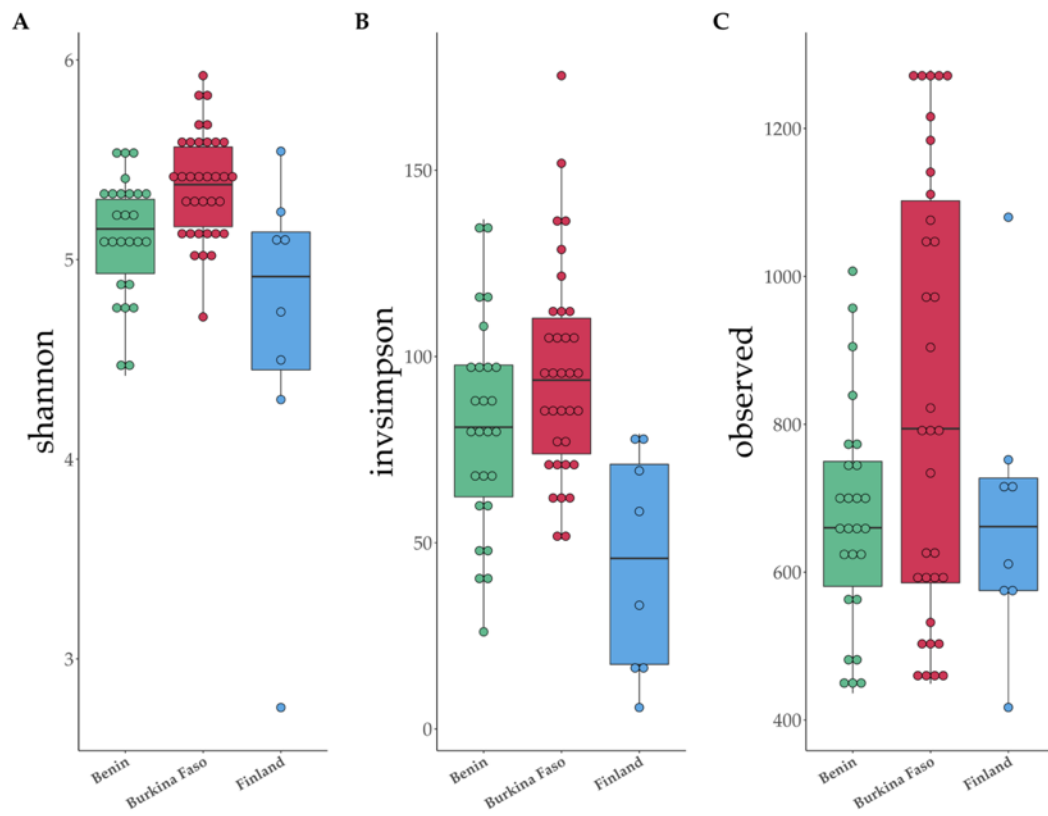
Figure S2. Examples of sampling sites. The samples collected in Benin included A) well water 100 m away from hospital A in Benin (sample BH11; BENN well water A); B) tap water in a distant village in Benin (sample BSE93; BENN tap water (drinking)); C) HWW from an open cesspit next to the wall at the hospital B (sample BH29; BENN HWW B); D) the central septic tank collecting wastewater from different sections of hospital C (sample BH44; BENN HWW C); E) water puddle surrounding the surgery room septic tank at hospital C (sample BH48; BENN puddle at yard C); F) water from a tank used for washing hands at hospital C (sample BH52; BENN hand-washing C). G) In Burkina Faso, the HWW septic tank is shown for hospital E (sample BFH1; BF HWW E). Photos by K. Haukka, A. Sarekoski and A. Butcher.

