3

4   **Title**

5   The genomic epidemiology of *Escherichia albertii* infecting humans and birds in Great Britain

6

7   **Author list**

8   Rebecca J. Bengtsson[1], Kate S Baker[#1], Andrew A. Cunningham[2], David R. Greig[3], Shinto K. John[2],

9   Shaheed K. Macgregor[4], Katharina Seilern-Moy[2], Simon Spiro[4], Charlotte C Chong[1], Malaka De Silva[1],

10   Claire Jenkins[3], Becki Lawson[2]

11

12   **Affiliations**

13   [1] Clinical Infection, Microbiology and Immunology, Institute of Infection, Veterinary and Ecological

14   Sciences, The University of Liverpool, Liverpool, UK

15   [2] Institute of Zoology, Zoological Society of London, Regent's Park, London, NW1 4RY

16   [3] Gastrointestinal and Food Safety (One Health) Division, UK Health Security Agency, Colindale,

17   London, UK

18   [4] Wildlife Health Services, Zoological Society of London, Regent's Park, London, NW1 4RY

19   [#] Corresponding author. Email: Kate Baker, kbaker@liverpool.ac.uk

20

21   **Abstract**

22   *Escherichia albertii* is a recently identified gastrointestinal bacterial pathogen of humans and animals

23   which is typically misidentified as pathotypes of diarrhoeagenic *Escherichia coli* or *Shigella* species,

24   and is generally only detected during genomic surveillance of other Enterobacteriaceae. The incidence

25   of *E. albertii* is likely underestimated and its epidemiology and clinical relevance are poorly

26   characterised. Here, we whole genome sequenced *E. albertii* isolates from humans (*n*=83) and birds

27   (*n*=79) isolated in Great Britain between 2000 and 2021 and analysed these alongside a broader public

28   dataset (*n*=475) to address these gaps. We found human and avian isolates typically (90%; 148/164)

29   belonged to host-associated monophyletic groups with distinct virulence and antimicrobial resistance

30   profiles. Overlaid patient epidemiological data suggested that human infection was likely related to

31   travel and possibly foodborne transmission. The Shiga toxin encoding *stx2f* gene was associated with

32   clinical disease (OR=10.27, 95% CI=2.98-35.45 *p* = 0.0002) in finches. Our results suggest that improved

33   future surveillance will further elucidate disease ecology and public and animal health risks associated

34   with *E. albertii*.

35

36

37

38

39 **Introduction**

40 *Escherichia albertii*, a Gram-negative gastrointestinal pathogen of humans and animals, was first

41 confirmed as a novel bacterium in 2003 (1-3). This pathogen is often mis-identified because it is

42 difficult to differentiate from *Shigella* species as they are morphologically, colonially, metabolically

43 and biochemically similar, for example both are non-lactose fermenting and lysine decarboxylase

44 negative (4, 5). Implementation of PCR for the detection of a wide range of gastrointestinal (GI)

45 pathogens including diarrhoeagenic *E. coli* (DEC), and the use of whole genome sequencing (WGS) for

46 identification and typing, has provided a more robust and reliable approach for the detection and

47 characterisation of *E. albertii* (6-8). As a result of the implementation of routine WGS for

48 microbiological surveillance by the United Kingdom Health Security Agency (UKHSA, formerly Public

49 Health England), we now have the capacity to accurately identify *E. albertii* in individuals presenting

50 to primary healthcare settings with gastrointestinal symptoms.

51

52 Although detection and speciation prior to the genomic era was challenging, the pathogenic traits of

53 *E. albertii* are well described (4). Like certain DEC pathotypes, specifically the enteropathogenic *E. coli*

54 (EPEC) and a subset of Shiga toxin-producing *E. coli* (STEC), the genome of *E. albertii* contains the locus

55 of enterocyte effacement (LEE) pathogenicity island encoding a type III secretion system involved in

56 attachment of the pathogen to the gut mucosa (9, 10). Colonisation of EPEC and *eae* gene positive (a

57 marker of LEE) STEC in both humans and animals can lead to the formation of attaching and effacing

58 (A/E) lesions on the intestinal epithelial cells (11). Cytolethal distending toxin (CDT) is encoded by the

59 *cdtABC* operon and is classified into five subtypes based on sequence variation of the *cdtB* gene (*cdtB*-

60 I to *cdtB*-V). Of these, *cdtB* subtypes I/II/III/V have been identified in *E. albertii* (10, 12). The *stx* gene

61 encoding for Shiga toxins, predominantly the *stx2f* subtype, has been found in certain strains of *E.*

62 *albertii* (13). Although these virulence determinants are well described in *E. albertii,* their distribution

63 and clinical relevance across species requires further elucidation.

64

65 Clinical symptoms in human patients caused by *E. albertii* infection are similar to those caused by EPEC

66 and typically include watery diarrhoea, dehydration, abdominal pain, vomiting and fever (13, 14). Over

67 the last decade, outbreaks of GI disease in people in Japan have been attributed to *E. albertii* following

68 re-examination of the original microbiological findings using genomic typing methods, such as

69   multilocus sequence typing (MLST) (15, 16). However, because of the challenges around detection and

70   identification which have hampered systematic surveillance, data on the epidemiology, source and

71   transmission routes of *E. albertii* infections are sparse. In England to date, commercial GI PCR panels

72   have been adopted by approximately 25% of diagnostic microbiology laboratories in the National

73   Health Service network (17). Furthermore, not all the commercial GI PCR panels target *eae*, and not

74   all diagnostic laboratories refer samples to the Gastrointestinal Bacteria Reference Unit (GBRU) at

75   UKHSA for further identification. These limitations of the current surveillance mechanisms likely result

76   in a considerable under ascertainment of cases, and the true burden of human infection caused by *E.*

77   *albertii* remains unknown.

78

79   In addition to infecting people, *E. albertii* can infect birds and other animals, in which the prevalence

80   and pathogenicity is also unclear. In the mid-1990s, multiple mortality incidents of Fringillidae (finch)

81   species were observed in Scotland with a bacterium, later identified as *E. albertii,* hypothesised to be

82   the cause of death (2, 18, 19). Similarly, in 2004, large-scale mortality of a finch species (*Carduelis*

83   *flammea*) occurred in Alaska, United States of America (USA), with *E. albertii* as the probable aetiology

84   (2). Active molecular surveillance studies for *E. albertii* have since detected the bacterium in dead and

85   apparently healthy birds of multiple orders and species from Australia, Asia, mainland Europe and

86   North America (2, 20-23). *Escherichia albertii* also has been detected in poultry faeces/GI tract

87   contents and meat (8, 24-26) and in domestic mammals (e.g. pig, cat) and both terrestrial and marine

88   wild mammal species (e.g. raccoon, seal, bat) (8, 27, 28). Although the occurrence and significance to

89   mammal host health remains uncertain, there is a growing body of evidence that avian hosts may act

90   as a reservoir of infection (21, 26). Thus, the extent of associated disease in birds and the relationship

91   of bird and human infections requires further investigation.

