1   **A preliminary investigation of major bacterial antibiotic resistance genes**

2   **present in Australian fecal samples using shotgun metagenomics and**

3   **targeted sequencing.**

4

5   Vanina Guernier-Cambert [1,2,†], Anthony Chamings [1,2], Fiona Collier [1,2,3], and Soren Alexandersen
6   [1,2,3]*
7

8   [1] Geelong Centre for Emerging Infectious Diseases, Geelong, VIC, Australia

9   [2] School of Medicine, Deakin University, VIC, Australia

10 [3] Barwon Health, University Hospital Geelong, VIC, Australia

11

12

13

14

15 [†] Current address: National Animal Disease Center, Agricultural Research Service, United States

16 Department of Agriculture, Ames 50010, IA, USA

17

18 *Corresponding author

19 Email: soren.alexandersen@deakin.edu.au

20

21

22

23 **Abstract**

24

25 The gut microbiota is an immense reservoir of antimicrobial resistance genes (ARGs); however, in

26 Australia the profile of the gut 'resistome', or ensemble of ARGs, has not been investigated. This

27 study provides a first preliminary mapping of the major bacterial ARGs present in human, domestic

28 dog and wild duck fecal samples collected from south-eastern Victoria, Australia; and evaluates

29 the use of shotgun metagenomics sequencing (SMS) and targeted amplification of ARGs. We

30 analysed SMS data using an in-house method and web-based bioinformatics tools: ResFinder and

31 KmerResistance. We examined targeted sequences using One Codex or the PanBacterialAnalysis

32 Torrent Suite plugin. All methods detected ARGs in all samples, with resistance to up to 13

33 classes of antibiotics detected overall. ARGs were more abundant in the human and dog samples

34 than the duck samples. They mostly conferred resistance to three classes of antibiotics that are

35 the most frequently prescribed in Australia: tetracycline, β-lactams and MLS_B (macrolide,

36 lincosamide, streptogramin B). Targeted sequencing significantly improved sensitivity for detection

37 of ARGs included in the panel; however, SMS provided quantitative information and allowed

38 tentative identification of the host bacteria. For SMS, web-based and in-house methods gave

39 comparable results, with discrepancies mostly due to different reference databases. The in-house

40 method allowed manually checking results and potential errors, while web-based methods were

41 user-friendlier and less time-consuming. More samples need to be investigated to fully describe

42 the resistome in humans and animals in Australia.

43

Introduction

Antimicrobial resistance (AMR) occurs naturally in bacteria and other microorganisms, but increases when antimicrobials (antibiotics, antivirals, antifungals) are used[1]. According to WHO, AMR accounts for an estimated 700,000 deaths per year and is one of the ten threats to global health in 2019[2]. Reservoirs of multi drug-resistant (MDR) bacteria are ubiquitous, and the concept of an "antibiotic resistome" was first coined to describe the large collection of bacterial antimicrobial resistance genes (ARGs) found in a specific environment[3]. Although AMR is most clinically relevant in pathogenic bacteria, ARGs exist in both pathogenic and non-pathogenic bacteria, and large reservoirs of these genes exist in all ecosystems, including in commensals at all sites of the human or animal body[4].

The adult human gastrointestinal tract harbours up to 1,000 molecular species or phylotypes often referred to as operational taxonomic units (OTUs) in 16S rRNA metagenomics[5,6]. This vast array of resident bacteria responds to environmental conditions inside the host, including altered dietary intake and antibiotic-induced disturbances. That can result in modifications in the microbial community structure and genetics, including the enrichment of ARGs within the gut microbiota. Recent studies have suggested potential links between the gut resistome and the spread of AMR: horizontal transfer of ARGs from commensal bacteria mediated by mobile genetic elements may affect potentially pathogenic organisms[7-9]. Although the development of culture-independent sequencing techniques has expanded our knowledge of the different reservoirs of ARGs, it is currently unclear how widespread commensal bacteria carrying ARGs are in the gut microbiota, and how they develop over time[10].

There is however considerable evidence that antibiotic exposure promotes the development of a resistome in the gut microbiota[11-13]. Different features such as class, potency, spectrum and regimen of the antibiotics can influence the observed pattern of resistance of the gut microbiota[13]. Geographical differences also exist between countries in relation to the ARGs detected in human fecal samples[14]. For example, the number and abundance of ARGs was higher in the fecal samples of Chinese individuals than in the Spanish and Danish groups studied[15]. Variation in the total usage of antibiotics between countries may be one explaining factor for those differences. Indeed, both antibiotic exposure and the number of ARGs were higher in individuals from Spain, France and Italy than in individuals from Denmark, Japan and the United States[16]. In other countries such as Australia, very little is known about the resistome of humans and animals, with available studies focusing primarily on ARGs in pathogenic bacteria (e.g. *Campylobacter* or *Escherichia coli*) or in environmental samples, e.g. studying the soil resistome after swine, cattle or poultry manure usage[17-19]. One recent study however assessed the diversity and abundance of ARGs transcribed in the fecal microbiota of water birds of Australia[20], but currently no study focused on the human fecal microbiota and its associated resistome in Australia. This clearly represents a significant knowledge gap. Regular surveillance of the resistome in Australia could be used to monitor trends in AMR over time and the emergence of resistance alleles. Resistome

87   surveillance could also inform future prescribing guidelines for doctors in hospitals and community

88   clinics, and veterinarians in companion and production animal settings, to help preserve the

89   effectiveness of valuable antibiotics.

90

91   The declining costs of metagenomic sequencing technologies has led to increased molecular

92   investigation of ARGs and has enabled a shift from phenotype to genotype-based approaches to

93   investigate AMR[21,22]. The ultimate goal of ARG sequencing and related bioinformatic analyses is

94   the accurate detection of the resistome and prediction of the antibiogram from genomic and

95   metagenomic data[23]. Several major databases provide collections of ARGs covering broad

96   categories of AMR mechanisms including the Comprehensive Antibiotic Resistance Database

97   (CARD)[24], Antibiotic Resistance Gene-ANNOTation (ARG-ANNOT)[25], the Bacterial Antimicrobial

98   Resistance Reference Gene Database[26] hosted by the National Center for Biotechnology

99   Information (NCBI) as part of the National Database of Antibiotic Resistant Organisms (NDARO),

100   and ResFinder[27] (see [23] for a more exhaustive list of available AMR bioinformatics software,

101   databases, and data-sharing resources). These resources and their associated prediction tools

102   enable detection and annotation of specific AMR elements in sequenced samples by comparing

103   user-submitted sequences against representative sets of ARGs sequences. While identifying

104   ARGs from a bacterial isolate from a sick patient is reasonably straightforward using whole

105   genome sequencing, the exhaustive description of a resistome in complex environments with high

106   bacteria diversity and relatively low abundances of ARGs is challenging. Two different strategies

107   have been developed to study AMR in complex environments like the gut. Deep shotgun

108   metagenomic sequencing is used to produce millions of reads that are filtered to map the ARGs.

109   However, because reads corresponding to ARGs can be low (e.g. less than 2% of the total

110   sequencing data in [28]), targeted approaches (including an initial amplification of a panel of ARGs)

111   with increased sensitivity have been developed. As new methods and new commercial panels

112   become available, it is currently unclear what method is best to study resistomes when no pre-

113   existing information is available, e.g. in Australia.

114

115   We investigated for the first time the diversity and abundance of ARGs in a broad range of fecal

116   microbiota samples (a number of infants, healthy pregnant women, domestic dog, wild ducks)

117   collected in south-eastern Victoria, Australia. We used both non-targeted NGS (analyzed in-house

118   and with web-based tools) and targeted NGS (using two commercially available panels) and

119   assessed and compared the performance of these tools for the analysis of bacterial resistomes in

120   an Australian context. This preliminary study will pave the way for a broader analysis of bacterial

121   resistomes in humans and animals in Australia.

122

123   **Methods**

124

125   **Sample collection and processing**

126

127 *Sample collection*

128 Human and animal samplings were performed in accordance with relevant guidelines and

129 regulations. We collected stool samples from two infants and seven adults from the southeast of

130 Australia (Table 1). The children were: one infant sampled at 4 weeks of age (ST5-1mo) and again

131 at 18 months (ST5-18mo), and another infant sampled at 3 months of age (ST4-3mo). The study

132 involving these samples was deemed negligible or low risk by the Barwon Health Human

133 Research Ethics Committee and therefore exempt from full committee review (HREC approval

134 17/119). The adult samples originated from pregnant women at 36-weeks gestation between 2010

135 and 2013 with informed consent for study participation as part of the Barwon Infant Study (BIS), a

136 birth cohort study–eligibility criteria and cohort profile are described elsewhere[29]. Ethics approval

137 for this study was granted by the Barwon Health Human Research and Ethics Committee (HREC

138 10/24). All human stool samples were stored at -80ºC prior to DNA extraction.

139

140 **Table 1. List of tested human and animal samples** (n = 13). Fecal samples were DNA extracted

141 using two different methods, as described.

Sample ID	Host	Extraction method
ST5-1mo	Human - child 1 month of age	PowerSoil® DNA Isolation Kit - Mo Bio
ST4-3mo	Human - child 3 months of age	QIAamp Fast DNA Stool Mini Kit
ST5-18mo	Human - child 18 months of age	QIAamp Fast DNA Stool Mini Kit
HS21	Human - adult	PowerSoil® DNA Isolation Kit - Mo Bio
HS22	Human - adult	PowerSoil® DNA Isolation Kit - Mo Bio
HS23	Human - adult	PowerSoil® DNA Isolation Kit - Mo Bio
HS24	Human - adult	PowerSoil® DNA Isolation Kit - Mo Bio
HS25	Human - adult	PowerSoil® DNA Isolation Kit - Mo Bio
HS26	Human - adult	PowerSoil® DNA Isolation Kit - Mo Bio
HS28	Human - adult	PowerSoil® DNA Isolation Kit - Mo Bio
ST1-MAD*	Pool from 6 Pacific black ducks (*Anas superciliosa*)	QIAamp Fast DNA Stool Mini Kit
ST3-MUD*	Muscovy duck (*Cairina moschata*)	QIAamp Fast DNA Stool Mini Kit
DFS**	Pool from 2 dogs	QIAamp Fast DNA Stool Mini Kit

142 HS: human sample; DFS: dog fecal sample. * See [30]. ** See [31].

143

144 Samples from ducks and dogs were collected in the south-eastern part of Victoria, Australia, as

145 part of other projects investigating fecal microorganisms. Fresh fecal samples from one Muscovy

146 duck (MUD) and a pooled sample from six juvenile Pacific black ducks (MAD) were collected in

147 November and December 2016 respectively[30]. Bird sample collection was approved under Deakin

148 University's Animal Ethics Committee project number B43–2016 and Department of Environment,

149 Land, Water and Planning permit number 1008206.

150 A pooled dog fecal sample (DFS) originating from two pups (7 weeks of age) was collected from

151 the ground immediately after being deposited, with the consent of the owner[31]. These puppies

152 were healthy at the time of collection, but they had been sick just 1-2 weeks before and an

153 astrovirus was detected in the 7-week sample[31]. However, they were originally thought to have a

154 *Campylobacter* infection and were likely treated with antibiotics at 5 weeks of age. Swabs from

155 dog samples were placed in Universal Transport Medium (UTM) and stored on ice overnight. The

156 following morning, swabs were aliquoted and frozen at -80ºC until DNA extraction.

157

158 *DNA extraction*

159 DNA was extracted from fecal samples and swabs using the Qiagen Powersoil® DNA Isolation Kit

160 (Cat#12888-100) (QIAGEN Pty Ltd., Victoria, Australia) or the QIAamp Fast DNA Stool Mini Kit

161 with MoBio bead shearing (Mo Bio, Carlsbad, CA USA). The manufacturer's protocol was slightly

162 modified following optimization: for the inhibitors' removal steps, 375 μL of Solution C2 and 335 μL

163 of Solution C3 were used instead of the prescribed volumes. DNA concentrations were measured

164 using a NanoDrop™ spectrophotometer (Thermo Fisher Scientific, Scoresby, VIC, Australia), and

165 the quantity and size distribution of the fragments after DNA extraction was visualized on an

166 Agilent Bioanalyzer DNA 7500 chip (Agilent Technologies) using the High sensitivity DNA kit

167 (Agilent Technologies).