65

66 2. Supporting results

67

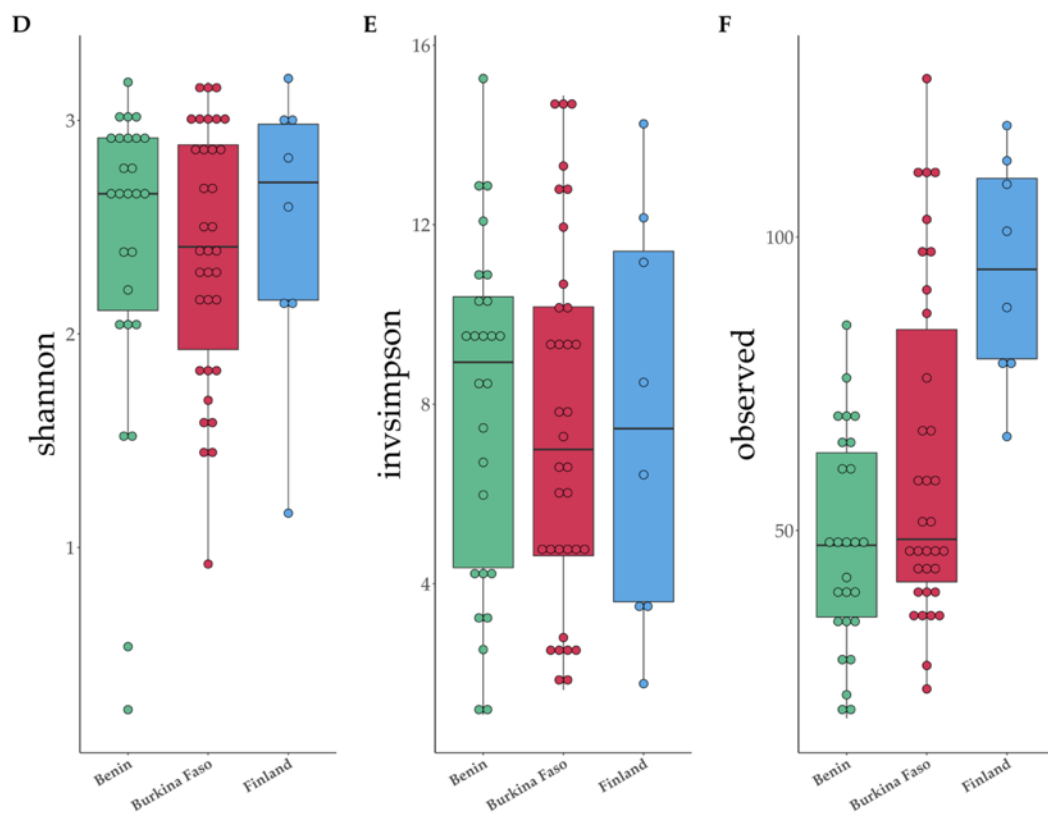
Diversity of ARGs



68

69

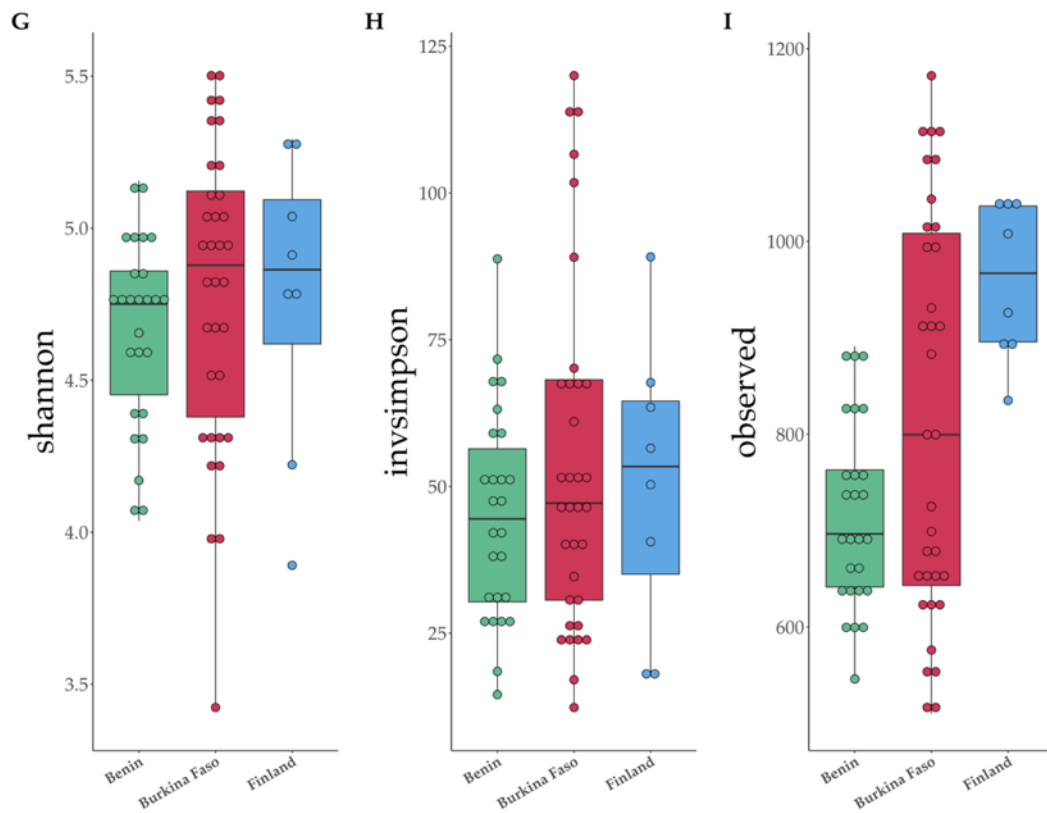
Diversity of taxa



70

71
72

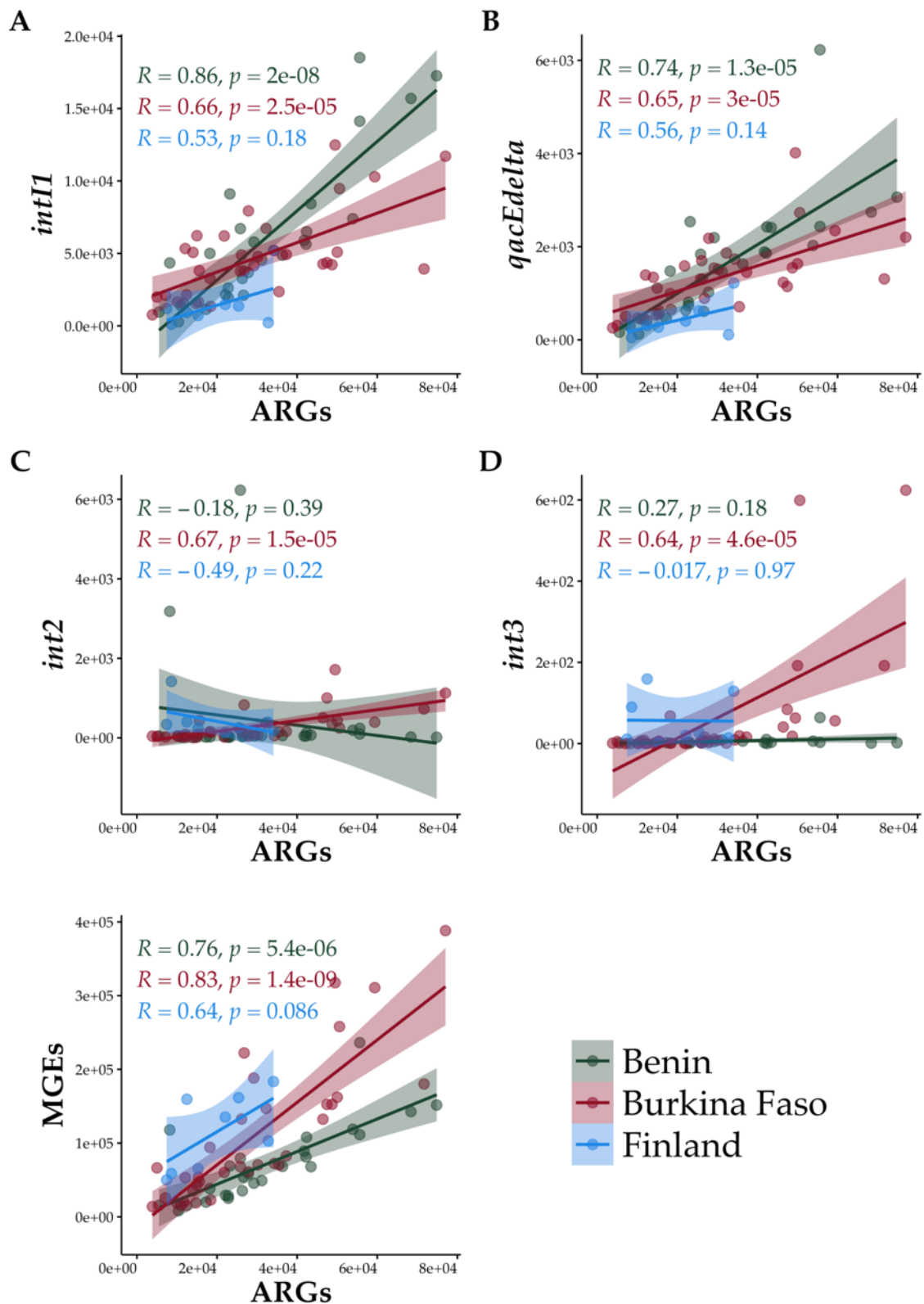
Diversity of MGEs



73

74 **Figure S3.** Diversity indexes (Shannon, Inverse Simpson, and Observed) for ARGs (A-C), taxa (D-F), and MGEs
75 (G-I). In the boxplots, country medians are shown as vertical lines and the interquartile ranges (25th and 75th
76 percentiles) as boxplot hinges. The horizontal lines represent the highest and lowest values. Outliers are values
77 higher or below 1.5 times the upper or lower percentiles, respectively. The outlier samples were FH6 (A), BFH35
78 (B), FH1 (C), BH01 and BH09 (D), and FH6 (G). Based on pairwise Wilcoxon rank sum tests adjusted by
79 Benjamini-Hochberg, the significant ($p < 0.05$) comparisons were: A) Finland – Burkina Faso and Benin –
80 Burkina Faso; B) Finland – Benin and Finland – Burkina Faso; F) Finland – Benin and Finland – Burkina Faso;
81 I) Benin – Finland. Diversity analyses of ARGs, taxa, and MGE were done with the function ‘diversity’ from the
82 ‘vegan’ package¹ (v 2.6.2). Pairwise Wilcoxon rank sum tests adjusted by Benjamini-Hochberg from the ‘stats’