92

93   Here, we performed WGS analysis on *E. albertii* isolates from humans and birds in Great Britain (GB)

94   from archives held at the UKHSA and the Zoological Society of London (ZSL), respectively, to

95   investigate the epidemiology of this recently identified pathogen. The aims of the study were to

96   integrate the phylogenetic and epidemiological data in order to gain insights into the ecology of *E.*

97   *albertii* among people and birds, to better understand the risk factors (e.g. recent international travel)

98   associated with human infection, and to infer the likely significance of *E. albertii* infection to avian

99   host health. Owing to the relative importance of *Enterobacteriaceae* as a reservoir for antimicrobial

100   resistance (AMR) genes, we also describe and compare genotypic AMR profiles recovered from the

101   two host groups.

102

Results

Summary of the human isolates

Between January 2014 (when routine WGS was first implemented at the GBRU) and December 2021, 83 isolates from human cases were confirmed as *E. albertii*. Over this 8-year period, between 4 and 23 isolates were identified per year (Supplementary Figure 1). Metadata regarding patient gender, age and history of recent travel were available for 82, 83 and 26 isolates, respectively (Supplementary Data 1). There was no statistical association of isolates with gender (39 males, 32 females) and little association with age group (Table 1/Figure 1/Supplementary Figure 1). A total of 24 (29%, *n*=24/83) patients stated they had recently travelled (within 7 days of onset of symptoms) outside the UK, of which the majority (*n*=21/24, 88%) reported travel to Asia. Travel status was unknown for the remaining cases (71%, *n*=59/83), as their travel history was not recorded.

Summary of the bird isolates

Seventy-four *E. albertii* isolates from wild birds were analysed over the period 2000-2019 inclusive. With a single exception (tawny owl *Strix aluco*), the hosts were Passeriformes from the following families in declining rank order: Fringillidae n=50, Passeridae n=8, Turdidae n=7, Paridae n=4 and single birds from the Hirundinidae, Motacillidae, Prunellidae, and Sturnidae (for species composition see Supplementary Data 2). Isolates were identified each year across the 20-year study period with two exceptions and from a total of 72 sites. Available data permitted determination of the inferred significance of *E. albertii* to host health (see Supplementary methods) for 69 wild birds; with 38% (*n*=26) being significant, 46% (*n*=32) being equivocal and 16% (*n*=11) being incidental. The wild birds for which *E. albertii* infection was considered significant to host health comprised Fringillidae (bullfinch *Pyrrula pyrrhula* n=1, chaffinch *Fringilla coelebs* n=4, greenfinch *Chloris chloris* n=9 and siskin *Spinus spinus* n=8), house sparrow *Passer domesticus* n=3 and a single blue tit *Cyanistes caeruleus*.

The five isolates from captive zoo birds were from a diverse range of species (Anseriformes, Passeriformes, Pelecaniformes, and Sphenisciformes). Inferred significance to host health was categorised as significant for one captive bird (black-footed penguin *Spheniscus demursus*), equivocal for two cases and incidental for two cases.

Fringillidae were more frequently associated with 'significant' inferred clinical significance than non-Fringillidae species combined (*p* = 0.0433, Fisher's exact test) for the wild and captive bird data, and this was also well supported statistically among wild birds alone (p=0.0612).

137 *Genomic epidemiology of Escherichia albertii from humans and bird isolates*

138 To explore the genomic epidemiology of *E. albertii* among the human and bird isolates from GB,

139 demographic features were overlaid on the bacterial population structure and statistical support for

140 associations with metadata variables were evaluated.

141

142 Specifically, to determine the population structure of *E. albertii* within our dataset, a maximum

143 likelihood phylogeny was constructed based on a SNP alignment of 26,594 bp (Figure 1). BAPS

144 identified eight clusters consistent with monophyletic clustering, with the exception of BAPS cluster

145 8, which was split across multiple regions of the tree (Table 1, Figure 1). Combining the epidemiological

146 information with this population structure revealed distinct and separate phylogenetic clustering of

147 bird and human isolates (p<0.0001, Chi-square test, 7 df), although statistical support varied for

148 individual clusters (see Table 1). Most bird isolates (*n*=74/79) belonged to BAPS clusters 6 and 7 in

149 which bird isolates were statistically over-represented, and these were termed Bird-Associated

150 Clusters (BACs, Table 1). To facilitate further high-level investigation, BAPS clusters 1,2,3,4,5 and 8

151 were termed Human-Associated Clusters (HACs). Intermixing between human and bird isolates was

152 observed within both BACs and one HAC. Specifically, the HAC BAPS 8 contained 6% (*n*=5/79) of

153 isolates from birds, 4/5 of which were from captive zoo birds. Within the BACs 6 and 7, 18% (*n*=16/90)

154 of isolates were from humans.

155

156 To investigate the association of *E. albertii* with human demographic features, we associated travel

157 history and patient age with the bacterial population structure. All 24 isolates from human patients

158 with a confirmed recent history of international travel belonged to HACs, and at least one travel-

159 associated isolate was identified in each of the six HACs (Figure 1). The travel status was not recorded

160 for any of the human cases with isolates that fell within the BACs. When associating human age groups

161 with population cluster assignation (BAC/HAC), we observed a significant difference between the BACs

162 and HACs (*p* = 0.0008, Fisher's exact, Figure 1, Supplementary Data 3). Within the BACs, infant (<2

163 years) and older people (60> years) were the predominant human age groups, comprising 44%

164 (*n*=7/16) and 31% (*n*=5/16) of human isolates respectively (where patient age information was

165 available, Supplementary Data 1). In contrast, the predominant age group within the HACs was adult

166 (16-60 years) comprising 55% (*n*=37/67) of human isolates.