168

169 **Library preparation and sequencing**

170 Extracted DNA was sequenced using three different methods relying on Next Generation

171 Sequencing (NGS). We used a non-targeted sequencing of sheared DNA, hereafter referred to as

172 "Shotgun Metagenomic Sequencing" (SMS) method, as well as a targeted metagenomic

173 sequencing method using two commercially available Ion Ampliseq™ (PCR) panels (Life

174 Technologies).

175

176 *Non-targeted sequencing: Shotgun Metagenomic Sequencing*

177 Genomic libraries were prepared separately for each genomic sample from 100 ng of DNA. DNA

178 was fragmented using the Ion Shear™ Plus Reagents (Life Technologies, Grand Island, NY).

179 Aiming for a 200 base-read library, the chosen reaction time for the enzymatic incubation at 37°C

180 was 15 minutes, as advised in the manufacturer's protocol. Sheared DNA was purified with

181 Agencourt AMPure® XP Reagent (1.8x sample volume) (Beckman Coulter, Lane Cove, NSW,

182 Australia). The quantity and size of sheared material was visualized on an Agilent Bioanalyzer

183 DNA 7500 chip (Agilent Technologies) using the High sensitivity DNA kit (Agilent Technologies)

184 assuming a targeted 200 base-read library.

185 The adapter ligation and nick repair were performed using the Ion Plus Fragment Library kit and

186 Ion Xpress barcode adapters following the manufacturer's recommendations (Life Technologies).

187 Ligated and nick repaired DNA was purified with Agencourt AMPure® XP Reagent (1.4x sample

188 volume) (Beckman Coulter) assuming a targeted 200 base-read library. The ligated and nick

189 repaired DNA was size-selected individually with the E-Gel® SizeSelect™ Agarose Gel (Life

190 Technologies). The size selected (and unamplified) libraries were quantified as per the

191 manufacturer's instructions using the Ion Library qPCR Quantitation Kit (Life Technologies) to

192 determine if library amplification was required.

193 If required (i.e. final library quantity was < 50 pM), the size selected libraries were further amplified

194 using Platinum® PCR SuperMix High Fidelity and Library Amplification Primer Mix as per

195 manufacturer's instructions (Life Technologies). This was necessary for four samples (ST4-3mo,

196 MUD, MAD and DFS), which were amplified for 10 PCR cycles prior to purification and

197 quantification.

198 All libraries were standardized to a concentration of 100 pM. Libraries were then pooled and

199 diluted to 50 pM prior to loading onto Ion 530™ or 540™ Chips using the Ion Chef Instrument and

200 Ion 530™ or Ion 540™ Kit (Thermo Fisher Scientific). Following template preparation, the chips

201 were run on the Ion Torrent S5xl System (Thermo Fisher Scientific) as per company protocols.

202 NGS and associated reactions were performed at the Geelong Centre for Emerging Infectious

203 Diseases (GCEID), Geelong, Victoria, Australia.

204

205 *Targeted sequencing: IonAmpliSeq™ panels*

206 We used two commercially available community panels designed to target AMR determinants,

207 both kindly donated by Life Technologies. The first panel used was the Ion AmpliSeq™ Pan-

208 Bacterial Research panel, consisting of two primer pools: the one pool contained 269 amplicons

209 targeting 21 specific bacterial species and 716 amplicons targeting 364 known ARGs; the other

210 pool was comprised of 24 amplicons for 16S rRNA gene profiling of up to approximately 400,000

211 16S rRNA sequences. From this panel, we only used the first primer pool targeting the ARGs.

212 The second panel used was the Ion AmpliSeq™ Antimicrobial Resistance (AMR) Research panel

213 that also consisted of two primer pools comprising a total of 815 amplicons targeting 478 known

214 ARGs[32]. Both primer pools were used from this panel. The list of primers is available upon request

215 from Life Technologies (https://www.thermofisher.com/).

216 Ten nanograms of extracted DNA were PCR-amplified with the Ion AmpliSeq™ Library Kit (Life

217 Technologies) and Ion AmpliSeq™ Panels, both panels including a 5X Ion AmpliSeq™ HiFi Master

218 Mix. With the Pan-Bacterial Research panel, PCR conditions were as follows: enzyme activation at

219 99ºC for 2 minutes followed by 18 cycles of 99ºC for 15 seconds and 60ºC for 8 minutes before

220 holding at 10ºC. With the AMR Research panel, PCR conditions were as follows: enzyme

221 activation at 99ºC for 2 minutes followed by 19 cycles of 99ºC for 15 seconds and 60ºC for

222 4 minutes before holding at 10ºC. The following steps (adaptors and barcodes ligation, purification,

223 and qPCR quantification) were done as per the manufacturer's instructions. None of the

224 AmpliSeq™ libraries required further amplification at this stage. Libraries were then pooled prior to

225 loading onto an Ion 530™ Chip and loading into the Ion Chef Instrument. Following template

226 preparation, the chip was run on the Ion Torrent S5xl System (Thermo Fisher Scientific) following

227 company protocols. We submitted reads obtained from the SMS and targeted sequencing to the

228 European Nucleotide Archive (ENA) under project accession number PRJEB36405.

229

230 **Next generation sequence analyses**

231 We obtained three sets of sequence data that we analysed using different tools and methods. A

232 schematic flowchart summarizing the different sequencing and analysis methods is presented in

233 Figure 1.

234

235 **Figure 1. Flowchart describing three methods to identify the presence of the most common**

236 **ARGs in clinically healthy human/animal samples using NGS without culturing.** (A) Non-

237 targeted sequencing of sheared DNA or "Shotgun Metagenomic Sequencing" (SMS) method,

238 versus (B) and (C) targeted sequencing using two commercially available AmpliSeq™ panels

239 targeting ARGs. Sequences obtained with the SMS method were analysed using different web-

240 based tools (ResFinder and KmerResistance, https://cge.cbs.dtu.dk/services/), and an in-house

241 method based on the comparison of reads and contigs with a NCBI-curated ARGs database

242 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA313047). *Targeted sequencing was tested on

243 human samples only: 4 with the AMR Research panel, 8 with the Pan-Bacterial panel. ** refers to

244 Ion Torrent Suite plugins available from the Thermo Fisher Scientific Plugin store. *** refers to the

245 "One Codex" online platform (available at: https://app.onecodex.com/) used with the AMR

246 Research panel. "ID": percentage identity threshold; "minL": minimum length: minimum percentage

247 of coverage compared to reference length or minimum number of nucleotides (nt); "ref": reference

248 sequences used for comparison with reads and/or contigs.

249

250 *Analysis of the sequences obtained by Shotgun Metagenomic Sequencing*

251 Raw sequences were analysed using the Ion Torrent Server linked to the Ion Torrent S5xl

252 instrument, based on each sample having a unique barcode. We developed an "in-house method"

253 based on the comparison of our data with references in the NCBI-curated AMR gene database

254 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA313047) containing a total of 4,528 sequences

255 (downloaded on the 2nd of August 2018).

256 The reads were queried against the NCBI's ARGs database using Magic-BLAST[33] with

257 parameters as follows: 98% minimum identity between the read and the reference, and 70

258 nucleotides minimum length. Concomitantly, the Torrent Suite AssemblerSPAdes 5.6.0 plugin was

259 used to generate contigs from the reads for each sample[34]. These contigs were also queried

260 against the NCBI's ARGs database but using nucleotide BLAST (BLASTN)[35,36] with parameters as

261 follows: 80% minimum identity between the contig and the reference and 1e-10 minimum E-value.

262 The contigs identified to carry ARGs were queried against the "Microbes" reference database

263 using BLAST Genomes in order to tentatively identify the bacterial species carrying ARGs.

264 After identifying the ARGs present, FASTA files were created by adding chromosomes X, 21 and

265 mitochondrion from the human genome reference sequence (hg19) to each pre-selected ARG

266 sequence (listed in Supplementary Table S1). This was done in order to artificially increase the

267 length of the reference used for the mapping, as it is our experience that mapping issues may

268 arise from mapping reads against small references such as e.g. ARGs of only around 1,000 to

269 2,000 bp due to the scaling factor algorithm used by the Torrent Suite TMAP plugin. The extended

270 FASTA references were added to the Ion Torrent Server and used for the mapping of sequence

271 reads using the Torrent Suite TMAP plugin. Mapped reads and consensus sequences were

272 visualised and analysed using the Integrative Genomics Viewer (IGV)[37].

273

274 *ARGs coverage and depth*

275   Individual ARGs coverage and depth metrics were calculated using the Torrent Suite

276   CoverageAnalysis v5.6.0.1 plugin using the filters: library type = whole genome; minimum aligned

277   length = 100; minimum mapping quality = 20. The "average base read depth" obtained as an

278   output was used to standardize reads counts. A "mean base depth per 5 million reads" (or BD5M)

279   was calculated for each ARG in each sample following the formula: (average base read depth *

280   5,000,000) / total reads. Contrary to the common RPKM (number of reads per kb of gene per

281   million sequenced reads), BD5M is 'per base' and not 'per read' and its interpretation is

282   independent of the size of the reads or the reference. As an example, if 5 million reads were

283   obtained for a sample, BD5M = 2 for a specific gene is equivalent to reads mapping the whole

284   gene twice or reads mapping half of the gene four times. To avoid confusion between the raw

285   coverage from the sequencer and our calculated BD5M, we hereafter refer to the latest as an

286   abundance rather than a coverage.

287

288   *Additional AMR analyses*

289   We analysed the sequences obtained with the SMS method using two web-based tools,

290   KmerResistance (available at https://cge.cbs.dtu.dk/services/KmerResistance/)[38,39] and ResFinder

291   3.1 (available at https://cge.cbs.dtu.dk/services/ResFinder/)[27], in order to compare the results with

292   those obtained with our "in-house method" (Figure 1). Both tools rely on the same reference

293   database (available at https://cge.cbs.dtu.dk/services/data.php)[27]. Raw sequences (FASTQ reads)

294   from each sample were screened for ARGs present in the ResFinder database with the default

295   settings of 70% identity threshold and 10% depth correction using "bacteria" as the host database

296   and "resistance genes" as the gene database. Scoring method was "species determination on

297   maximum query coverage". Contigs (FASTA files) from each sample were also screened for ARGs

298   present in the ResFinder database using the settings of 90% minimum identity and 20% minimum

299   length.

300

301   *Microbiota analysis and host identification*

302   The Torrent Suite Ion Reporter™ software was used to process reads for metagenomics analysis

303   of the gut microbiota. Data files (Binary Alignment Map, BAM) generated by the Ion Torrent S5xl

304   with the SMS method were uploaded to Ion Reporter (https://ionreporter.thermofisher.com/ir/). All

305   analyses were run using a "150-90-3" setting, i.e. minimum read length = 150; minimum alignment

306   coverage between hit and query = 90%; minimum number of unique reads (or 'read abundance

307   filter') = 3. This means that only reads found as triplicates or more were included in the analysis. A

308   minimum percentage identity of 97% was required for genus identification of a ribosomal

309   sequence, and a minimum percentage identity of 99% for species identification. When there was a

310   difference of less than 0.2% match between the top hit and the next best hit, the sequence was

311   identified as 'species slash ID' (unidentified species). The relative abundance of the species/genus

312   identified in the gut microbiota was determined from Ion Reporter™.

313   The contigs generated from SMS allowed tentative identification of host bacteria carrying the

314   ARGs. This was achieved by running the contigs against the "Microbes" database using BLAST

315   Genomes or the ResFinder tool.

316   *Analysis of the sequences obtained by Ion AmpliSeq^{TM} panels*

317   Following targeted amplification, sequences obtained from the Ion Torrent S5xl were analysed

318   based on each sample having a unique barcode and using dedicated platforms or plugins, i.e. (i)

319   the "One Codex" web-based platform (available at: https://app.onecodex.com/) for the sequences

320   obtained from using the AMR Research panel, or (ii) the "PanBacterialAnalysis" Torrent Suite

321   plugin from the Ion Torrent Server for the sequences obtained from using the Pan-Bacterial

322   Research panel (see Figure 1). One Codex reported the percentage of the ARG sequence

323   covered by reads in the sample, the percentage identity to the reference ARG sequence, and the

324   depth of read coverage. The "PanBacterialAnalysis" plugin reported the "read counts" matching

325   each ARG in the database.