83 package (78) (v 4.2.0) were performed to determine the significant comparisons.

Correlation between ARGs and MGEs in HWWs

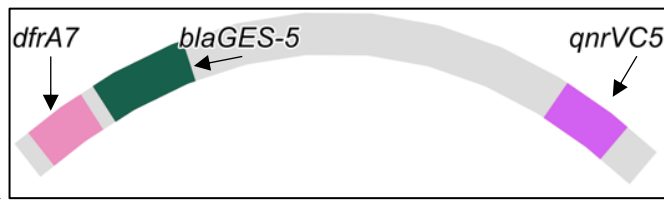


84

85 **Figure S4.** Pearson correlations between ARGs and A) *intI1*, B) *qacEA*, C) *int2*, and D) *int3* genes, and E) all
 86 MGEs. Correlations were computed using package `ggpubr`² (v 0.4.0.999).

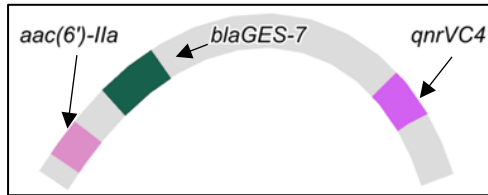
87

88



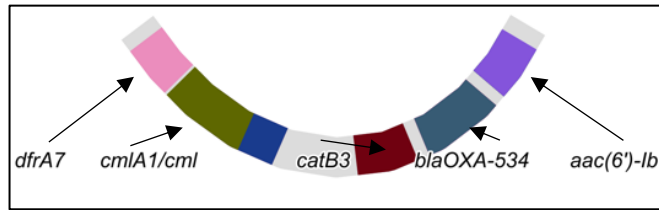
89

A



90

B



91

C

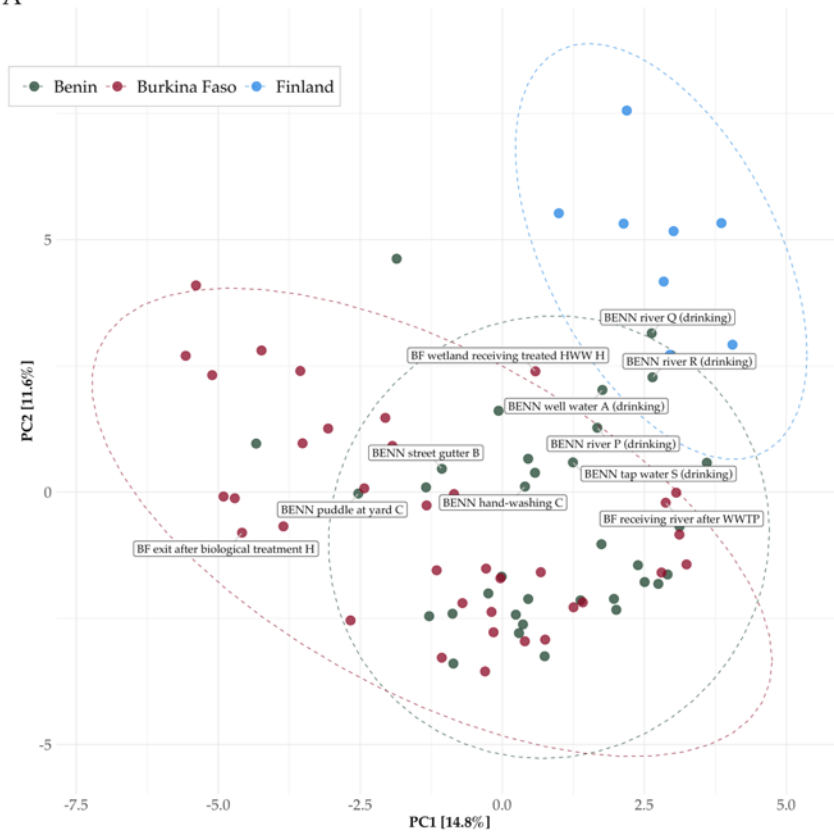
92 **Figure S5.** Examples of assembly graphs showing multidrug-resistance fragments putatively encoded by class 1
93 integron gene cassettes in samples A) BH02, B) BH48, and C) BFH19. Contigs were assembled with MEGAHIT
94 ³ (v1.2.8) with parameters --min-contig-len 1000 -m 32000000000. To visualize the co-localization of multiple
95 ARGs with Bandage tool ⁴, the assembled contigs were converted into assembly graphs using the MEGAHIT
96 command 'megahit_toolkit contig2fastg'. Contigs (putative class 1 integron gene cassettes) carrying multiple
97 ARGs were browsed visually with Bandage tool ⁴ and ResFinder database ⁵.