167

168 *Virulence profiles and associations with disease in bird hosts*

169 The *eae* gene was present in all but one isolate within the dataset, and the *cdtA*, *cdtB*, *cdtC* genes were

170 present in >94% (*n*=153/162) isolates (Figure 2, Supplementary Figure 2). The *stx2f* gene was detected

171 in 38 isolates, the majority (*n*=37/38, 97%) of which were from wild birds in BACs, except for one

172 human isolate (SRR6144114) belonging in BAPS 8. Among the wild birds, *stx2f* resulted in an increased

173 odds of inferred clinical significance of infection (relative to equivocal and incidental combined)

174 (OR=10.27, 95% CI=2.98-35.45 *p* = 0.0002). There was little evidence for confounding of the disease

175 association by bird family (Fringillidae/Non-Fringillidae, Adjusted OR 10.25 95% CI 2.66 – 92.78), a

176 possible effect modification of the bird family (Strata specific OR: OR=12.68, 95% CI=2.66 – 877.38, p-

177 value <0.001 (Fringillidae), OR=0.64 95% CI 0.03 – 16.03, p-value=1). This was challenging to evaluate

178 further as the *stx2f* was over-represented among the Fringillidae (vs non-Fringillidae OR=25.67, 95%

179 CI=5.35-123.23. *p* = 0.0001), specifically of 37 *stx2f*-positive bird isolates, 35 were from Fringillidae

180 species.

181

182 *Antimicrobial resistance profiles in human and bird isolates*

183 To investigate the genotypic predictors of AMR among *E. albertii* isolates in this dataset, we looked

184 for the presence of genetic determinants of AMR. Both horizontally acquired antimicrobial resistance

185 genes (ARGs) and vertically inherited point mutations known to confer resistance or reduced

186 susceptibility to various antimicrobials in *E. coli* were identified. ARGs were exclusively identified in

187 human isolates, except for one captive zoo bird isolate in HAC BAPS cluster 8 (SRR13092475). Overall,

188 human isolates were observed to carry more AMR genetic determinants compared to bird isolates,

189 including a total of 25 ARGs and five point mutations associated with resistance or reduced

190 susceptibility to 10 different antimicrobial drug classes. In contrast, only three point mutations were

191 identified among the bird isolates, with the exception of the aforementioned captive bird isolate

192 (SRR13092475) carrying an additional 6 ARGs associated with resistance to 7 antimicrobial drug

193 classes. Point mutations were more frequent than ARGs, but the implications less clear. Specifically,

194 *uhpT* E350Q and I355T (Figure 3a), predicted to confer resistance against fosfomycin and quinolone,

195 respectively (48), were identified in all human and bird isolates, with the multidrug-resistance

196 associated *marR* S3N point mutation being identified in the majority (*n*=157/162, 97%) of isolates.

197

198 There were 18 unique genotype profiles, including three dominant profiles identified in 80% (*n*=66/83)

199 of the isolates (Figure 3b). Correlating ARGs with the phylogeny revealed that the majority (14/16) of

200 isolates within the HAC BAPS 4 had the ARGs *blaDHA-1*, *blaTEM-1*, *dfrA17*, *mph(A)*, *qnrB4*, *sul1* and

201 *tet(A)* (Figure 1). Among these, *mph(A)*, *sul1*, *blaDHA-1* and *qnrB4* were present on a single contig in

202 multiple isolates, the longest of which was 14,961 bp. A BLASTn search of this contiguous sequence

203 revealed 100% coverage and identity with plasmids from multiple *E. coli* strains, *Shigella sonnei* and *S.*

204 *flexneri* (Supplementary Data 3). Single contiguous sequences containing the 4 ARGs were identified

205 in 13 isolates, all belonging to BAPS cluster 4. A single point mutation in the quinolone resistance

206 determining region (QRDR) of gyrA, S83L, was present in 60% (43/72) of HACs isolates (though this

207 was not present in BAPS 3) and only 1% (1/90) of isolates in BACs (Figure 1).

208

209 The genotypic AMR profile among human isolates was further explored through phenotypic testing.

210 We selected 11 *E. albertii* (HAC *n*= 7, BAC *n*= 4) isolates that captured the lineage and genotypic AMR

211 diversity across the phylogenetic tree and determined their antimicrobial resistance profiles against

212 cefoxitin, ceftriaxone, fosfomycin, tetracycline, ampicillin, ciprofloxacin, chloramphenicol and

213 rifampicin, to review the phenotypic consequences of mutations identified in this study (Figure 1 & 3).

214 The presence of ARGs *tet(A)*, *blaTEM-1* and *qnrB4* conferred resistance to tetracycline, beta-lactam

215 and fluoroquinolone class antibiotics, respectively (Table 2). The presence of ARG *blaDHA-1* did not

216 confer resistance to the cephalosporin class antibiotics, cefoxitin and ceftriaxone, in this isolate set.

217 Point mutations *uhpA_G97D** and *uhpT_E350Q**, when present together, as well as point mutations

218 in *gyrA_S83L*,* were associated with resistance/decreased susceptibility to their related antimicrobial

219 classes (fosfomycin and fluroquinolones respectively). Point mutations in *marR_S3N** and

220 *parE_I355T** were present in the majority of isolates tested in this study set, and resistance profiles

221 were consistent across the dataset and impacted by the additional presence of other ARGs or point

222 mutations (Table 2).

223

224 *Global contextualisation of E. albertii from GB*

225 To place the human and bird *E. albertii* isolates from GB within the global context, we expanded the

226 analysis to include additional isolates retrieved from publicly available data (n=475, Methods,

227 Supplementary Data 4). A cgMLST tree was generated based on hierarchical clustering of 2513 gene

228 loci. These additional isolates were derived from diverse sources (22% human; 47% Avian ['poultry',

229 'non-poultry' and 'not defined']; 7% mammal (e.g. livestock, wild species and companion species); 1%

230 food; 2% water and 21% undescribed sources) and locations (18% Americas, 16% Europe, 41% Asia,

231 1% Africa, 1% Oceania, and remaining 23% unknown). We correlated the position of BAPS clustered

232 isolates from the current study in this broader context (meaning, notably, that BAPS notation is

233 specific to the BACs and HACs groupings of *E. albertii* isolates from GB**).** We observed that isolates

234 from GB were dispersed across most parts of the cgMLST tree, indicating that these isolates capture

235 much of the known diversity of *E. albertii*. The cgMLST tree also revealed that while isolates belonging

236 to the HACs BAPS 2, 4 and 5 remained largely within individual clades of the tree alongside other

237 human-derived isolates (Figure 4), isolates from HAC BAPS 3 clustered with poultry-derived isolates

238 from Asia and the USA. The majority of isolates from the wild BAC BAPS 7 were similarly embedded

239 within a cluster, this time dominated by poultry isolates from Asia. However, isolates from BAC BAPS
240 6 and HAC BAPS 8 appeared in multiple clades intermixed with isolates derived from various sources.
241 This is consistent with their greater phylogenetic distance relative to other BAPS clusters (particularly
242 the polyphyletic BAPS 8, Table 1) and suggests that the association of these two BAPS clusters as bird-
243 and human-associated may be less clear.