326

327   **Results**

328

329   All 13 samples were processed using the SMS method; with eight samples processed using the

330   Pan-Bacterial Research panel, and four of these eight also processed using the AMR panel (Table

331   2). SMS generated a total of 3.8 to 11 million reads per sample, with average read lengths of 162-

332   201 nucleotides (Table 2). Whatever the method, all samples had reads that mapped to ARGs,

333   with abundance and composition varying between samples. SMS overall identified ARGs to 2-11

334   classes of antibiotics in individual samples (Figure 2 and Table 3), with ARGs to two additional

335   classes identified using targeted sequencing methods.

336

337   **Table 2. Summary of NGS sequencing results using three methods:** (1) shotgun

338   metagenomics sequencing (SMS) or (2) targeted sequencing using community panels from Life

339   Technologies: (2a) the Ion AmpliSeq^{TM} Antimicrobial Resistance (AMR) Research panel and (2b)

340   the Ion AmpliSeq^{TM} Pan-Bacterial Research (PBR) panel. BC: barcode adapter identifier; Reads:

341   number of short read sequences per sample; MRL: mean read length in nucleotides. The four

342   human samples tested with all three methods are in dark grey, the four human samples tested with

343   two methods are in light grey and the five samples tested only with the SMS method are in white.

344

Sample ID	Host	SMS method			AMR panel		PBR panel	
		BC	Reads	MRL	BC	Reads	BC	Reads
ST4-3mo	Human - child	BC062	5,057,157	184	-	-	-	-
ST5-1mo	Human - child	BC002	10,998,563	201	BC049	1,486,710	BC053	505,192
ST5-18mo	Human - child	BC063	5,656,858	195	-	-	-	-
HS21	Human - adult	BC021	3,829,581	182	BC050	523,118	BC054	524,591
HS22	Human - adult	BC022	5,721,672	184	-	-	BC055	377,819
HS23	Human - adult	BC023	9,313,335	197	-	-	BC056	262,060
HS24	Human - adult	BC024	7,303,016	195	BC051	709,047	BC057	491,234

HS25	Human - adult	BC025	6,986,354	193	-	-	BC058	174,895
HS26	Human - adult	BC026	9,411,554	198	BC052	730,272	BC059	431,487
HS28	Human - adult	BC028	7,952,415	180	-	-	BC060	110,975
ST1-MAD	Pacific black ducks	BC065	5,671,145	162	-	-	-	-
ST3-MUD	Muscovy duck	BC066	8,438,563	169	-	-	-	-
DFS	Pool from 2 dogs	BC070	8,570,346	192	-	-	-	-

Figure 2. Overview of antimicrobial resistance genes (ARGs) abundance and composition in human and animal samples (n = 13) using the shotgun metagenomics sequencing (SMS) method. (A) Major *versus* (B) minor ARGs. ARGs abundance is calculated as a "mean base depth per 5 million reads" (BD5M) for each reference gene in each sample, and then summed for ARGs belonging to the same class of antibiotics.

351 **Table 3. Antimicrobial resistance genes (ARGs) identified from the shotgun metagenomics sequencing (SMS) using the "in-house" analysis** (based

352 on queries against the NCBI's "Bacterial Antimicrobial Resistance Reference Gene Database" BioProject PRJNA313047)**.** Mean depth coverage per 5 million

353 reads (BD5M) between 1 and 10 is highlighted in yellow, BD5M > 10 is highlighted in red. Additional ARGs identified with the ResFinder and/or

354 KmerResistance tools but not the in-house method are not listed here (see text).

355

Class of antibiotics	Gene	ST4-3mo	ST5-1mo	ST5-18mo	HS21	HS22	HS23	HS24	HS25	HS26	HS28	MAD	MUD	DFS
Aminoglycosides	aac(6')	0	0	0	0	0	0.9R	0	0	0	0	0	0	0
	aadA1	0	0	0	0	0	0	0	0	0	0.9	9*	1.2	0
	aadE (or ant6-Ia)	0	0	0	6.7KR	1.3K	3.4R	10.2KR	0.4	1.7KR	2.6K	0	0	1.1R
	aadS §	0	0	0	0	0	1.0§	41.7§	0	0	0	0	0	18.2§
	spw for ant9 §	0	0	0	0	0	0.1§	8.9§	0	2.2§	0	0	1.3§	0
	aph(3')-Ia	0	0	0	0	0	0	0	0	0	0	0	0	2.7R
	aph(3')-IIIa	0	0	0	1.0	1.4	1.2	0.2	1.0R	0.1	0.2	0	0	0
	aph(6)-Id	0	0	0	0	0	0	0	0	0	0	0	0	0.6
	aac(6')-Ie / aph(2'')-Ia §	0	0	0.5§	0	0.1§	0	0	0	0.9§	0.2§	0	0	0.1§
Beta-lactams	blaEC §	46.8§	95.9§	0	2.9§	0	0.4§	0	0.6§	0.3§	0	0.5§	0.1§	0
	blaOXA	0	0	0	0	0	0	0	0	0	0	0	1.2K	0
	blaTEM	53.7KR	0	0	0	0	0	0	0	0.1	0	0.2	0	2.7
	cblA §	0	0	32.4§	1.6§	4.4§	0.7§	0	1.2§	0.5§	2.8§	0	0	0
	cepA	0	69.5KR	11.8KR	0	0.3	0.1	0.9	0.1	0.2	0	0	0	0
	cfxA	0	0	35.9KR	16.2K	4.0K	1.7R	0	1.5	1.5R	0	0	0.5	66.5
Chloramphenicol	catD §	0	0	0.2§	0	0.7§	4.6§	0	0	0.1§	0.8§	0	0	0
	catP	0	0	0.2	0	0.7	6.4	0	0	0	0.8	0	0	0
Fosfomycin	fosA	1.0R	0	0	0	0	0	0	0	0	0	0	0	0
Glycopeptides	vanG §	0	0	0	0	0.5§	0	0	0	0	0	0	0	0
	vanS-D	0	0	0.4	0	0.1	0.2	1.1	0.2	0	0	0	0	0
	vanY-D	0	0	0.7	0	0	0	0	0	0	0	0	0	0
MLS_B	eat(A) §	0	0	0	0	0	1.0§	0	0	0	0	0	0	0.1§
	ere(D)	0	0	0	0	0	1.0R	0	0	0	0	0	0	0
	erm(B)	0	0	1.4	0.3	2.8KR	1.2R	20.8KR	5.2KR	21.4KR	13.4KR	0	0.2	0
	erm(F)	0	0	0	0	0	0	29.4KR	0.3	0	0	0.2	4.2KR	0
	erm(G)	0	0	0	0	0	0.1	0	0	0	0.1	0	0.5	26.4KR
	erm(Q)	0	0	0	0	0	0	0	0	0	0	0	2.2KR	0.6
	lnu(A)	0	0	0	0	0	0	0	0	0	0	0	0	2.3R
	lnu(AN2) §	0	63.2§	29.5§	0	0.7§	1.2§	14.6§	0.2§	25.0§	1.0§	0	1.1§	56.6§
	lnu(C)	0	0	0	0	1.1	0.4	0	4.9KR	0.7	0.2	0	5.8KR	64.2KR

Category	Gene													
	lnu(P)	0	0	0	0	0	0	0	0	0.3	0	0	3.0[KR]	0
	lsa(A)	0	0	0	0	0	0	0	0	0	0	6.1[KR]	0	0
	mef(A)	0.5	1.1[KR]	0.1	0	0.1	0.7	14.4[KR]	0.6	3.7[KR]	0.1	0	10.0[KR]	20.5[KR]
	mef(En2) [§]	0	61.8[§]	32.8[§]	0.4[§]	0.3[§]	1.3[§]	15.7[§]	0.2[§]	17.3[§]	2.2[§]	0	2.7[§]	65.9[§]
	msr(C)	0	0	0	0	0	1.3[R]	0	0	0	0	0	0	0.3
	msr(D) (or mel)	0.3	0.6[KR]	0.2	0	0.1	0	0	0.1	0.1	0.1	0.1	1.3	23.9[KR]
Quinolone	oqxA	0.6	0	0	0	0	0	0	0	0	0	0	0	0
Sulfonamide	sul1	0.2	0	0	0	0	0	0	0	0	0.4	0	0	0
	sul2	0	0	0	0	0	0	0	0.1	0	0	0	0	0.8[R]
Tetracyclines	tet(32)	0	0	24.8[KR]	3.5[K]	12.1[KR]	43.9[KR]	14.7[R]	12.5[K]	25.4[KR]	38.0	0	0.3	7.2
	tet(40)	0	0	3.3[KR]	1.6	9.1[KR]	2.6[R]	7.5[KR]	1.4	13.0[KR]	4.1[KR]	0	0.1	1.7[R]
	tet(44)	0	0	0	0.2	0.2	0.1	0.3	0.4	1.1	1.4[R]	0	0.1	0.7
	tet(A)	279.8[KR]	0	0	0	0.1	0	0	0	0	0.2	0	0	0
	tetA(P)	0	3.4[KR]	0	0.7	0.2	0	0	0	0.3	0	0	40.8[KR]	1.1[R]
	tetB(P)	0	0	0	0.6	0.4	0.6	0	0	0.3	0	0	42.3[KR]	1.2[R]
	tetB(46)	0.2	0	0	0.2	0.5	0	0	1.3[R]	0.1	0	0	0	0
	tet(C)	0	0	0	0	0	0	0	0	0	0.6[R]	0	0	0
	tet(M)	0.2	0	1.4[R]	0.2	0.7	1.5	0.9	0.2	1.5	2.0	0	0.1	0.9[R]
	tet(O)	0.1	0.1	34.2[K]	5.2	16.1[KR]	73.9[KR]	19.4[KR]	5.9[KR]	29.8[K]	67.5[KR]	0	0.5	15.6[R]
	tet(Q)	0.2	54.5[KR]	52.9[KR]	29.8[KR]	33.7[KR]	6.2[KR]	68.4[KR]	9.7[K]	11.7[KR]	2.7[KR]	0	5.1[KR]	14.9[KR]
	tet(W)	0	0	13.3[KR]	7.4[KR]	43.3[KR]	76.6[KR]	55.6[KR]	43.7[KR]	13.1[KR]	67.5[KR]	0	3.8[KR]	12.2[KR]
	tet(X)	0	0	0	0	0	0	96.2[KR]	0	0	0	0	0	20.5[KR]
Trimethoprim	dfrF [§]	0	0	17.6[§]	0	2.2[§]	2.0[§]	1.9[§]	2.9[§]	5.3[§]	0.9[§]	0	0	0.2[§]
Trimethoprim/ streptothricin	estX/sat1 [§]	0	0	0	0	0	0	0	0	0	0	0	1.5[§]	0

356 HS: human sample; MAD: Pacific black duck; MUD: Muscovy duck; DFS: dog fecal sample. [K] ARGs detected from the SMS that were also detected with the KmerResistance tool. [R] ARGs detected
357 from the SMS that were also detected with the ResFinder tool. [§] ARGs that were absent from the ResFinder database. * Gene aadA1 was calculated to be present in bird MAD at BD5M = 9 but the
358 mapping shows that only 23% (226 / 989 bp) of the gene (reference NG_047324) is present.

359 **SMS-based analysis of ARGs**

360 All 13 samples were processed using the SMS method, allowing comparison of ARGs diversity

361 between the human samples; three young infants and seven adults; as well as the pooled

362 domestic dog fecal sample (DFS), and two bird samples (MAD being a pool of six birds). We

363 identified ARGs from both the reads and the contigs (see Table S1).

364

365 *Resistance - antibiotic classes*

366 Using the "in-house method", we detected genes conferring resistance to 11 classes of antibiotics

367 (out of 16 classes included in the NCBI's ARGs database). ARGs conferring resistance to at least

368 one class of antibiotics at BD5M >50 were found in 12 samples (Figure 2 and Table 3). ARGs

369 conferring resistance to tetracycline (BD5M 49 to 280), macrolide, lincosamide, streptogramin B

370 (MLS$_B$) (BD5M 0.7 to 260), β-lactams (BD5M 1.9 to 165) and aminoglycoside (BD5M up to 61)

371 were found at higher abundances in all 13 samples (Figure 2A, Table 3). ARGs to six other

372 classes of antibiotics were found in ten samples at lower abundance (Figure 2B): trimethoprim

373 (BD5M up to 17.6), chloramphenicol (BD5M up to 11), glycopeptides, quinolones, fosfomycin and

374 sulfonamide (all last four classes with BD5M ≤ 1) (Figure 2B and Table 3). An eleventh class of

375 antibiotics was only detected as a bifunctional gene estX/sat1 conferring resistance to

376 trimethoprim/streptothricin.