98

PCA, clr-transformed

ARGs

A

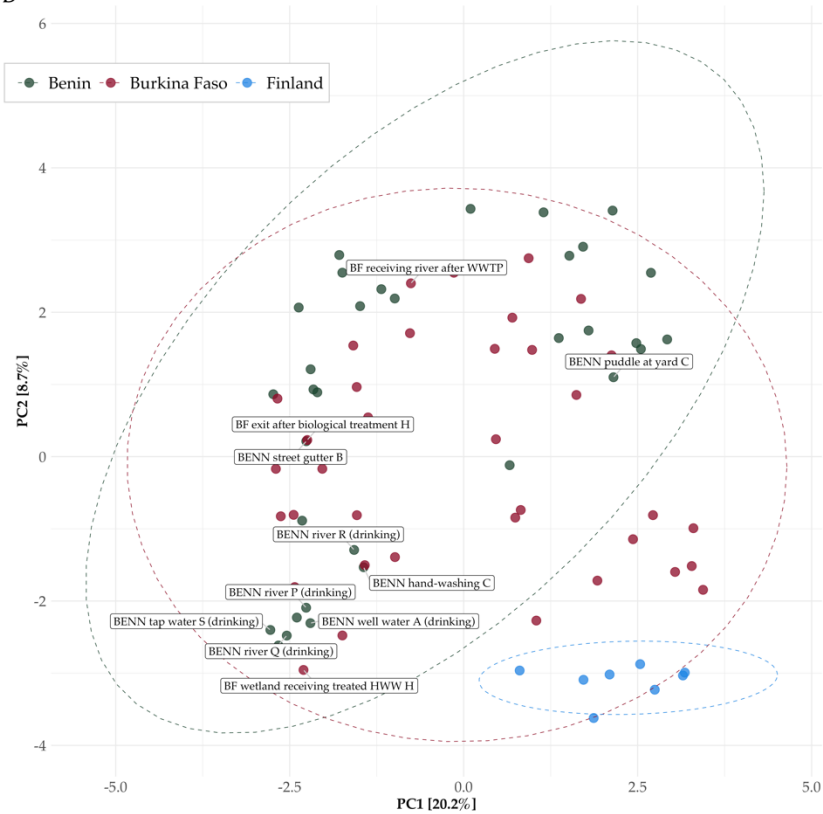


99

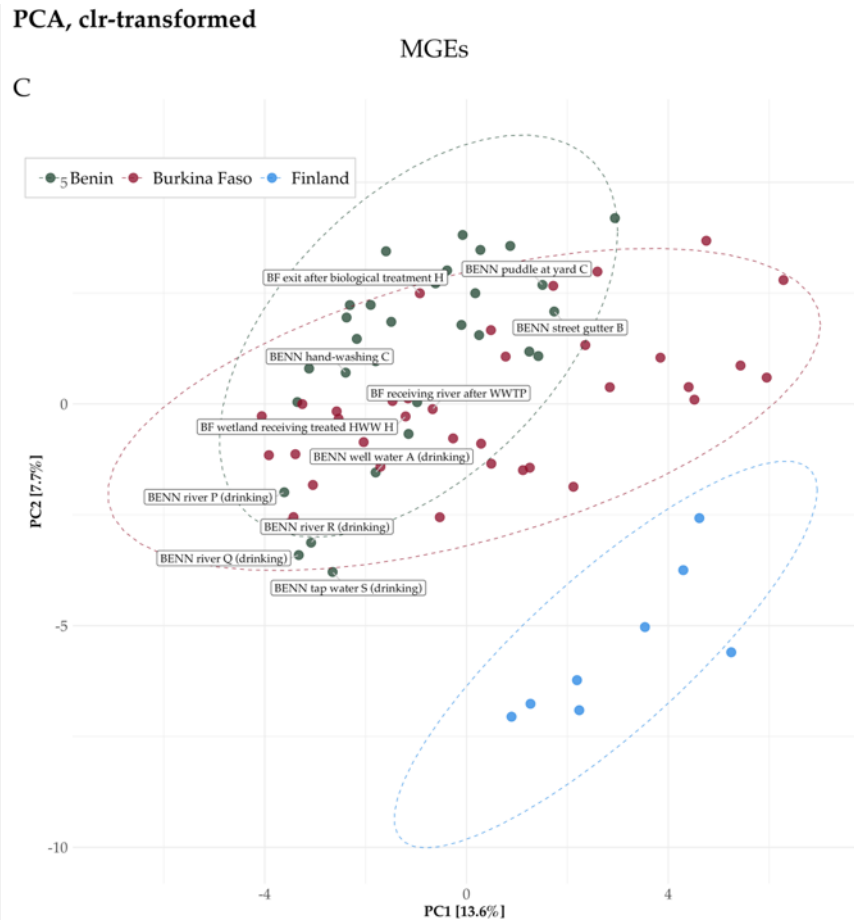
PCA, clr-transformed

Taxa

B



100



101

102 **Figure S6.** Principal components analysis (PCA) showing the significant dissimilarities of A) resistomes, B)
 103 taxonomical composition Metaphlan3, and C) mobilomes among all collected (HWW and other water samples)
 104 samples from Benin, Burkina Faso, and Finland. Non-HWW, samples are labeled with a text box, while HWW
 105 samples are unlabeled. Count data were clr-transformed and visualized with ‘microViz’⁶ (v0.9.1). Confidence
 106 ellipses represent 95 % confidence levels.