244

245 **Discussion**

246 The notification of cases of GI disease caused by *E. albertii* in GB in both humans and animals is
247 currently low compared to other well-established pathogens, such as *Campylobacter* and *Salmonella*
248 species (49, 50). However, it is likely that the number of *E. albertii* diagnoses will increase in line with
249 improvements in molecular diagnostics and the wider adoption of PCR and WGS as tools for GI
250 pathogen surveillance. Thus, analysing current data to understand the potential public health burden,
251 clinical significance, and risk factors in human and animal hosts will guide future research and
252 surveillance.

253

254 Although epidemiological follow-up is not conducted for *E. albertii*, the patterns we observed for *E.
255 albertii* infection in people are consistent with similar transmission routes and risk factors as other GI
256 pathogens. Specifically, a similar proportion of reported travel-association (31%, 24/83) with other
257 travel-associated *Enterobactericeae*, including *Shigella* (19 - 50% for the years 2005 - 2014) (51) and
258 *Salmonella* (19 - 32% for the years 2005 - 2014) (52). We also explored whether, like *Salmonella* and
259 *Campylobacter* species (53), zoonotic infection might contribute to disease transmission. Our
260 observations that GB human and bird isolates belonged primarily to host-associated monophyletic
261 groups, and had distinct and convergent accessory genome features (e.g. with HACs containing or
262 acquiring ARGs and the occurrence of *stxf2* in BACs) do not support substantial cross-species
263 transmission (i.e. zoonotic or anthroponotic) between birds and humans. The acquisition of ARGs in
264 HACs of *E. albertii*, however, may have been confounded by geography as many patients had recently
265 travelled to Asia, a known risk factor for enteric pathogen ARG acquisition, and where convergent
266 evolution of QRDR mutations is reported (54, 55).

267

268 Although our data are not supportive of extensive zoonosis for *E. albertii,* the existing evidence
269 supports reinforcing public health messaging. Specifically, the human isolates grouped in the BAC
270 BAPS 6 (n=13) were not very closely related with bird isolates in BAPS 6 (Figure 1). Comparatively, the
271 human isolates that grouped in BAPS 7 (n=4) had higher similarity with avian isolates in BAPS 7,
272 possibly indicating the occurrence of zoonotic transmission (Figure 1, Table 1). Supplementary feeding

273 of garden birds is a common pastime in GB that results in a close human-wildlife interface (56) and
274 zoonotic infection has been suggested for other bacterial pathogens of wild birds (57, 58).
275 Furthermore, humans infected with isolates belonging to BAC were typically very young or older
276 people, consistent with bias towards infant infection; previously described for wildlife-associated
277 *Salmonella* Typhimurium and *S*. Enteritidis biotypes (58, 59). Hence, some (n=4) human isolates
278 conceivably represent zoonotic infections, reinforcing the need for good hygiene measures (e.g. hand
279 washing after handling bird feeders) when feeding garden birds (58).
280
281 Four of five captive zoo bird isolates clustered within monophyletic subclades of the HAC BAPS8.
282 However, similar to the human/bird mixing observed within BAPS 6 (see above), the large genomic
283 divergence among isolates in BAPS 8 is not indicative of direct anthroponotic transmission (Table 1)
284 and there are other potential sources of *E. albertii* infection for captive birds (e.g. diet, wild birds).
285
286 Our study did not strongly support evidence of zoonotic infection in contrast to recent studies from
287 China, Japan and the USA that highlighted the potential for foodborne transmission of *E. albertii* to
288 humans via the consumption of poultry (8, 24, 26). Incorporating public data revealed that one HAC
289 (BAPS 3) admixed with poultry-associated isolates from China and the USA (Figure 4), indicating the
290 possibility that *E. albertii* infections may be a foodborne illness linked to eating poultry either
291 domestically or overseas. The cluster supporting potential zoonotic infection from our study (BAPS 7)
292 also encompassed a broader group of poultry isolates, possibly indicating longer-term transmission
293 among wild birds, poultry, and humans for some lineages. In contrast, HACs BAPS 2 and BAPS 4 were
294 on long branches without close associations with other hosts or regions (Figure 3), potentially
295 indicating an unobserved reservoir of infection, either overseas and/or in non-human hosts. The
296 emerging picture of *E. albertii* as a travel-associated pathogen with a potential reservoir in poultry
297 parallels other enteric pathogens, including *Salmonella* and *Campylobacter* (49, 50). Therefore,
298 genomic surveillance of *E. albertii* in more locations and potential reservoir hosts is needed to further
299 elucidate this pathogen's ecology.
300
301 Results from this study, combined with the published literature (2, 21, 24, 60), indicate that avian
302 hosts are likely to play a larger role in the epidemiology of *E. albertii* than other (e.g. mammalian)
303 hosts. Analysis of publicly available isolates revealed that comparatively few isolates were derived
304 from non-human mammals relative to birds (7% vs 47% respectively). Although public data are not a
305 reflection of representative surveillance, unpublished data from the ZSL provide a similar picture.
306 While the same microbiological protocol has been used across clinical and routine health check

307   samples from a diverse taxonomic range of birds and mammals held in the ZSL zoological collections
308   since 1991, *E. albertii* has only been identified from five captive birds and not from mammals.
309   Furthermore, there have been no confirmed *E. albertii* detections from livestock or wildlife species in
310   disease surveillance conducted by the Animal Plant & Health Agency (APHA) in England and Wales for
311   23 years. Although there are limitations to the APHA and ZSL *E. albertii* surveillance (e.g. APHA routine
312   microbiology relies primarily on phenotypic and biochemical characterisation meaning *E. albertii* may
313   be present but not detected; the ZSL captive collections are limited to two sites; the ZSL national wild
314   bird surveillance was skewed towards passerines), an absence of isolation from non-human mammals
315   supports a primary avian reservoir. However, targeted surveillance with broad spatial and taxonomic
316   coverage is required to further explore this hypothesis.

317

318   Our study also identified implications of *E. albertii* for bird health, with infection being more frequently
319   associated with significant disease in finch than in non-finch species. This is consistent with historical
320   investigations of multiple mortality incidents of finches in Scotland and the USA (2, 19) and supports
321   the hypothesis that it acts as a primary pathogen in these birds. This familial bias may relate to host
322   or environmental factors (e.g. differential exposure or susceptibility) as well as pathogen factors. Our
323   data support the latter, with a possible role for differences in virulence determinant components
324   among circulating *E. albertii* strains affecting infection outcomes. Specifically, isolates containing the
325   *stx2f* virulence factor were associated with finch hosts (Fringillidae), and infection in these birds was
326   significantly more likely to be associated with disease. This relationship could not be disentangled
327   further owing to the low occurrence of *stx2f*-bearing strains from non-finch species, but it is possible
328   that finches may act as a reservoir of *stx2f*-positive *E. albertii,* as is hypothesised for garden bird-
329   associated biotypes of *Salmonella* Typhimurium (30). Infection with *E. albertii* was also inferred as
330   having possible health impacts on other bird species; further surveillance is required to explore this
331   further.