377 The wild ducks (MAD) and the 1-month old human (ST5-1mo) carried genes conferring resistance

378 to the fewest classes of antibiotics –three classes (or two classes at BD5M ≥ 1 for sample MAD)–

379 while HS23, HS24, duck MUD and dog DFS carried ARGs attributable to the most classes of

380 antibiotics, i.e. seven classes (or six classes at BD5M ≥ 1) (Table 3).

381

382 *Resistance - gene level (ARGs)*

383 At the gene level, 62 ARGs were detected from the SMS-generated reads: 54 using the "in-house

384 method" (or 45 at BD5M ≥ 1), 43 using the ResFinder tool and 28 using the KmerResistance tool

385 (Figure 3B). Within a same class of antibiotics, the lowest diversity of ARGs was observed for

386 fosfomycin, quinolones and trimethoprim/streptothricin (n = 1), while the highest diversity of ARGs

387 was observed for MLS$_B$ (n = 15) and tetracycline (n = 13) (Table 3).

388 Out of 54 ARGs identified with the in-house method, 18 could not be detected using the web-

389 based tools. Of these, twelve were absent from the ResFinder database and thus could not have

390 been detected using the web-based tools (Figure 3B). Conversely, eight ARGs identified from the

391 SMS reads using the ResFinder and KmerResistance tools could not be detected using the in-

392 house method (Figure 3B).

393

394 *Incongruent results between methods at the individual level*

395 ARGs identified by all three methods at the gene level could sometimes be identified by only two

396 or one method at the individual level (see supplementary Table S2 for detailed results).

397 Gene cfxA (β-lactam resistance) was detected with all three methods in sample ST5-18mo but

398 was detected only with the in-house method in dog DFS (BD5M = 66.5). The best match in the

399　　NCBI's AMR reference database (NG_047636.1) was only 89% similar to the consensus

400　　sequence, while the best match by BLASTN against the full NCBI nucleotide database (98.9%

401　　similarity) was gene cfxA6 (GQ342996), which was absent from the NCBI's AMR database.

402　　Gene tet(32) (tetracycline resistance) was detected only with the in-house method in HS28 (BD5M

403　　= 38.0) with the best match in the NCBI's AMR reference database being NG_048124.1. However,

404　　reads coverage against NG_048124.1 was very high for the first 350 bp and last 700 bp only,

405　　probably related to recombination events with other 'tet' genes (see Discussion). The best match

406　　with BLASTN (97.8% similarity) was a recombinant tet(O/32/O) gene (JQ740052). Even though

407　　the ResFinder database includes five recombinant tet(O/32/O) gene references, it was not

408　　detected in HS28 with the web-based tools.

409

410　　**Figure 3. Resistance patterns identified from shotgun metagenomics sequencing (SMS)**

411　　**using three analysis methods (A) per gene and per sample or (B) per gene.** SMS reads were

412　　screened for the presence of antimicrobial resistance genes using (i) an "in-house method" (based

413　　on queries against the NCBI's AMR reference gene database BioProject PRJNA313047), (ii)

414　　ResFinder and/or (iii) KmerResistance tools.

415

416　　At the individual level, ARGs identified using the web-based tools but not the in-house method

417　　were: aadA24 and aph(3')-III (aminoglycosides); cfxA6 (β-lactams); catS (chloramphenicol);

418　　mdf(A) (MLS$_B$); tet(O/W), tet(O/32/O) and tet(W/32/O) (tetracycline). Out of these, seven genes

419　　were absent from the NCBI's AMR database, and thus could not have been detected using the in-

420　　house method (Table S2). The remaining gene (aadA24) was detected only with the web-based

421　　tools despite being present in both reference databases, but further analysis showed that this gene

422　　was detected as an aadA1 gene with the in-house method.

423

424　　*Resistance profiles between and within species*

425　　Different species (human / birds / dogs) and different samples showed different resistance profiles

426　　(Table 3). With the in-house method, HS23 showed the highest AMR diversity within the human

427　　samples, with 30 ARGs (or 19 at BD5M ≥ 1) belonging to six classes of antibiotics detected,

428　　followed by HS26 (28 ARGs or 15 at BD5M ≥ 1), and HS22 (28 ARGs or 12 at BD5M ≥ 1). HS24

429　　showed the highest abundances of individual ARGs, with 12 out of 20 ARGs detected at BD5M ≥

430　　10. When accounting for the additional genes detected using the ResFinder and KmerResistance

431　　tools, the number of ARGs increased for most samples, HS22 becoming the sample with the

432　　highest AMR diversity (34 ARGs) (Table S2).

433　　Most of the adult samples were similar to each other in their resistance profile, except for HS24

434　　(Figure 2). The three infants showed higher abundances for fewer ARGs compared to the adults,

435　　e.g. three ARGs at BD5M ≥ 1 in infant ST4-3mo: two ARGs conferring resistance to β-lactams

436　　(BD5M = 47 and 54) and one to tetracycline (BD5M = 280) (Table 3).

437

438 The sample MAD (pool from 6 Pacific black ducks) had the lowest AMR diversity, with two ARGs
439 at BD5M ≥ 1 (BD5M = 9 for aminoglycosides; 6.4 for MLS$_B$), while the sample MUD (from a
440 Muscovy duck) showed the presence of 16 ARGs at BD5M ≥ 1 (resistance to aminoglycoside,
441 MLS$_B$, β-lactams and tetracycline). ARGs in highest abundance in the sample MUD were tetA(P)
442 and tetB(P) (tetracycline resistance; BD5M > 40), two genes also found in humans and dog DFS
443 but at very low levels (Table 3).

444

445 The pooled DFS showed one of the highest AMR diversity, with 29 ARGs detected (or 20 at BD5M
446 ≥ 1) belonging to six classes of antibiotics; this number increased to 32 ARGs when accounting for
447 additional genes detected using the ResFinder and KmerResistance tools. The dog sample also
448 contained 12 ARGs at high abundance (BD5M ≥ 10), including 6 MLS$_B$ genes. This is comparable
449 to 5 MLS$_B$ genes at BD5M ≥ 10 found in HS24, but with only three ARGs in common, and those in
450 the dog DFS typically at higher abundances than in the human HS24.

451

452 **Identification of bacteria carrying ARGs**
453 The metagenomics analysis of the gut microbiome for selected samples (n = 7) using the Torrent
454 Suite Ion Reporter^{TM} software is summarized in Table 4. Because adult human microbiotas are
455 diverse and complex, we are detailing one example for which the ARGs carrying bacteria were
456 likely identified (HS24).

457

458 **Table 4. Percentages of major bacterial Orders in the fecal microbiome of selected samples**
459 **(n = 7).** Bacterial orders likely carrying antimicrobial resistance genes (ARGs) identified in each
460 sample are highlighted as a grey box.

461

Phylum	Order	ST5-1mo	ST5-18mo	ST4-3mo	HS24	DFS	ST1-MAD	ST3-MUD
Firmicutes	Clostridiales	< 1%	35%		41%	4%	9%	54%
	Erysipelotrichales						4%	18%
	Lactobacillales	< 1%				42%	74%	
	Selenomonadales					12%		
	Veillonellales			2%				2%
Bacteroidetes	Bacteroidales	34%	51%		56%	30%		
Proteobacteria	Burkholderiales		5%					
	Campylobacterales						2%	19%
	Enterobacteriales	66%		46%				
	Gammaproteobacteria					5%		
	Kopriimonadales						1%	
	Legionellales							2%
Verrucomicrobia	Verrucomicrobiales		6%		2%			
Actinobacteria	Bifidobacteriales		1%	51%				
Cyanobacteria	Synechococcales						5%	
	Nostocales						3%	
Deinococcus-Thermus	Deinococcales						1%	

462 ST5-1mo: infant sample one month old; ST5-18mo: infant sample 18 months old; ST4-3mo: infant sample three months
463 old; HS: human sample; DFS: dog fecal sample; MAD: pool from 6 Pacific black ducks; MUD: Muscovy duck.

464

465 *Human infants samples*

466 The composition of the humans gut microbiota was age-related (Table 4). At the taxonomic level

467 'Order', there were two overwhelmingly dominant bacterial populations in ST5-1mo:

468 Enterobacteriales (66%) and Bacteroidales (34%). ARGs were identified in both populations:

469 Order Enterobacteriales (possibly Species *Escherichia coli*) carrying the blaEC gene for

470 cephalosporin-hydrolyzing class C β-lactamase; and Order Bacteroidales (possibly Species

471 *Bacteroides fragilis*) carrying three ARGs, one to β-lactams, and two to MLS$_B$ antibiotics.

472 In ST5-18mo (same child as ST5-1mo but older) Order Enterobacteriales was absent, but new

473 Orders appeared: Bacteroidales (51%), Clostridiales (35%), Verrucomicrobiales (6%) and

474 Burkholderiales (5%). ARGs to three β-lactams and two MLS$_B$ antibiotics were most likely carried

475 by bacteria in the Order Bacteroidales, while ARGs to tetracyclines were most likely carried by

476 bacteria in various taxa, predominantly Order Clostridiales and Order Bacteroidales.

477 In ST4-3mo, we detected no or little bacteria in Order Bacteroidales, but many bacteria in

478 Bifidobacteriales (51%) and Enterobacteriales (46%). ARGs were identified in bacteria belonging

479 to Order Enterobacteriales only: two ARGs to β-lactams, one to fosfomycin, one to MLS$_B$

480 antibiotics and one to tetracycline. Of note, genes blaTEM and tet(A) were likely carried by a

481 plasmid, as identified by a BLASTn query of the contigs carrying these genes.

482

483 *Human adult samples*

484 In sample HS24, the two main Orders were Bacteroidales (56%) and Clostridiales (41%). The

485 most abundant ARGs appeared to be carried by bacteria in the Order Bacteroidales (resistance to

486 aminoglycosides, MLS$_B$ and tetracyclines). Of note, gene erm(B) (conferring resistance to

487 macrolides) was likely carried by a plasmid.

488

489 *Animal samples*

490 The gut microbiota and resistant bacterial species in the dog sample DFS were somewhat similar

491 to the human samples, with bacteria in the Order Bacteroidales likely carrying one β-lactam-

492 resistance gene, while 6 MLS$_B$-resistance genes were likely carried by bacteria in the Phylum

493 Firmicutes (Order unidentified). Bacteria carrying ARGs to tetracycline and aminoglycosides could

494 not be definitively identified. The aph(3')-Ia gene (conferring resistance to aminoglycosides) was

495 likely sitting on a plasmid also carrying a transposase, and the aadS gene (aminoglycosides) was

496 likely sitting on a transposon.

497 The birds' microbiotas were different from the mammals (Table 3). Bacteria from the Order

498 Lactobacillales (possibly Species *Enterococcus faecalis*) were likely carrying gene lsa(A)

499 (conferring multidrug resistance) detected in the wild pacific black ducks (MAD). Bacteria from the

500 Order Clostridia (possibly Species *Clostridium perfringens*) was likely carrying genes erm(Q)

501 (resistance to MLS$_B$) and tetA(P) and tetB(P) (resistance to tetracycline) detected in the Muscovy

502 duck (MUD), with other less abundant MLS$_B$-resistance genes likely carried by unidentified

503 bacteria (Phylum Firmicutes).