107

108

109

110

111

112

113

114

115

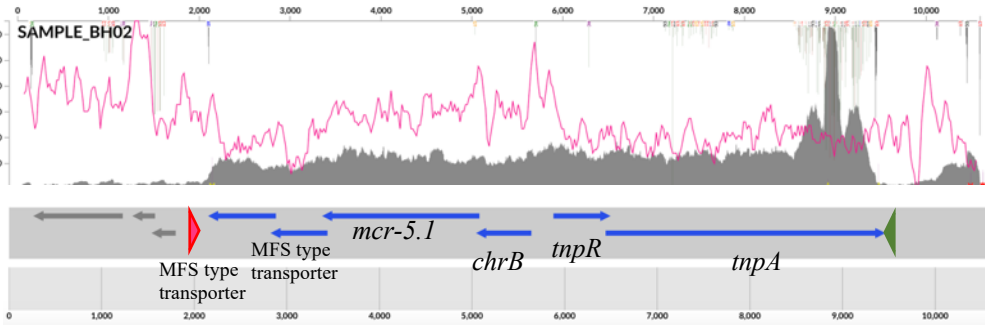
116

117

118

119

A

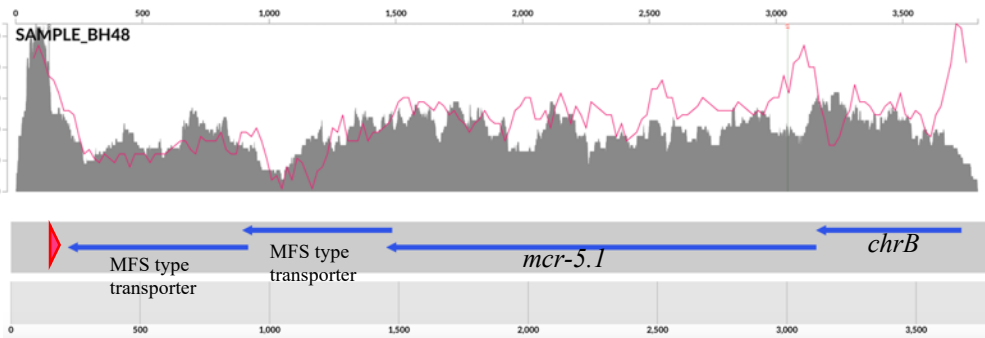


120

121

122

B

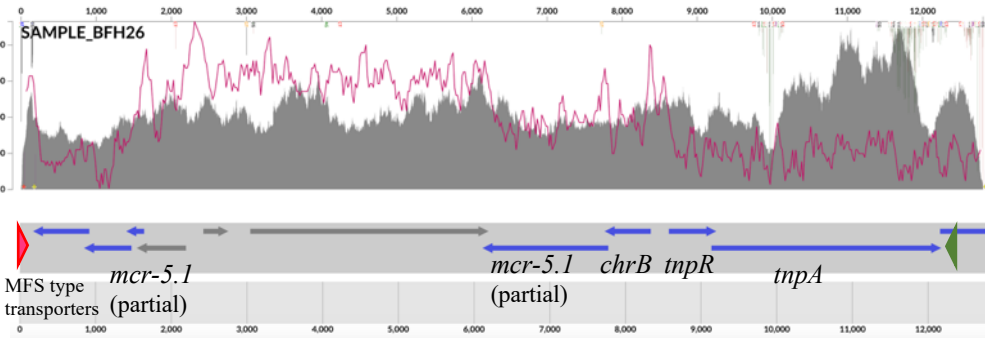


123

124

125

C

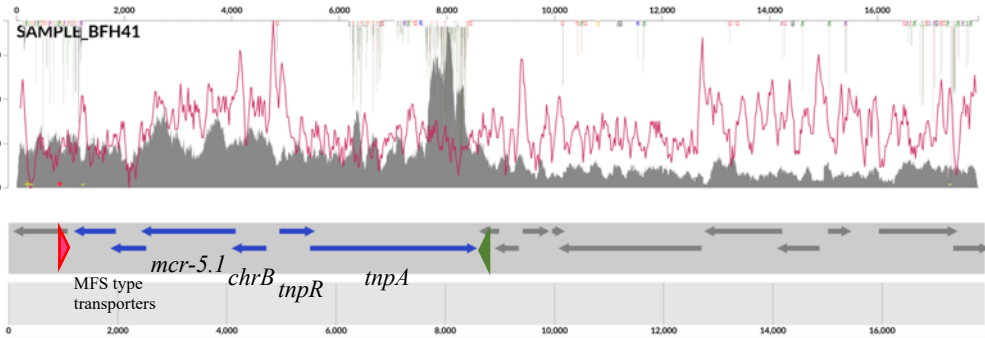


126

127

128

D



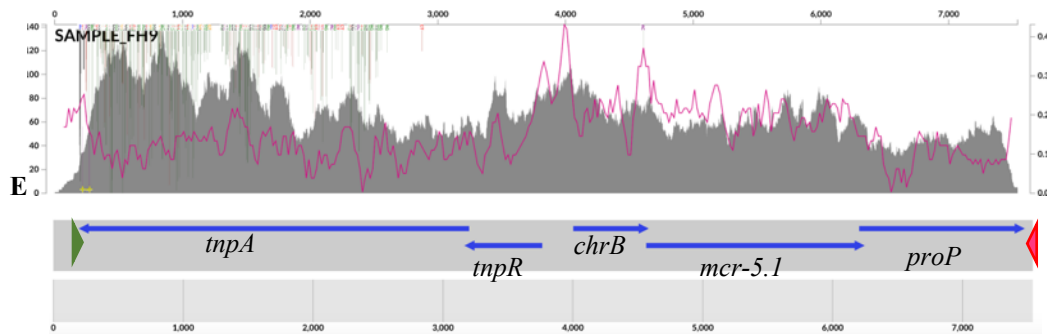
129

130

131

132

133



134

135 **Figure S7.** The genetic environments of *mcr-5* genes in four different metagenomic samples A) BH02 B) BH48,
 136 C) BFH26 D) BFH41, and E) FH9. Metagenomic contigs were assembled with MEGAHIT (v1.2.8) with
 137 parameters --min-contig-len 1000 -m 32000000000³. Visualizations of the contigs were created using Anvi'o
 138 (v7) metagenomics workflow for single profiles⁷. Blue arrows indicate putative encoding genes. The gene names
 139 were determined based on BLAST search results. The recruited reads are shown as a dark grey histogram with an
 140 overlay line (pink) denoting the GC content. Single nucleotide variation is shown on top of the panel with colorful
 141 vertical lines. Inverse repeat sequences of Tn3-like element sequences (38 bp) are shown with dark pink (IRL: 5'
 142 GGGGTCGTCTCAGAAAACGGAAAAATCGTACGCTAAG 3'⁸) and green triangles (IRR: 5'
 143 CTTAGCGTACGATTTTTTCCGTTTTCTAGACGACCCC 3'⁸).