332

333   In conclusion, poor molecular diagnostic capabilities for *E. albertii* in human and animal health
334   laboratories mean the true burden of *E. albertii* infection is likely underestimated and the lack of
335   systematic surveillance data means that clinical severity and exposure risks are largely unknown.
336   However, we leveraged available data to highlight the likely relevance of travel to regions with a high
337   risk of GI infections, including an association with AMR, and a potential zoonotic component that is
338   likely bird-associated, apparently more so with poultry than with wild bird species. To improve
339   surveillance for *E. albertii*, we recommend increased deployment of molecular diagnostics in medical
340   and veterinary laboratories, in conjunction with the systematic collection of epidemiological data.

341 Maintaining close collaborations between public health and veterinary institutions is essential to
342 better understand the source, transmission and risks to animal and public health of this recently-
343 identified pathogen.

344

345 **Methods**
346 *Human isolates and epidemiological data collection*
347 Diagnostic algorithms for the detection of *E. albertii* are not included in the UK Standard Microbiology
348 Investigation of Gastroenteritis protocols used by local hospital diagnostic laboratories
349 (https://www.gov.uk/government/publications/smi-s-7-gastroenteritis-and-diarrhoea).   Between
350 2014 and 2021, isolates from routine gastrointestinal surveillance including faecal specimens from
351 hospitalised cases or cases in the community, were either submitted to the GBRU at UKHSA from local
352 hospital diagnostic laboratories in England having been mis-identified as *Shigella* species or DEC, or
353 were cultured from faecal specimens sent to GBRU for molecular testing. At GBRU, bacteria cultured
354 from faecal specimens on MacConkey agar following aerobic incubation overnight, were tested for
355 virulence genes that define the different pathotypes of DEC using PCR, including *eae* which is a
356 characteristic of EPEC, STEC and *E. albertii* (7).

357

358 All *eae*-positive isolates were genome sequenced and bacterial identification was confirmed from the
359 genome using a kmer-based approach, as described previously (29). In total, all 83 isolates identified
360 as *E. albertii* using this approach were included in this study (Supplementary Data 1). Where available,
361 human isolates were linked to demographic data, including age category, gender, and travel history
362 (Supplementary Data 1).

363

364 *Bird isolates and epidemiological data collection*
365 Wild bird derived *E. albertii* isolates (n=74) were obtained through scanning surveillance of dead wild
366 birds conducted by ZSL over the period 2000-2019 inclusive (Supplementary Data 2). Members of the
367 public reported observations of wild bird mortality, typically in the vicinity of garden bird feeding
368 stations, consequently, the species coverage was predominantly small passerines (e.g. Fringillidae,
369 Paridae, Passeridae, Turdidae) and columbids, which commonly visit peri-domestic habitats in Great
370 Britain. Carcasses were submitted from a subset of mortality incidents for post-mortem examination.
371 Coverage was across Great Britain, although the majority of wild bird submissions and those from
372 which *E. albertii* was isolated were from England (England n=63 isolates, Wales n=6, Scotland n=5).
373 Post-mortem investigations were conducted following a standardised protocol, supported by
374 parasitological and microbiological examination as routine, combined with histological examination

375 and other ancillary diagnostic testing as indicated based on macroscopic abnormalities. Liver and small

376 intestinal tract contents were routinely sampled for microbiological examination using a standardised

377 protocol (30). Semitranslucent, butyrous, non-lactose fermenting and oxidase negative colonies of

378 Gram-negative rods to coccobacilli were subjected to an Analytical Profile Index 20E biochemical test

379 (bioMerieux): isolates tentatively identified as *E. albertii* were cryo-archived at -80 degrees C. Where

380 *E. albertii* was isolated from multiple wild birds examined from the same mortality incident, a single

381 isolate was submitted to GBRU with two exceptions where two isolates were typed. An available

382 archive of similarly identified *E. albertii* isolates from clinical examinations (n=2) and post-mortem

383 examinations (n=3) of captive birds in the zoological collection at ZSL was also included

384 (Supplementary Data 2). Additionally, a single *E. albertii* isolate was identified from a sample of small

385 intestinal tract contents collected from a dead wild bird examined post-mortem using the UKHSA

386 diagnostic algorithm for human faecal samples.

387

388 The inferred significance of *E. albertii* infection to wild and captive zoo bird health (i.e. its likely

389 contribution to the cause of death) was classified as significant, equivocal, or incidental, based on

390 review of the incident history and the pathological, microbiological and parasitological findings for

391 those examined post mortem (see Supplementary Methods for full definitions). For the two captive

392 zoo birds with *E. albertii* isolated from clinical samples, the history and ancillary diagnostic test results

393 were also appraised to infer likely isolate significance to host health.

394
395 *Genome sequencing and quality control*

396 Isolates of *E. albertii* from UKHSA and ZSL were sequenced at GBRU according to previously described

397 protocols (29) and deposited in the Sequence Read Archive (SRA) under the bioproject accession

398 PRJNA315192 with the SRA accession numbers of individual isolates listed in Supplementary Data 1.

399 Short read sequences were retrieved from the SRA and processed using Trimmomatic v0.38 (31) to

400 trim adaptors and filter low-quality bases. FastQC v0.11.6

401 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and MultiQC v1.7 (32) were used to

402 assess the quality of reads.

403

404 *Phylogenetic and clustering analysis*

405 Processed reads were mapped to the *E. albertii* strain 1551-2 reference genome (GenBank accession

406 CP025317) (33) using BWA mem v0.7.17 (34). Alignment files were sorted and filtered using the

407 SAMtools suite v1.9-47 (35), PCR duplicates were marked using Picard v2.21.1-SNAPSHOT

408 MarkDuplicates (http://broadinstitute.github.io/picard/). The BCFtools suite v1.9-80 (35) was used to

409 identify sequence variants and filter variant files, in which low quality single nucleotide polymorphisms

410   (SNPs) were removed if mapping quality <60, Phred-scaled quality score <30, read depth <10 and

411   variant allele frequency <0.7.

412

413   BCFtools consensus was used to generate reference-based pseudogenomes for each isolate from the

414   filtered SNP variants. Regions containing insertion sequences and phages (identified using the

415   PHASTER web server https://phaster.ca/) were identified from the reference genome and masked

416   using BEDTools v2.28.0 maskfasta (36). Regions with read depth of <10 were also masked. The masked

417   pseudogenomes were concatenated and provided as an alignment for Gubbins v2.3.4 (37) to identify

418   and mask regions of putative recombination (Supplementary Figure 2). Following Gubbins, SNP-sites

419   v2.4.1 (38) was used to extract variant sites, producing a final SNP-alignment of 26,594 bp in length.