504

505 **Comparison of the SMS method and the targeted sequencing methods to detect ARGs**

506   Four human samples were analysed using all three methods. For the SMS method, we focused on
507   the in-house analysis and ARGs at BD5M ≥ 1. The results are summarized in Table 5 (see Table
508   S3 for detailed information per sample and per gene).
509   Combining all methods, we detected ARGs conferring resistance to 13 different classes of
510   antibiotics, with two additional classes not detected with the SMS-based in-house analysis:
511   Mupirocin and Quaternary Ammonium Compounds (QAC). The number of ARGs identified from
512   the four individuals was 29 to 54 using the Ion AmpliSeq^{TM} AMR panel, 14 to 30 using the Ion
513   AmpliSeq^{TM} Pan-Bacterial panel, and 9 to 28 using the SMS-based in-house analysis (Table 5).
514   Two to seven ARGs were found with all three methods, while four to 16 ARGs were found with
515   both the SMS-based in-house analysis and the AMR panel (Table S3). The AMR panel was the
516   most sensitive, identifying 21 to 38 ARGs not found with any other method, representing 50-66.7%
517   of the overall identified genes. Only two to three ARGs were identified with the Pan-Bacterial panel
518   alone (4.4-7.1% of all genes) while 5 to 11 genes were identified with the SMS-based in-house
519   analysis alone (11.9-16.2% of all the genes).
520
521   **Table 5. Number of antimicrobial resistance genes (ARGs) identified with three sequencing**
522   **methods from human samples (n = 4).** Reads obtained from the shotgun metagenomics
523   sequencing (SMS) were queried against the NCBI's AMR reference gene database BioProject
524   PRJNA313047. Other methods were targeted sequencing using the Ion AmpliSeq^{TM} AMR
525   Research panel or the Ion AmpliSeq^{TM} Pan-Bacterial panel.
526

	Method	ST5-1mo	HS21	HS24	HS26
Number of ARGs identified with each method	AMR Research panel *	29 (14)	39 (33)	34 (28)	54 (44)
	Pan-Bacterial panel	14	26	23	30
	SMS + in-house **	9 (7)	17 (10)	20 (16)	28 (15)
Number of ARGs identified with any method		34	45	39	64
Number of ARGs identified exclusively with one method	AMR Research panel only	24 (66.7%)	27 (55.1%)	21 (50.0%)	38 (55.9%)
	Pan-Bacterial panel only	2 (5.6%)	3 (6.1%)	3 (7.1%)	3 (4.4%)
	SMS + in-house only	5 (13.9%)	6 (12.2%)	5 (11.9%)	11 (16.2%)

527
528   * The number of genes with a depth >1 is given in brackets. ** The number of genes with a mean base depth per 5
       million reads (BD5M) >1 is given in brackets.
529
530   General performances of two methods (the SMS-based in-house analysis and the AMR panel)
531   were compared in Figure 4.
532
533   **Figure 4. Performance summary** of the in-house analysis from the shotgun metagenomics
534   sequencing (SMS), compared to one targeted sequencing method using the Ion AmpliSeq^{TM} AMR
535   panel (Life Technologies). Results obtained with the AMR panel were analyzed with the One
536   Codex online platform (available at: https://app.onecodex.com/); depth is defined as the average
537   number of reads piled up across the entire gene.
538
539   **Discussion**

540

541   Although only including a relatively low number of samples (n = 13), we identified a high number of

542   ARGs (34 to 64) predicted to confer resistance to 13 classes of antibiotics. All sequencing

543   methods were able to detect the most abundant genes, showing that ARGs were common in the

544   gut microbiota of humans and dogs in Australia. Australian birds also carried ARGs in their gut, at

545   a very low level in six wild Pacific black ducks from a non-urban environment, and at higher

546   abundances in a Muscovy duck from an environment close to humans (semi-urban).

547

548   **ARGs conferring resistance to tetracycline, β-lactams and MLS$_B$ are widespread**

549   The most abundant ARGs present in the samples conferred resistance to tetracycline, β-lactams

550   and MLS$_B$ antibiotics. Interestingly, antibiotics within these classes are used frequently in human

551   and veterinary medicine in Australia. Between 2015 and 2017, in human medicine, the most

552   commonly dispensed antibiotics in community-based outpatient practice were β-lactams (57.3%),

553   tetracycline (8.0%) and macrolides (5.1%), similar to those prescribed in hospitals (63.3% β-

554   lactams; 8.2% tetracycline; 4.0% macrolides)[40]. Antimicrobial use in veterinary medicine in

555   Australia is not as well documented, however the same three classes of antibiotics made up the

556   majority (50.5 to 65.4%) of antimicrobials sold for animal use between 2005 and 2010 but ranked

557   differently: macrolides and streptogramins (18.8 to 27.1%), followed by tetracyclines (16.3 to

558   28.6%) and β-lactams (9.2 to 14.6%)[41]. Studies have reported that subjects from countries with

559   tighter policies on antibiotic usage in humans and animals have considerably less ARG levels[14]. In

560   line with this, we would argue that the common usage of specific classes of antibiotics as noted

561   above is likely driving the high abundance of matching ARGs in the gut microbiota of people and

562   animals in Australia.

563

564   A large proportion of ARGs in our human samples were likely carried by bacteria within the Order

565   Bacteroidales that is common in the human lower gut[42]. ARGs conferring resistance to tetracycline

566   are one of the most commonly found in human gut microbiota[15] and their prevalence has steadily

567   increased in Bacteroidales since the 1970's[43]. A study suggested that once acquired, ARGs to

568   tetracycline can persist in Bacteroidales in the absence of antibiotic selection[43]. This could explain

569   why ARGs to tetracycline were the most abundant in our data (detected in 12 samples out of 13,

570   with up to 11 tetracycline ARGs in a single sample), despite tetracycline only being the third most

571   common class of antibiotics prescribed in the Australian community behind several β-lactams[40].

572

573   Mosaic tetracycline resistance genes, presumably arising by recombination between wild-type

574   genes, have been discovered recently, and studies have shown that mosaic genes comprising

575   tet(O), tet(W), and tet(32) sequences were abundant in DNA extracted from pig and human fecal

576   samples[44]. In our data, mosaic genes tet (O/W), tet(O/32/O) and tet(W/32/O) were identified using

577   the ResFinder tool. There is probably more undetected, as reads mapping against tet genes

578   sometimes showed incomplete coverage with clear truncations. Targeted sequencing techniques

579   might be required in order to properly assess tetracycline resistance that might otherwise be

580   underestimated.

581

Detected ARGs in humans and animals are diverse

583   The diversity of ARGs detected in humans was generally higher in adults compared to infants.

584   This might be correlated with higher diversity of the gut microbiota in adults (the range of

585   microorganisms found in the intestine changes dramatically between birth and the age of 3

586   years[45]) and/or a growing antibiotic selection pressure with age. Although the ARGs in infants

587   were less diverse, they were at very high relative abundances, having the 2nd, 3rd and 4th highest

588   mean coverage of ARGs (BD5M) of all the human samples. In Australia, antibiotic usage in

589   children aged 0 to 9 years is almost equivalent to one antibiotic prescription per child and per

590   year[46], which is relatively high when compared to similar countries[47]. Although, the youngest infant

591   still had ARGs evident when it had never been exposed to antibiotics.

592

593   The ARGs detected in the dog sample were somewhat similar to the ones detected in adult

594   humans. This might reflect similarities in the diversity and composition of the dog and human fecal

595   microbiota, with microbial exchange facilitated by direct contact[45] as well as a similarity in the

596   antibiotics used (as previously mentioned). Very few ARGs were detected in the Pacific black duck

597   MAD (a pool of faeces from wild ducks collected in rural areas) compared to the Muscovy duck

598   MUD (common duck collected in a semi-urban area). This is consistent with previous studies

599   demonstrating that proximity to human activity increases the number of the antibiotic-resistant

600   bacteria that are associated with wild birds. For example, gulls and geese nesting near waste or

601   agricultural water harbor more antibiotic-resistant *E. coli* than do birds associated with unpolluted

602   water[48,49]. Interestingly, ARGs to macrolide and tetracycline were the most abundant in the

603   Muscovy duck MUD sample, overlapping with the antimicrobials used for therapeutic purposes in

604   farm animals in Australia[41]. Most of the macrolides and tetracyclines are administered in food or

605   water to pigs and poultry in Australian farms [41], and environmental contamination near farms with

606   both antibiotics and ARGs has been demonstrated, as well as the higher abundance of ARGs in

607   wild animals residing in the vicinity of farms[50].

608

The different sequencing methods have advantages and limitations

610   We tested three sequencing techniques and analysed the SMS-based data with three methods.

611   The SMS-based in-house method and the AMR panel detected the greatest range of ARGs, with

612   the highest diversity of ARGs obtained with a combination of the two methods. However, each

613   method had advantages and disadvantages. The SMS-based in-house analysis and the AMR

614   panel both detected abundant ARGs, while discordant results occurred for ARGs at low

615   abundances, because of different detection sensitivities of the methods.

616   Discordant results between methods also occurred from the use of different AMR reference

617   databases. Of the ARGs identified with the SMS-based in-house analysis only, many were absent

618   from the AMR panel, and therefore could not have been detected by this method. Similarly, many

619 of the ARGs identified with the AMR panel only were absent from the NCBI's AMR database,

620 therefore could not have been detected using the in-house analysis (Figure 4).

621

622 ***The two targeted sequencing methods relying on AmpliSeq^{TM} panels*** were very sensitive

623 because of the pre-amplification step by PCR. The AMR panel identified the highest number of

624 ARGs and captured >50% more genes not identified by the other methods. This included genes

625 present in the NCBI's AMR database but nonetheless undetected with the in-house method, e.g.

626 erm(G), sat4, tet(M), lnu(C). Despite amplification, those genes were found with the AMR panel at

627 a depth below 6,000 i.e. ≤ 1% of the total amplified reads per sample (see Table S3 for details). As

628 a comparison, in sample HS24, all ARGs represented 0.2% of total SMS reads. With a non-

629 targeted method, sensitivity can only be increased by increasing the number of reads.

630

631 On the contrary and less intuitively, four genes were detected at relatively high abundances with

632 the SMS-based in-house method but at a low depth with the AMR panel: mef(A), aadE, aadS, tetQ

633 (Figure 4). This might indicate poor efficiency of some primers. In sample HS24, the "One Codex"

634 online platform reported a mean depth of 328 for tet(M) gene when in reality two sets of primers

635 were amplified and detected at different depths (38 and 608). Other sets of primers may be

636 working poorly, leading to false negative results or underestimated depths. Also, genes used to

637 design the primers might also have been being quite different from the one present in our samples.

638 This seemed to be the case for gene aadE detected in sample HS24, with the best match using

639 the SMS-based in-house method (NG_047378.1 *Pediococcus acidilactici* aadE gene for

640 aminoglycoside 6-adenylyltransferase) being only 69% similar to the aadE gene reference

641 AF516335 used to design the AMR panel.

642

643 Targeted methods are limited by their inability to identify novel AMR determinants, whereas a

644 functional SMS approach might provide more detail[52]. Using an in-solution probe-and-capture

645 strategy[53], researchers suggested that designed probes should target sequences with up to 15%

646 nucleotide sequence divergence from a reference sequence, which would widen their applicability

647 and target capacity toward newly characterized members of AMR gene families, which often differ

648 from other members by only a few nucleotides[28].

649

650 Because the depth analysis follows an amplification step, targeted methods are not fully

651 quantitative compared to SMS. Nevertheless, it can be considered semi-quantitative as the depth

652 generally gives a good idea of the initial quantity of the ARGs present.

653

654 ***The SMS-based in-house analysis*** was not as sensitive as the targeted methods, being that the

655 sequencing is random with no amplification step. However, (i) it is quantitative and (ii) it potentially

656 allows identification of the bacteria carrying ARGs. Limitations of the SMS-based in-house analysis

657 include reliance on manual checking of the ARGs detected. In the NCBI database used for the

658 current study, references were constructed to include the ARG but also flanking regions of

659   approximately 100 bp each side (if such regions were available in the original sequence report).

660   This could trigger identification of false-positives for the presence of ARGs. That was the case for

661   nimJ gene in HS24 that was originally identified at a BD5M = 0.3, but after manual checking, all

662   reads mapped against the first 100 bp only of reference NG_048017 (*Bacteroides fragilis* nimJ

663   gene for nitroimidazole resistance), i.e. the bacterial host genome.