144

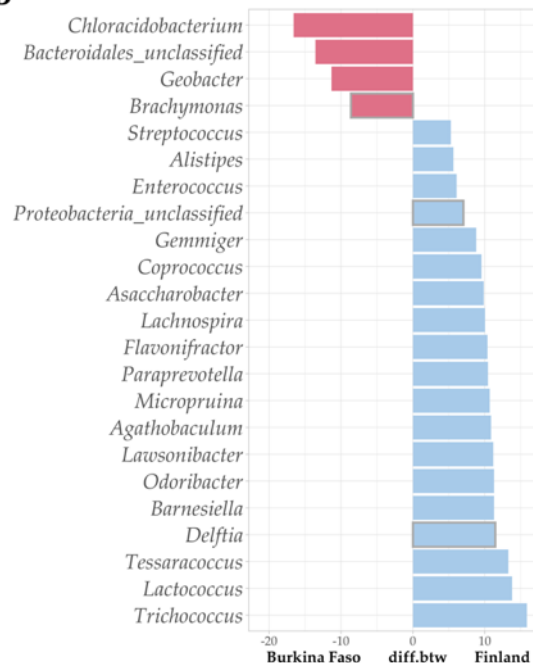
Divergent taxa in HWWs (ALDEx2, clr-transformation)