420   This SNP-alignment was used to construct a maximum-likelihood phylogenetic tree using IQ-TREE

421   v2.0-rc2 (39), constructed based on the FreeRate nucleotide substitution, invariable site, and

422   ascertainment bias correction model with 1000 bootstrap replicates. The phylogenetic tree was

423   midpoint rooted and visualised using interactive Tree of Life (iTOL) v6.5 (40).

424

425   RhierBAPS v1.1.3 (41) was used to identify clusters of genetically similar isolates among the SNP-

426   alignment, termed Bayesian Analysis of Population Structure (BAPS) clusters.

427

428   *Construction of cgMLST tree with publicly available data*

429   To deepen the insights gained from the UKHSA and ZSL *E. albertii* isolates, we analysed their genome

430   sequences in the context of publicly available *E. albertii* sequence data. Specifically, additional publicly

431   available *E. albertii* genome sequences accessible through Enterobase on the 7th of February 2022

432   (n=475) were constructed alongside the data above into a core genome Multi Locus Sequence Type

433   (cgMLST) tree using hierarchical clustering (HeirCC) (42). Minimal metadata on source and country of

434   origin was extracted from Enterobase alongside HeirCC level classifications and visualised over the

435   unrooted cgMLST tree using interactive Tree of Life (iTOL) v6.5 (40). Metadata on isolate origin was

436   manually curated into the following categories: human, avian (poultry, non-poultry and not defined);

437   mammal (livestock, wildlife and companion species); food, water and undescribed sources.

438

439   *AMR and virulence gene analysis*

440   Draft genomes were assembled de novo from processed short read sequences using Unicycler v0.4.7

441   (43) with –min_fasta_length set to 200. Qualities of the draft assemblies were assessed with QUAST

442   v5.0.2 (44) and were all within the assembly quality standards of EnteroBase for *Escherichia* (42).

443   Prokka v1.13.3 (45) was used to annotate draft genome sequences.

444

445 The presence of known genetic determinants of AMR was detected using AMRFinderPlus v3.9.3 (46)

446 and screened against the Pathogen Detection Reference Gene Catalog

447 (https://www.ncbi.nlm.nih.gov/pathogens/). AMRFinderPlus was run with the organism-specific

448 option for *Escherichia* and screening for both point mutations and genes (with 80% coverage and 90%

449 identity threshold applied). AMR resistance profiles were visualised with UpSetR v2.1.3 (47).

450

451 Association of known AMR genes with related plasmid sequence were identified by extracting AMR-

452 gene containing contiguous sequences from draft genome assemblies and comparison against the

453 NCBI nonredundant database using MegaBlast.

454

455 Detection of virulence genes was performed using ABRicate (https://github.com/tseemann/abricate),

456 by which draft genomes were screened against the Virulence Factor Database with minimum

457 nucleotide identity of 80% and minimum coverage of 60%. This screen comprised of virulence genes

458 associated with *E. albertii* including *stx*, *eae* and *cdtABC* genes that encode Shiga toxin, intimin and

459 CDT.

460

461 *Statistical testing*

462 Statistical support for phylogenetic clustering of bird and human isolates was evaluated with chi-

463 square testing on: 1] the proportion of human isolates in individual clusters (Table 1) and 2]

464 associations of human associated clusters (HACs) and bird associated clusters (BACs) with patient age

465 (categorised into infant [<2 yrs], children [2-15 yrs], adult [16-60 yrs] and older people [60>yrs]). In

466 the finch (Fringillidae) hosts, any significance between the presence of *stx*2f and clinically significant

467 *E. albertii* infection was also evaluated using the Fisher's exact test. Adjusted and strata-specific odds

468 ratios for the effect of bird family on the association between *stx2f* presence and inferred significant

469 disease was conducted using the Mantel-Haenszel Test. All statistical tests were performed using R

470 v4.0.3.

471

472 *Phenotypic antimicrobial resistance testing*

473 Minimum Inhibitory Concentration (MIC) determination was carried out using Lioflichem→ MIC test

474 strips (Liofilchem, Italy) following the manufacturer's guidelines. Bacterial inoculum for MIC testing

475 was prepared, following the EUCAST guidelines for *Enterobacterales* standard broth microdilution

476 (https://www.eucast.org/fileadmin/src/media/PDFs/EUCAST_files/Breakpoint_tables/v_11.0_Break

477 point_Tables.pdf) and was spread on Mueller Hinton Agar plates (Bio-Rad, France) using sterile cotton

478   swabs after which the MIC test strip was applied. Plates were incubated at 37 °C for 18 hours before

479   the readings were recorded.

480
481   **Ethics statement**

482
483   Samples were collected during post-mortem examination of wild birds found dead or euthanased for

484   welfare reasons under the Veterinary Surgeons Act 1966. Samples from humans . For data relating to

485   isolates from the United Kingdom Health Security Agency: No individual patient consent was required

486   or sought as UKHSA has authority to handle patient data for public health monitoring and infection

487   control under section 251 of the UK National Health Service Act of 2006 (previously section 60 of the

488   Health and Social Care Act of 2001

489

490   **Data availability**

491   Individual accessions numbers for isolates used in this study area available in Supplementary Data 1,
492   2, and 4. Phylogenetic trees from this study have been deposited in FigShare (DOI:
493   10.6084/m9.figshare.20894854.v1). The authors recognise that this study opens up important
494   further avenues for functional research of *E. albertii* and are happy to make isolates available on
495   request.

496

497   **Code availability**

498   No custom code was used in the analysis of this data.

499
500

501   **References**

502
503   1.   Huys G, Cnockaert M, Janda JM, Swings J. Escherichia albertii sp. nov., a
504   diarrhoeagenic species isolated from stool specimens of Bangladeshi children. Int J Syst Evol
505   Microbiol. 2003;53(Pt 3):807-10.
506   2.   Oaks JL, Besser TE, Walk ST, Gordon DM, Beckmen KB, Burek KA, et al.
507   Escherichia albertii in wild and domestic birds. Emerging infectious diseases.
508   2010;16(4):638.
509   3.   Ooka T, Seto K, Kawano K, Kobayashi H, Etoh Y, Ichihara S, et al. Clinical
510   significance of Escherichia albertii. Emerg Infect Dis. 2012;18(3):488-92.
511   4.   Hyma KE, Lacher DW, Nelson AM, Bumbaugh AC, Janda JM, Strockbine NA, et al.
512   Evolutionary genetics of a new pathogenic Escherichia species: Escherichia albertii and
513   related Shigella boydii strains. J Bacteriol. 2005;187(2):619-28.
514   5.   Nimri LF. Escherichia albertii, a newly emerging enteric pathogen with poorly
515   defined properties. Diagn Microbiol Infect Dis. 2013;77(2):91-5.
516   6.   Zhang X, Das S, Dunbar S, Tang YW. Molecular and non-molecular approaches to
517   etiologic diagnosis of gastroenteritis. Adv Clin Chem. 2020;99:49-85.