664

665   Manual checking was also essential to avoid overestimation of a gene's abundance. Gene tet(C)

666   was identified in ST4-3mo at a BD5M = 13, but all reads mapped against a limited region of the

667   gene and the consensus sequence was only 80.8% similar to the best reference in the NCBI's

668   AMR database NG_048174 (*Francisella tularensis* tet(C) gene for tetracycline efflux MFS

669   transporter). The best match overall using BLASTN was gene tet(A) (98.9% similarity; reference

670   MG904997). A tet(A) gene (reference NG_048158) was actually identified in ST4-3mo at a BD5M

671   = 279.8, confirming that the reads mapped to NG_048174 / tet(C) were in fact not the best match.

672

673   Library post-amplification (done when the initial library quantity was too low) triggered unexpected

674   problems and should be avoided when possible. Reference NG_047324 (*Escherichia coli* aadA1

675   gene for ant(3")-Ia family aminoglycoside nucleotidyltransferase) was identified in MAD at a BD5M

676   = 9 but the mapping result showed that all reads matched the same 200 bp region in the middle of

677   the gene, so that the aadA1 gene coverage was only 22.7%. This specific sample (as well as ST4-

678   3mo, MUD and DFS) was post-amplified, meaning that there might have only been a small amount

679   of aadA1 gene in the initial DNA extract, and one fragment only was post-amplified and

680   sequenced, leading to overestimating the depth of this gene in MAD.

681

682   In contrast, the depth may have been underestimated for some genes. In HS24, reference

683   NG_047625 (*Bacteroides fragilis* cepA gene for beta-lactamase) was identified at a BD5M = 0.9

684   using a Q-value = 20, with NG_047625 being only 80.9% similar to the consensus sequence. This

685   could be verified at a Q-value = 0, with good mapped reads, only quite different from the reference

686   used for mapping; BD5M increased to 4.5, which is probably more accurate. One way to overcome

687   this issue and obtain the real depth is to use the consensus sequence as the reference for re-

688   mapping the reads. Identification from reads or contigs relies heavily on the reference databases,

689   and this example shows that the closest match with a non-perfect reference can still allow the

690   detection of a hypothetical ARG. But because only distantly related to the original curated ARG,

691   further explorations are required to assess if the gene is actually a functional resistance gene.

692

693   Genes for which there is no good match in the reference database or in the NCBI general

694   nucleotide database might be totally missed. When conducting read mapping coverage analyses

695   at a Q-value = 0 instead of 20, we detected in some samples mixed populations of reads, i.e. one

696   population matching the reference quite well, and the second population being more dissimilar. A

697   coverage analysis at a Q-value = 20, showed the real depth of population 1, but discarded

698   population 2 for which no better reference was available.

699

700   **The ResFinder tool,** especially when combined with the **KmerResistance tool**, had a sensitivity

701   similar to the SMS-based in-house analysis, albeit with the ARGs detected being slightly different.

702   Even though the ResFinder and KmerResistance tools relied on the same database, results did

703   not fully overlap, probably because they used different input data, i.e. contigs versus reads

704   respectively. Compared to the in-house analysis, the online tools are more user-friendly and

705   provide quick and relatively accurate results. Provided that the reference database is regularly

706   updated to include missing genes (in our case the obvious ones being aminoglycoside/aadS,

707   lincosamide/lnu(AN2), macrolide/mef(En2) and β-lactam/blaEC) these are valuable tools for

708   scientists and diagnosticians requiring fast results. However, they don't allow manually rechecking

709   the mapping quality, and do not provide information on whether the genes are chromosomal or on

710   plasmids, nor the taxa of the host bacteria. Even though the ResFinder tool provides an accession

711   number corresponding to a good match for the host bacteria (from the contig sequence), it does

712   provide a single match even though several other (potentially very different) species might show

713   the same percentage of identity with the contig. Also, because we used the ResFinder tool with the

714   contigs only, abundance is not provided, while the KmerResistance tool provides a depth result for

715   the ARGs identified. The SMS-based in-house analysis provided the best accuracy regarding the

716   abundance of ARGs, our calculated "BD5M" coverage taking into account the size of the gene as

717   well as the depth of coverage.

718

719   **All the methods tested** shared some common limitations. First, they rely on reference databases

720   that are non-exhaustive, precluding the detection of ARGs very different to those included in the

721   database, and impacting the abundance estimations of distantly related genes or when multiple

722   similar genes are present. ARGs that show many different alleles were challenging to detect (e.g.

723   cfxA / cfxA6 in sample DFS or aadA1 / aadA24 in sample MUD). For targeted methods, new

724   (larger) panels are constantly being released, e.g. 37,826 probes targeting over 2,000 nucleotide

725   sequences associated with AMR using a probe-and-capture strategy[28], or 78,600 non-redundant

726   genes (including 47,806 putative ARGs) using targeted metagenomics[51].

727   Second, genotypic resistance does not necessarily reflect a phenotypic resistance. More studies

728   comparing molecular data with phenotypic resistance are needed to assess how much of the

729   detected resistome is linked to a functional resistant gene. It will also be important in the future to

730   assess if those genes persist over time in an individual, and how much ARGs in the gut microbiota

731   can be horizontally transferred to other bacteria species, potentially pathogenic ones.

732

733   This study is the first to describe the bacterial resistome in humans and animals from Australia,

734   comparing and assessing various methods. A high diversity of ARGs was identified, as well as a

735   high variability between samples, even with a limited number of samples analysed. We show that

736   resistomes from (human) adults and infants are dissimilar, and that resistomes from adults, dogs

737   and urban birds are more similar than that from rural birds. Ongoing surveillance of ARGs

738   abundance in people and animals (including livestock and wild animals) in Australia using the

739   techniques described here could help inform policy makers and health care professionals on the

740   most prudent use of these important drugs to ensure we have access to effective treatments of

741   bacterial infections well into the future.

742

743   **Ethics statement**

744   Ethics approval for the use of seven pregnant mothers samples was granted by the Barwon Health

745   Human Research and Ethics Committee (HREC 10/24). The study involving the three infant

746   samples was deemed negligible or low risk by the Barwon Health Human Research Ethics

747   Committee and therefore exempt from full committee review (HREC approval 17/119). Bird sample

748   collection was approved under Deakin University's Animal Ethics Committee project number B43–

749   2016 and Department of Environment, Land, Water and Planning permit number 1008206. Fecal

750   samples from the dogs were environmental samples collected on the ground and did not require

751   ethics approval.

752

753   **Data availability**

754   All data generated or analysed during this study are included in the published article.

755

756   **References**

757

758   1   Zumla, A. *et al.* Emerging novel and antimicrobial-resistant respiratory tract infections: new
759   drug development and therapeutic options. *Lancet Infect Dis* **14**, 1136-1149,
760   doi:10.1016/S1473-3099(14)70828-X (2014).
761   2   Organization, W. H. Draft thirteenth general programme of work, 2019–2023. Promote
762   health, keep the world safe, serve the vulnerable. (WHO, Geneva, Switzerland, 2018).
763   3   D'Costa, V. M., McGrann, K. M., Hughes, D. W. & Wright, G. D. Sampling the antibiotic
764   resistome. *Science* **311**, 374-377, doi:10.1126/science.1120800 (2006).
765   4   Aarts, H. & Margolles, A. Antibiotic resistance genes in food and gut (non-pathogenic)
766   bacteria. Bad genes in good bugs. *Front Microbiol* **5**, 754, doi:10.3389/fmicb.2014.00754
767   (2014).
768   5   Claesson, M. J. *et al.* Comparative analysis of pyrosequencing and a phylogenetic
769   microarray for exploring microbial community structures in the human distal intestine.
770   *PLoS One* **4**, e6669, doi:10.1371/journal.pone.0006669 (2009).
771   6   Lozupone, C. A., Stombaugh, J. I., Gordon, J. I., Jansson, J. K. & Knight, R. Diversity,
772   stability and resilience of the human gut microbiota. *Nature* **489**, 220-230,
773   doi:10.1038/nature11550 (2012).
774   7   Vignaroli, C., Zandri, G., Aquilanti, L., Pasquaroli, S. & Biavasco, F. Multidrug-resistant
775   enterococci in animal meat and faeces and co-transfer of resistance from an Enterococcus
776   durans to a human Enterococcus faecium. *Curr Microbiol* **62**, 1438-1447 (2011).
777   8   Boerlin, P. & Reid-Smith, R. J. Antimicrobial resistance: its emergence and transmission.
778   *Anim Health Res Rev* **9**, 115-126, doi:10.1017/S146625230800159X (2008).
779   9   Salyers, A. A., Gupta, A. & Wang, Y. P. Human intestinal bacteria as reservoirs for
780   antibiotic resistance genes. *Trends Microbiol* **12**, 412-416, doi:10.1016/j.tim.2004.07.004
781   (2004).
782   10   Marshall, B. M., Ochieng, D. J. & Levy, S. B. Commensals: underappreciated reservoir of
783   antibiotic resistance. *Microbe* **4** (2009).
784   11   Francino, M. P. Antibiotics and the Human Gut Microbiome: Dysbioses and Accumulation
785   of Resistances. *Front Microbiol* **6**, 1543, doi:10.3389/fmicb.2015.01543 (2015).
786   12   Jernberg, C., Lofmark, S., Edlund, C. & Jansson, J. K. Long-term impacts of antibiotic
787   exposure on the human intestinal microbiota. *Microbiology* **156**, 3216-3223,
788   doi:10.1099/mic.0.040618-0 (2010).

789   13   Perez-Cobas, A. E. *et al.* Differential effects of antibiotic therapy on the structure and
790   function of human gut microbiota. *PLoS One* **8**, e80201,
791   doi:10.1371/journal.pone.0080201 (2013).
792   14   Forslund, K. *et al.* Country-specific antibiotic use practices impact the human gut
793   resistome. *Genome Res* **23**, 1163-1169, doi:10.1101/gr.155465.113 (2013).
794   15   Hu, Y. F. *et al.* Metagenome-wide analysis of antibiotic resistance genes in a large cohort
795   of human gut microbiota. *Nat Commun* **4**, doi:10.1038/ncomms3151 (2013).
796   16   Donskey, C. J. Antibiotic regimens and intestinal colonization with antibiotic-resistant
797   gram-negative bacilli. *Clin Infect Dis* **43**, S62-S69, doi:Doi 10.1086/504481 (2006).
798   17   Gou, M. *et al.* Aerobic composting reduces antibiotic resistance genes in cattle manure
799   and the resistome dissemination in agricultural soils. *Sci Total Environ* **612**, 1300-1310,
800   doi:10.1016/j.scitotenv.2017.09.028 (2018).
801   18   Hu, H. W. *et al.* Temporal changes of antibiotic-resistance genes and bacterial
802   communities in two contrasting soils treated with cattle manure. *FEMS Microbiol Ecol* **92**,
803   doi:10.1093/femsec/fiv169 (2016).
804   19   Zhang, Y. J. *et al.* Temporal succession of soil antibiotic resistance genes following
805   application of swine, cattle and poultry manures spiked with or without antibiotics. *Environ*
806   *Pollut* **231**, 1621-1632, doi:10.1016/j.envpol.2017.09.074 (2017).
807   20   Marcelino, V. R. *et al.* Meta-transcriptomics reveals a diverse antibiotic resistance gene
808   pool in avian microbiomes. *BMC Biol* **17**, 31, doi:10.1186/s12915-019-0649-1 (2019).
809   21   Bertelli, C. & Greub, G. Rapid bacterial genome sequencing: methods and applications in
810   clinical microbiology. *Clin Microbiol Infect* **19**, 803-813, doi:10.1111/1469-0691.12217
811   (2013).
812   22   Didelot, X., Bowden, R., Wilson, D. J., Peto, T. E. A. & Crook, D. W. Transforming clinical
813   microbiology with bacterial genome sequencing. *Nat Rev Genet* **13**, 601-612,
814   doi:10.1038/nrg3226 (2012).
815   23   McArthur, A. G. & Tsang, K. K. Antimicrobial resistance surveillance in the genomic age.
816   *Ann N Y Acad Sci* **1388**, 78-91, doi:10.1111/nyas.13289 (2017).
817   24   McArthur, A. G. *et al.* The comprehensive antibiotic resistance database. *Antimicrob*
818   *Agents Chemother* **57**, 3348-3357 (2013).
819   25   Gupta, S. K. *et al.* ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance
820   genes in bacterial genomes. *Antimicrob Agents Chemother* **58**, 212-220,
821   doi:10.1128/AAC.01310-13 (2014).
822   26   NCBI. Bacterial antimicrobial resistance reference gene database. (2016). Accessed on
823   the 02 August 2018. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA313047
824   27   Zankari, E. *et al.* Identification of acquired antimicrobial resistance genes. *J Antimicrob*
825   *Chemother* **67**, 2640-2644, doi:10.1093/jac/dks261 (2012).
826   28   Guitor, A. K. *et al.* Capturing the Resistome: a Targeted Capture Method To Reveal
827   Antibiotic Resistance Determinants in Metagenomes. *Antimicrob Agents Chemother* **64**,
828   doi:10.1128/AAC.01324-19 (2019).
829   29   Vuillermin, P. *et al.* Cohort Profile: The Barwon Infant Study. *Int J Epidemiol* **44**, 1148-
830   1160, doi:10.1093/ije/dyv026 (2015).
831   30   Vibin, J. *et al.* Metagenomics detection and characterisation of viruses in faecal samples
832   from Australian wild birds. *Sci Rep* **8**, 8686, doi:10.1038/s41598-018-26851-1 (2018).
833   31   Bhatta, T. R., Chamings, A., Vibin, J. & Alexandersen, S. Detection and characterisation of
834   canine astrovirus, canine parvovirus and canine papillomavirus in puppies using next
835   generation sequencing. *Sci Rep* **9**, 4602, doi:10.1038/s41598-019-41045-z (2019).
836   32   Urbaniak, C. *et al.* Detection of antimicrobial resistance genes associated with the
837   International Space Station environmental surfaces. *Sci Rep* **8**, 814, doi:10.1038/s41598-
838   017-18506-4 (2018).
839   33   Boratyn, G. M., Thierry-Mieg, J., Thierry-Mieg, D., Busby, B. & Madden, T. L. Magic-
840   BLAST, an accurate RNA-seq aligner for long and short reads. *BMC Bioinformatics* **20**,
841   405, doi:10.1186/s12859-019-2996-x (2019).
842   34   Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to
843   single-cell sequencing. *J Comput Biol* **19**, 455-477, doi:10.1089/cmb.2012.0021 (2012).
844   35   Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment
845   search tool. *J Mol Biol* **215**, 403-410, doi:10.1016/S0022-2836(05)80360-2 (1990).
846   36   Mount, D. W. Using the Basic Local Alignment Search Tool (BLAST). *CSH Protoc* **2007**,
847   pdb top17, doi:10.1101/pdb.top17 (2007).