A

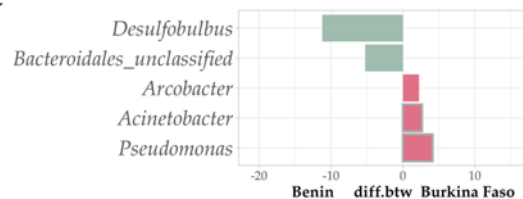


145

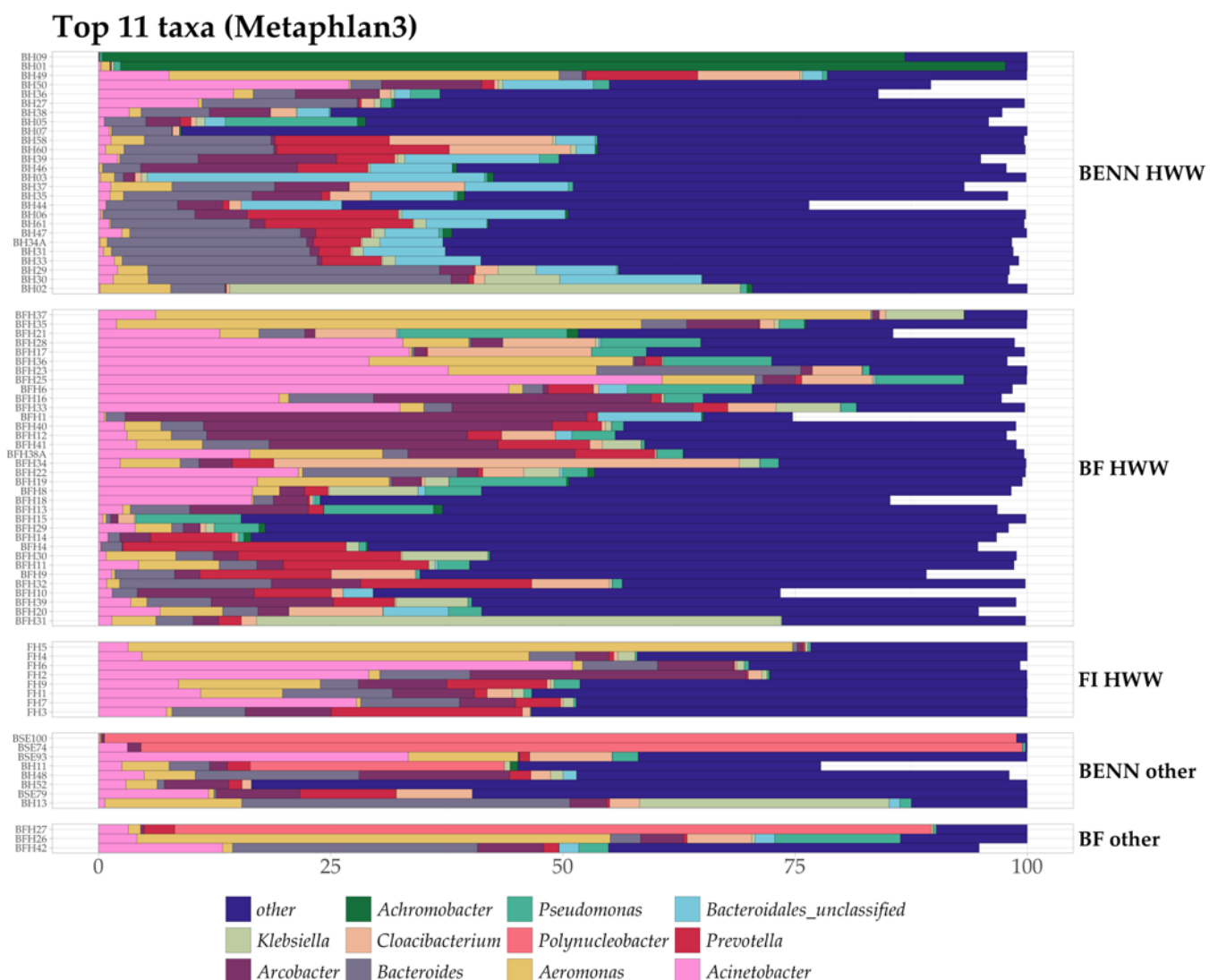
B



C

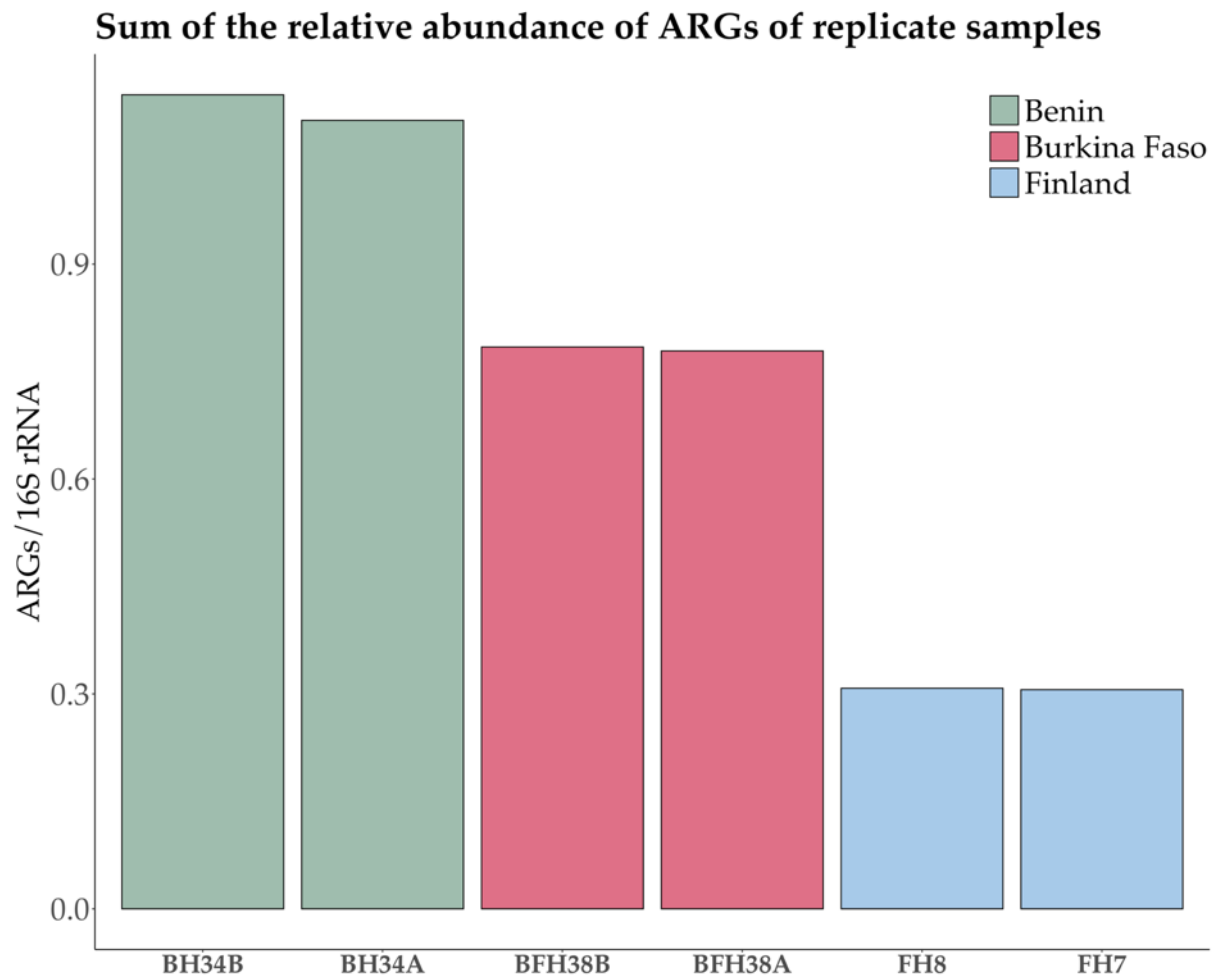


146 **Figure S8.** Divergent barplot shows the differentially abundant taxa in HWW from Benin, Burkina Faso, and
 147 Finland defined by the ALDEx2 tool ⁹ (v1.28.0). Metaphlan3 count data (obtained by running Metaphlan3 with
 148 the parameter ‘-t rel_ab_w_read_stats’) was clr-transformed in the analysis. The pairwise comparisons were A)
 149 Benin – Finland, B) Burkina Faso – Finland, and C) Benin – Burkina Faso. Significantly more abundant taxa in
 150 one country have positive ‘diff.btw’ values, while taxa significantly more abundant in the other country have
 151 negative ‘diff.btw’ values. Significantly differentially abundant taxa (BH corrected *p*-value < 0.05) were filtered
 152 for visualization to include only taxa with ‘diff.btw’ value < -5 or > 5, except for the Benin-Burkina Faso, in
 153 which all significant results are shown in the plot. All unfiltered results are listed in Tables S8, S9, and S10.
 154 Countries are denoted with distinct colors (Benin: green; Burkina Faso: red; Finland: blue). Genera belonging to
 155 the phylum *Pseudomonadota* are highlighted with a grey borderline to show the genera considered typical carriers
 156 of MGEs such as plasmids and integrons and contribute to the carriage of ARGs ¹⁰.
 157



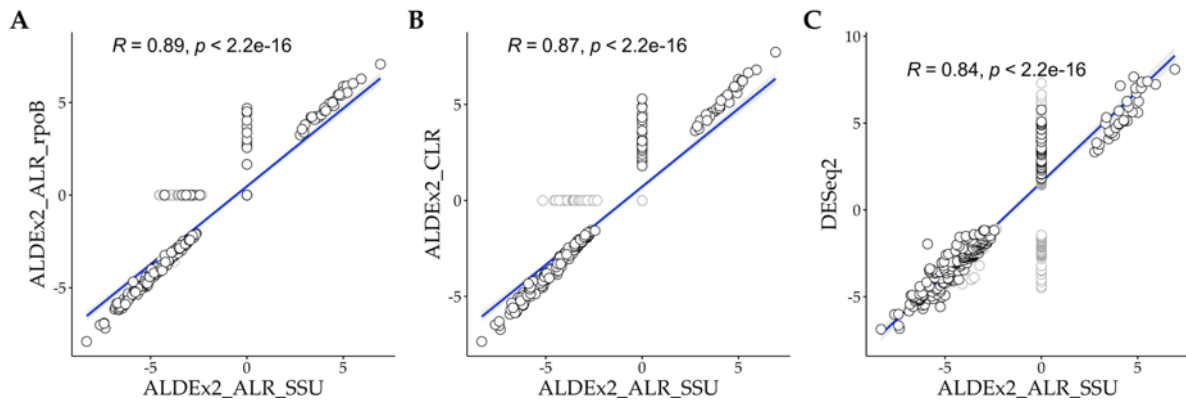
158
 159 **Figure S9.** Top 11 most abundant genera across all collected samples (HWW and other water samples). Sample
 160 codes are on the left, and the sample groups are on the right side of the plot (BENN HWW = HWW from Benin,
 161 BF HWW = HWW from Burkina Faso, FI HWW = HWW from Finland, BENN other (from top to bottom) =
 162 BENN river P (drinking), BENN river Q (drinking), BENN tap water S (drinking), BENN well water A (drinking),