518 7.   Boxall MD, Day MR, Greig DR, Jenkins C. Antimicrobial resistance profiles of
519 diarrhoeagenic Escherichia coli isolated from travellers returning to the UK, 2015-2017. J
520 Med Microbiol. 2020;69(7):932-43.
521 8.   Luo L, Wang H, Payne MJ, Liang C, Bai L, Zheng H, et al. Comparative genomics of
522 Chinese and international isolates of Escherichia albertii: population structure and evolution
523 of virulence and antimicrobial resistance. Microb Genom. 2021;7(12).
524 9.   Hassan J, Awasthi SP, Hatanaka N, Okuno K, Hoang PH, Nagita A, et al.
525 Development of a multiplex PCR targeting eae, stx and cdt genes in genus Escherichia and
526 detection of a novel cdtB gene in Providencia rustigianii. Pathog Dis. 2018;76(9).
527 10.   Hinenoya A, Ichimura H, Awasthi SP, Yasuda N, Yatsuyanagi J, Yamasaki S.
528 Phenotypic and molecular characterization of Escherichia albertii: Further surrogates to avoid
529 potential laboratory misidentification. Int J Med Microbiol. 2019;309(2):108-15.
530 11.   Slater SL, Sagfors AM, Pollard DJ, Ruano-Gallego D, Frankel G. The Type III
531 Secretion System of Pathogenic Escherichia coli. Curr Top Microbiol Immunol.
532 2018;416:51-72.
533 12.   Hinenoya A, Yasuda N, Mukaizawa N, Sheikh S, Niwa Y, Awasthi SP, et al.
534 Association of cytolethal distending toxin-II gene-positive Escherichia coli with Escherichia
535 albertii, an emerging enteropathogen. Int J Med Microbiol. 2017;307(8):564-71.
536 13.   Gomes TAT, Ooka T, Hernandes RT, Yamamoto D, Hayashi T. Escherichia albertii
537 Pathogenesis. EcoSal Plus. 2020;9(1).
538 14.   Bhatt S, Egan M, Critelli B, Kouse A, Kalman D, Upreti C. The Evasive Enemy:
539 Insights into the Virulence and Epidemiology of the Emerging Attaching and Effacing
540 Pathogen Escherichia albertii. Infect Immun. 2019;87(1).
541 15.   Ooka T, Tokuoka E, Furukawa M, Nagamura T, Ogura Y, Arisawa K, et al. Human
542 gastroenteritis outbreak associated with Escherichia albertii, Japan. Emerging infectious
543 diseases. 2013;19(1):144.
544 16.   Masuda K, Ooka T, Akita H, Hiratsuka T, Takao S, Fukada M, et al. Epidemiological
545 Aspects of Escherichia albertii Outbreaks in Japan and Genetic Characteristics of the
546 Causative Pathogen. Foodborne Pathog Dis. 2020;17(2):144-50.
547 17.   Vishram B, Jenkins C, Greig DR, Godbole G, Carroll K, Balasegaram S, et al. The
548 emerging importance of Shiga toxin-producing Escherichia coli other than serogroup O157 in
549 England. J Med Microbiol. 2021;70(7).
550 18.   Foster G, Ross HM, Pennycott TW, Hopkins GF, McLaren IM. Isolation of
551 Escherichia coli O86:K61 producing cyto-lethal distending toxin from wild birds of the finch
552 family. Lett Appl Microbiol. 1998;26(6):395-8.
553 19.   Pennycott TW, Ross HM, McLaren IM, Park A, Hopkins GF, Foster G. Causes of
554 death of wild birds of the family Fringillidae in Britain. Vet Rec. 1998;143(6):155-8.
555 20.   Murakami K, Maeda-Mitani E, Kimura H, Honda M, Ikeda T, Sugitani W, et al. Non-
556 biogroup 1 or 2 Strains of the Emerging Zoonotic Pathogen Escherichia albertii, Their
557 Proposed Assignment to Biogroup 3, and Their Commonly Detected Characteristics. Front
558 Microbiol. 2019;10:1543.
559 21.   Hinenoya A, Awasthi SP, Yasuda N, Nagano K, Hassan J, Takehira K, et al.
560 Detection, Isolation, and Molecular Characterization of Escherichia albertii from Wild Birds
561 in West Japan. Jpn J Infect Dis. 2022;75(2):156-63.
562 22.   Oh JY, Kang MS, Hwang HT, An BK, Kwon JH, Kwon YK. Epidemiological
563 investigation of eaeA-positive Escherichia coli and Escherichia albertii strains isolated from
564 healthy wild birds. J Microbiol. 2011;49(5):747-52.
565 23.   Barmettler K, Biggel M, Treier A, Muchaamba F, Vogler BR, Stephan R. Occurrence
566 and Characteristics of Escherichia albertii in Wild Birds and Poultry Flocks in Switzerland.
567 Microorganisms. 2022;10(11).