848 37 Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV):
849   high-performance genomics data visualization and exploration. *Brief Bioinform* **14**, 178-
850   192, doi:10.1093/bib/bbs017 (2013).
851 38 Clausen, P., Aarestrup, F. M. & Lund, O. Rapid and precise alignment of raw reads
852   against redundant databases with KMA. *BMC Bioinformatics* **19**, 307, doi:10.1186/s12859-
853   018-2336-6 (2018).
854 39 Clausen, P. T., Zankari, E., Aarestrup, F. M. & Lund, O. Benchmarking of methods for
855   identification of antimicrobial resistance genes in bacterial whole genome data. *J
856   Antimicrob Chemother* **71**, 2484-2488, doi:10.1093/jac/dkw184 (2016).
857 40 ACSQHC. AURA 2019 - Third Australian report on antimicrobial use and resistance in
858   human health. 261 (Australian Commission on Safety and Quality in Health Care, Sydney,
859   Australia, 2019).
860 41 APVMA. Quantity of antimicrobial products sold for veterinary use in Australia - July 2005
861   to June 2010. 73 (Australian Pesticides and Veterinary Medicines Authority, Canberra,
862   Australia, 2014).
863 42 Salyers, A. A. Bacteroides of the human lower intestinal tract. *Annu Rev Microbiol* **38**, 293-
864   313, doi:10.1146/annurev.mi.38.100184.001453 (1984).
865 43 Shoemaker, N. B., Vlamakis, H., Hayes, K. & Salyers, A. A. Evidence for extensive
866   resistance gene transfer among Bacteroides spp. and among Bacteroides and other
867   genera in the human colon. *Appl Environ Microbiol* **67**, 561-568 (2001).
868 44 Patterson, A. J., Rincon, M. T., Flint, H. J. & Scott, K. P. Mosaic tetracycline resistance
869   genes are widespread in human and animal fecal samples. *Antimicrob Agents Chemother*
870   **51**, 1115-1118, doi:10.1128/AAC.00725-06 (2007).
871 45 Song, S. J. *et al.* Cohabiting family members share microbiota with one another and with
872   their dogs. *Elife* **2**, e00458, doi:10.7554/eLife.00458 (2013).
873 46 Australian Commission on Safety and Quality in Health Care *The third Australian Atlas of
874   Healthcare variation 2018* (ACSQHC, Sydney, Australia 2018).
875 47 Organisation for Economic Co-operation and Development. Health at a glance 2017:
876   OECD indicators. (OECD, Paris, 2017).
877 48 Cole, D. *et al.* Free-living Canada geese and antimicrobial resistance. *Emerg Infect Dis*
878   **11**, 935-938, doi:10.3201/eid1106.040717 (2005).
879 49 Dolejska, M., Cizek, A. & Literak, I. High prevalence of antimicrobial-resistant genes and
880   integrons in Escherichia coli isolates from Black-headed Gulls in the Czech Republic. *J
881   Appl Microbiol* **103**, 11-19, doi:10.1111/j.1365-2672.2006.03241.x (2007).
882 50 Kraemer, S. A., Ramachandran, A. & Perron, G. G. Antibiotic Pollution in the Environment:
883   From Microbial Ecology to Public Policy. *Microorganisms* **7**,
884   doi:10.3390/microorganisms7060180 (2019).
885 51 Lanza, V. F. *et al.* In-depth resistome analysis by targeted metagenomics. *Microbiome* **6**,
886   11, doi:10.1186/s40168-017-0387-y (2018).
887 52 Crofts, T. S., Gasparrini, A. J. & Dantas, G. Next-generation approaches to understand
888   and combat the antibiotic resistome. *Nat Rev Microbiol* **15**, 422-434,
889   doi:10.1038/nrmicro.2017.28 (2017).
890 53 Gnirke, A. *et al.* Solution hybrid selection with ultra-long oligonucleotides for massively
891   parallel targeted sequencing. *Nat Biotechnol* **27**, 182-189, doi:10.1038/nbt.1523 (2009).
892

Acknowledgements

903

Author Contributions

905 Project design and coordination was performed by SA. Samples were collected by and selected by
906 SA and FC, with adult human samples collection coordinated by the BIS investigator group.
907 Nucleic acid was extracted by VG and FC. VG and FC performed the sample library and NGS
908 DNA sequencing. VG, AC and SA performed the reads analyses. VG wrote the initial manuscript
909 with input from SA, and later versions were based on input and suggestions from all.
910

Additional information

912 **Competing Interests:** The authors declare no competing interests.
913

Legends to Supplementary Tables

915

916 **S1 Table: Detailed results from the mapping of Shotgun Metagenomic Sequencing (SMS)**
917 **against the NCBI's AMR reference gene database using the Torrent Suite TMAP plugin.**
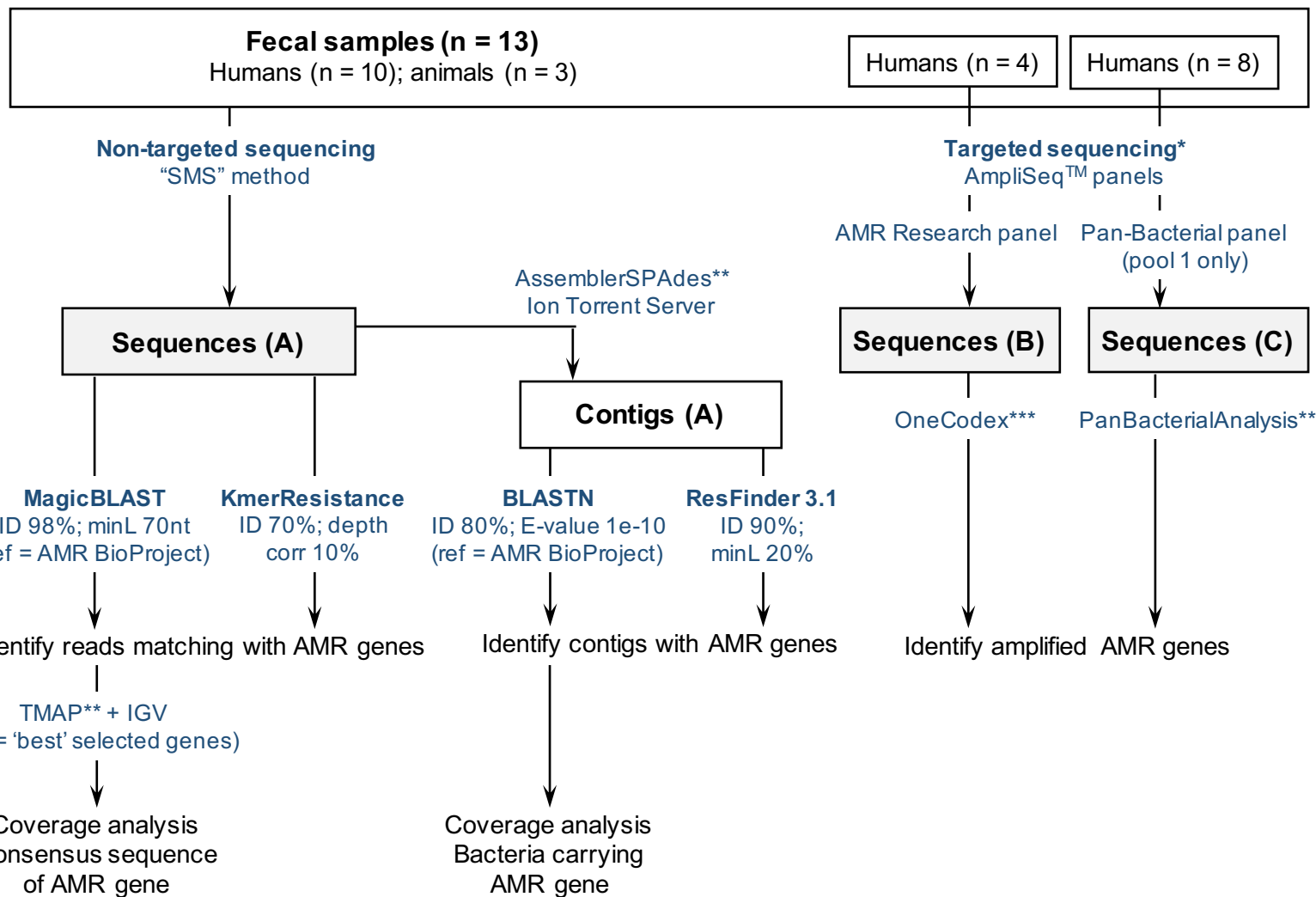918 Reads were queried against the database using Magic-BLAST and contigs were queried against
919 the database using BLASTN. For each reference (line), red squares indicate the samples for which
920 the reference was selected from the reads and/or contigs as the best (or one of the best), while the
921 other samples were mapped against the reference just as a comparison with the other samples.
922 Mean base depth per 5 million reads (BD5M) is highlighted in yellow when between 1 and 10, or
923 red when > 10. When a reference is the best mapping result but the number of reads mapped to
924 the reference is low (< 20), the exact number of reads is provided in brackets.
925