163 BENN puddle at yard C, BENN hand-washing C, BENN river R (drinking), BENN street gutter, BF other (from
164 top to bottom) = BF wetland receiving treated HWW H, BF exit after biological treatment H, BF receiving river
165 after WWTP B). Relative abundances were obtained using Metaphlan3 ¹¹ and visualized with 'comp_barplot'
166 from package 'microViz' ⁶ without any transformation ('tax_transform_for_plot = "identity"'). Samples are
167 ordered according to Bray-Curtis distance.
168



169
170 **Figure S10.** The sum of the relative abundance of ARGs in the replicated samples. Sample pairs from Benin
171 (samples BH34 and BH34B), Burkina Faso (samples BFH38A and BFH38B), and Finland (samples FH7 and
172 FH8) were collected from the same septic tanks (Benin and Burkina Faso) or drains (Finland) at the same time.
173

Correlations between results using different approaches to study divergent ARGs



174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

Figure S11. Significantly differentially abundant ARGs were studied using different approaches, and the results were compared with each other. The results from ALDEx2⁹ (v1.28.0) using 16S rRNA as the reference were compared to A) those obtained using the *rpoB* gene as a reference, B) those without a reference, and to C) the results from Differential gene expression analysis based on the negative binomial distribution (DESeq2)¹² (v1.36.0). The *rpoB* counts were obtained using the translated sequence reads described above as input for hmmsearch¹³ against a *rpoB*-specific HMM profile (pf04563) from the Pfam-A database¹⁴. Hits or forward reads (R1) were counted. In DESeq2 analysis, ARGs filtered using a variance threshold of 50 were screened for significantly different ARGs between the compare groups after Benjamini-Hochberg adjustment ($p < 0.05$). The DESeq2 algorithm is designed for unnormalized count data from gene expression studies. Thus, to take into account the amount of bacterial DNA in the samples (counts of detected 16S rRNA), the relative abundances were multiplied by 10^5 , and a pseudo count of one was added to avoid zero values unsuitable for integer transformation. HWW from Benin and Finland were included in these comparisons. Insignificant results from one of the two compared approaches were set to zero. Overall, the results conducted using each approach generated similar results considering the specific ARGs but the number of significantly differentially abundant ARGs reported by one method differed from another (zero data points in the correlations). The most drastic difference in the results was seen between ALDEx2 with SSU as reference and DESeq2 (C).

3. References

1. Oksanen J, Blanchet G, Friendly M, et al. vegan: Community Ecology Package. <https://cran.r-project.org/package=vegan>.
2. Kassambara A. ggcorrplot: Visualization of a Correlation Matrix using “ggplot2”. <https://cran.r-project.org/package=ggcorrplot>.
3. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*. 2015;31(10):1674–1676. doi:10.1093/bioinformatics/btv033
4. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics*. 2015;31(20):3350–3352. doi:10.1093/bioinformatics/btv383
5. Bortolaia V, Kaas RS, Ruppe E, et al. ResFinder 4.0 for predictions of phenotypes from genotypes. *Journal of Antimicrobial Chemotherapy*. 2020;75(12):3491-3500. doi:10.1093/jac/dkaa345
6. Barnett D, Arts I, Penders J. microViz: an R package for microbiome data visualization and statistics. *J Open Source Softw*. 2021;6(63):3201. doi:10.21105/joss.03201
7. Murat Eren A, Kiefl E, Shaiber A, et al. Community-led, integrated, reproducible multi-omics with anvi'o. *Nat Microbiol*. 2021;6:3-6. doi:10.1038/s41564-020-00834-3
8. Borowiak M, Fischer J, Hammerl JA, Hendriksen RS, Szabo I, Malorny B. Identification of a novel transposon-associated phosphoethanolamine transferase gene, *mcr-5*, conferring colistin resistance in d-

211 tartrate fermenting *Salmonella enterica* subsp. *enterica* serovar Paratyphi B. *Journal of Antimicrobial*
212 *Chemotherapy*. 2017;72(12):3317-3324. doi:10.1093/jac/dkx327

213 9. Fernandes AD, Reid JNS, Macklaim JM, McMurrugh TA, Edgell DR, Gloor GB. Unifying the
214 analysis of high-throughput sequencing datasets: Characterizing RNA-seq, 16S rRNA gene sequencing
215 and selective growth experiments by compositional data analysis. *Microbiome*. 2014;2(1).
216 doi:10.1186/2049-2618-2-15

217 10. Rice EW, Wang P, Smith AL, Stadler LB. Determining Hosts of Antibiotic Resistance Genes: A
218 Review of Methodological Advances. *Environ Sci Technol Lett*. 2020;7(5):282-291.
219 doi:10.1021/acs.estlett.0c00202

220 11. Beghini F, McIver LJ, Blanco-Míguez A, et al. Integrating taxonomic, functional, and strain-level
221 profiling of diverse microbial communities with biobakery 3. *Elife*. 2021;10. doi:10.7554/eLife.65088

222 12. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data
223 with DESeq2. *Genome Biol*. 2014;15:550. doi:10.1186/s13059-014-0550-8

224 13. Eddy SR. Accelerated Profile HMM Searches. *PLoS Comput Biol*. 2011;7(10):1002195.
225 doi:10.1371/journal.pcbi.1002195

226 14. El-Gebali S, Mistry J, Bateman A, et al. The Pfam protein families database in 2019. *Nucleic Acids Res*.
227 2018;47:427-432. doi:10.1093/nar/gky995

228