926 **S2 Table: Detailed comparison of the analyses with the SMS-based method using (i) the**
927 **KmerResistance tool, (ii) the ResFinder tool and (iii) our in-house method.** For ResFinder
928 and KmerResistance the percentages provided are the template coverage. For the in-house
929 method, the calculated "mean base depth per 5 million reads" is provided. * gene not in the NCBI's
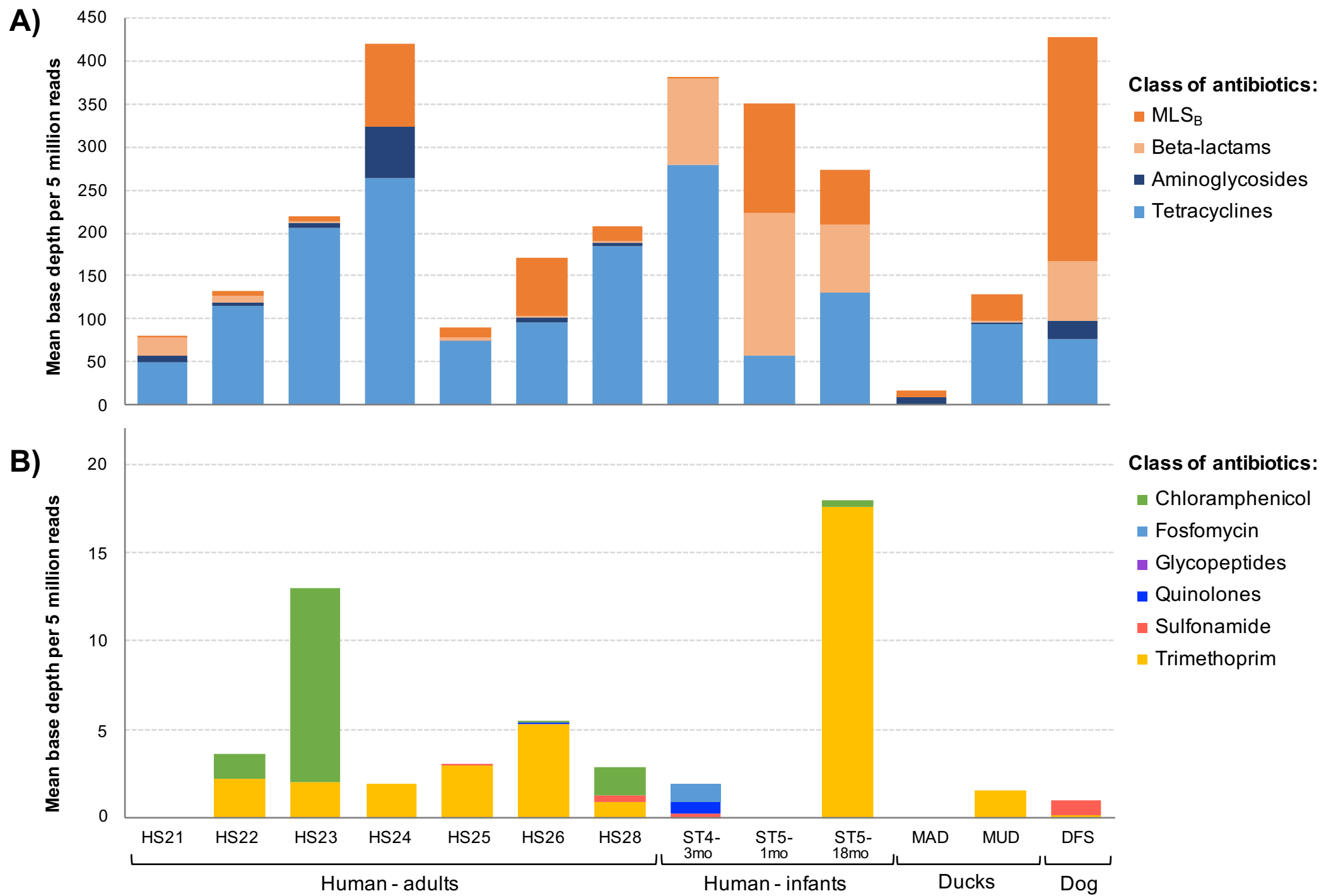930 AMR gene database; § gene not in the ResFinder database.
931

932 **S3 Table: Results from a SMS-based in-house method and a targeted metagenomic**
933 **sequencing method using the Ion AmpliSeq^{TM} AMR panel (Life Technologies).** Results
934 obtained with the AMR panel were analysed with the One Codex online platform (available at:
935 https://app.onecodex.com/); depth is defined as the average number of reads piled up across the
936 entire gene. SMS: shotgun metagenomics sequencing; BD5M: mean base depth per 5 million
937 reads; MLS: macrolide, lincosamide, streptogramin B; QAC: quaternary ammonium compounds;
938 gni: gene not included in Bioproject database; gnp: gene not included in the AMR panel. Genes
939 detected with both methods are highlighted in green; genes in blue are detected with the AMR
940 panel only; genes in yellow are detected with the SMS method only.
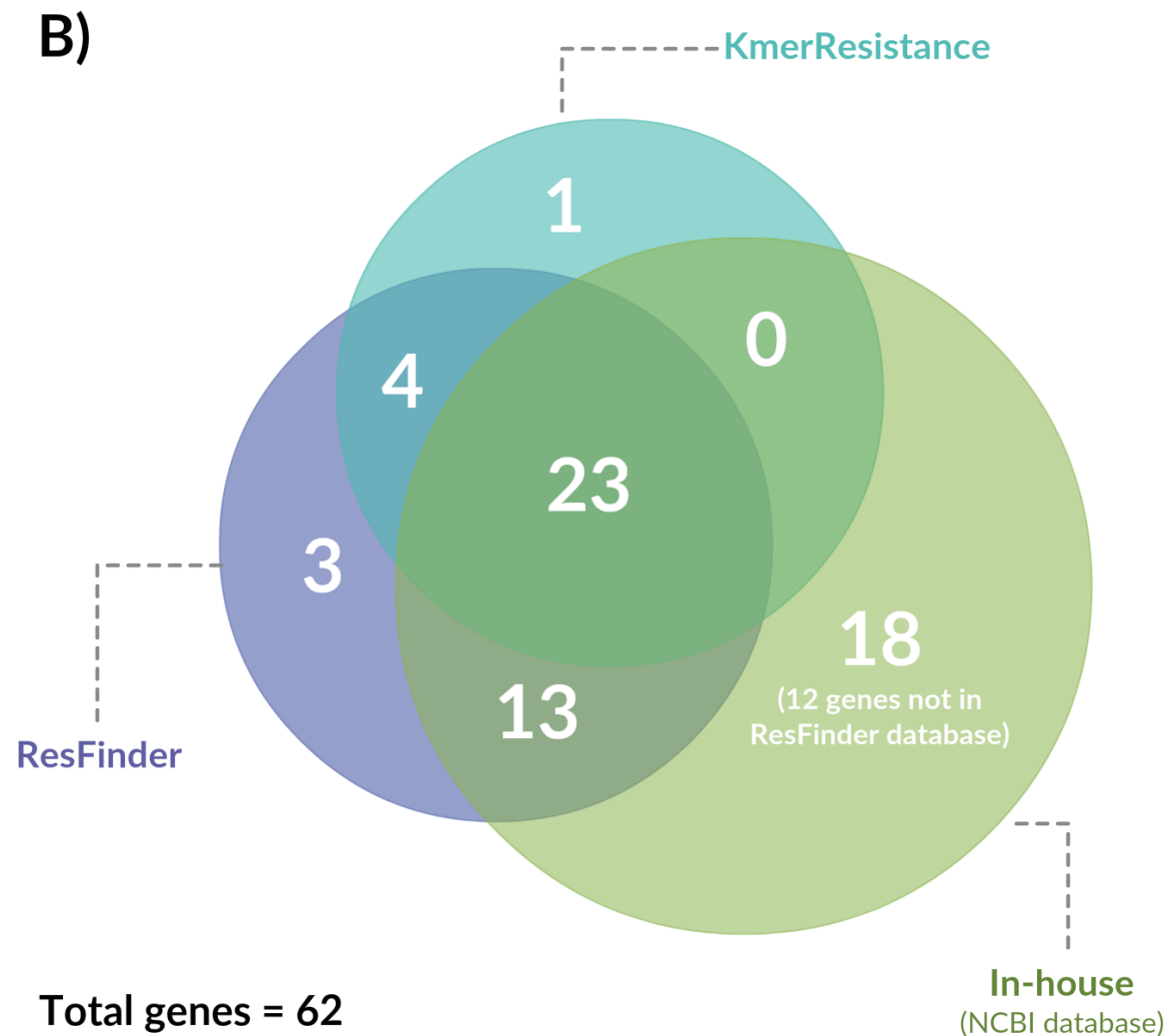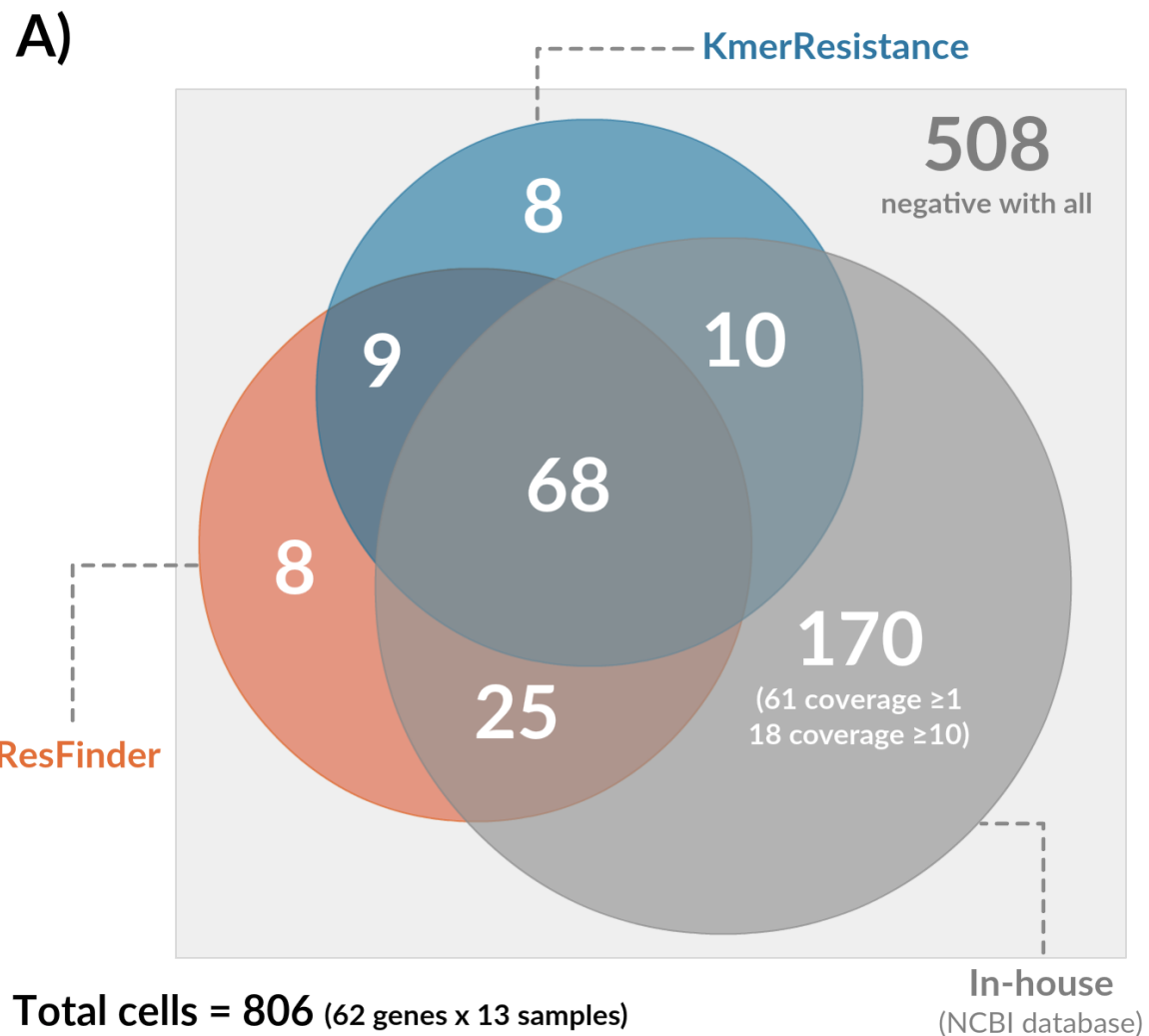
Figure 1.

Figure 2.

Figure 3.

A)

KmerResistance

508
negative with all

8

9

10

68

ResFinder

8

170
(61 coverage ≥1
18 coverage ≥10)

25

In-house
(NCBI database)

Total cells = 806 (62 genes x 13 samples)

B)

KmerResistance

1

4

0

23

ResFinder

3

13

18
(12 genes not in
ResFinder database)

In-house
(NCBI database)

Total genes = 62

Figure 4.