568   24.   Hinenoya A, Li XP, Zeng X, Sahin O, Moxley RA, Logue CM, et al. Isolation and
569   characterization of Escherichia albertii in poultry at the pre-harvest level. Zoonoses Public
570   Health. 2021;68(3):213-25.
571   25.   Sonnevend A, Alali WQ, Mahmoud SA, Ghazawi A, Bharathan G, Melegh S, et al.
572   Molecular Characterization of MCR-1 Producing Enterobacterales Isolated in Poultry Farms
573   in the United Arab Emirates. Antibiotics (Basel). 2022;11(3).
574   26.   Wang H, Zhang L, Cao L, Zeng X, Gillespie B, Lin J. Isolation and characterization
575   of Escherichia albertii originated from the broiler farms in Mississippi and Alabama. Vet
576   Microbiol. 2022;267:109379.
577   27.   Grillova L, Sedlacek I, Pachnikova G, Stankova E, Svec P, Holochova P, et al.
578   Characterization of four Escherichia albertii isolates collected from animals living in
579   Antarctica and Patagonia. J Vet Med Sci. 2018;80(1):138-46.
580   28.   Hinenoya A, Nagano K, Awasthi SP, Hatanaka N, Yamasaki S. Prevalence of
581   Escherichia albertii in Raccoons (Procyon lotor), Japan. Emerg Infect Dis. 2020;26(6):1304-
582   7.
583   29.   Chattaway MA, Schaefer U, Tewolde R, Dallman TJ, Jenkins C. Identification of
584   Escherichia coli and Shigella Species from Whole-Genome Sequences. J Clin Microbiol.
585   2017;55(2):616-23.
586   30.   Lawson B, Howard T, Kirkwood JK, Macgregor SK, Perkins M, Robinson RA, et al.
587   Epidemiology of salmonellosis in garden birds in England and Wales, 1993 to 2003.
588   Ecohealth. 2010;7(3):294-306.
589   31.   Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina
590   sequence data. Bioinformatics. 2014;30(15):2114-20.
591   32.   Ewels P, Magnusson M, Lundin S, Kaller M. MultiQC: summarize analysis results
592   for multiple tools and samples in a single report. Bioinformatics. 2016;32(19):3047-8.
593   33.   Romao FT, Hernandes RT, Ooka T, Hayashi T, Sperandio V, Gomes TAT. Complete
594   Genome Sequence of Escherichia albertii Strain 1551-2, a Potential Extracellular and
595   Intracellular Pathogen. Genome Announc. 2018;6(9).
596   34.   Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-
597   MEM. arXiv preprint arXiv:13033997. 2013.
598   35.   Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence
599   Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9.
600   36.   Quinlan AR. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Curr
601   Protoc Bioinformatics. 2014;47:11 2 1-34.
602   37.   Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid
603   phylogenetic analysis of large samples of recombinant bacterial whole genome sequences
604   using Gubbins. Nucleic Acids Res. 2015;43(3):e15.
605   38.   Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, et al. SNP-sites:
606   rapid efficient extraction of SNPs from multi-FASTA alignments. Microb Genom.
607   2016;2(4):e000056.
608   39.   Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective
609   stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol.
610   2015;32(1):268-74.
611   40.   Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new
612   developments. Nucleic Acids Res. 2019;47(W1):W256-W9.
613   41.   Tonkin-Hill G, Lees JA, Bentley SD, Frost SDW, Corander J. RhierBAPS: An R
614   implementation of the population clustering algorithm hierBAPS. Wellcome Open Res.
615   2018;3:93.

616 42. Zhou Z, Alikhan NF, Mohamed K, Fan Y, Agama Study G, Achtman M. The
617 EnteroBase user's guide, with case studies on Salmonella transmissions, Yersinia pestis
618 phylogeny, and Escherichia core genomic diversity. Genome Res. 2020;30(1):138-52.
619 43. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome
620 assemblies from short and long sequencing reads. PLoS Comput Biol. 2017;13(6):e1005595.
621 44. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for
622 genome assemblies. Bioinformatics. 2013;29(8):1072-5.
623 45. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics.
624 2014;30(14):2068-9.
625 46. Feldgarden M, Brover V, Haft DH, Prasad AB, Slotta DJ, Tolstoy I, et al. Validating
626 the AMRFinder Tool and Resistance Gene Database by Using Antimicrobial Resistance
627 Genotype-Phenotype Correlations in a Collection of Isolates. Antimicrob Agents Chemother.
628 2019;63(11).
629 47. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of
630 intersecting sets and their properties. Bioinformatics. 2017;33(18):2938-40.
631 48. Takahata S, Ida T, Hiraishi T, Sakakibara S, Maebashi K, Terada S, et al. Molecular
632 mechanisms of fosfomycin resistance in clinical isolates of Escherichia coli. Int J Antimicrob
633 Agents. 2010;35(4):333-7.
634 49. LV DEK, Pires SM, Hald T. Attributing foodborne salmonellosis in humans to animal
635 reservoirs in the European Union using a multi-country stochastic model. Epidemiol Infect.
636 2015;143(6):1175-86.
637 50. Kaakoush NO, Castano-Rodriguez N, Mitchell HM, Man SM. Global Epidemiology
638 of Campylobacter Infection. Clin Microbiol Rev. 2015;28(3):687-720.
639 51. Travel-associated Shigella spp. infection in England, Wales and Northern Ireland:
640 2014. UK Health Security Agency; 2017.
641 52. Travel-associated non typhoidal Salmonella infection in England, Wales and Northern
642 Ireland: 2014. UK Health Security Agency; 2017.
643 53. Wang F, Zhang W, Niu D. Editorial: Foodborne Enterobacteriaceae of Animal Origin.
644 Front Cell Infect Microbiol. 2021;11:772359.
645 54. Sridhar S, Turbett SE, Harris JB, LaRocque RC. Antimicrobial-resistant bacteria in
646 international travelers. Curr Opin Infect Dis. 2021;34(5):423-31.
647 55. Kantele A, Kuenzli E, Dunn SJ, Dance DAB, Newton PN, Davong V, et al. Dynamics
648 of intestinal multidrug-resistant bacteria colonisation contracted by visitors to a high-endemic
649 setting: a prospective, daily, real-time sampling study. Lancet Microbe. 2021;2(4):e151-e8.
650 56. Davies ZG, Fuller RA, Loram A, Irvine KN, Sims V, Gaston KJ. A national scale
651 inventory of resource provision for biodiversity within domestic gardens. Biological
652 Conservation. 2009;142(4):761-71.
653 57. Rehn M, Ringberg H, Runehagen A, Herrmann B, Olsen B, Petersson AC, et al.
654 Unusual increase of psittacosis in southern Sweden linked to wild bird exposure, January to
655 April 2013. Euro Surveill. 2013;18(19):20478.
656 58. Lawson B, de Pinna E, Horton RA, Macgregor SK, John SK, Chantrey J, et al.
657 Epidemiological evidence that garden birds are a source of human salmonellosis in England
658 and Wales. PLoS One. 2014;9(2):e88968.
659 59. Lawson B, Franklinos LHV, Rodriguez-Ramos Fernandez J, Wend-Hansen C, Nair S,
660 Macgregor SK, et al. Salmonella Enteritidis ST183: emerging and endemic biotypes affecting
661 western European hedgehogs (Erinaceus europaeus) and people in Great Britain. Sci Rep.
662 2018;8(1):2449.
663 60. Murakami K, Etoh Y, Tanaka E, Ichihara S, Horikawa K, Kawano K, et al. Shiga
664 toxin 2f-producing Escherichia albertii from a symptomatic human. Jpn J Infect Dis.
665 2014;67(3):204-8.

666   61.   Lawson B, Robinson RA, Toms MP, Risely K, MacDonald S, Cunningham AA.
667   Health hazards to wild birds and risk factors associated with anthropogenic food
668   provisioning. Philos Trans R Soc Lond B Biol Sci. 2018;373(1745).
669

670

699
700
701   **Author contributions statement**

19

702 According to the Contributor roles taxonomy, author contributions were as follows:
703 Conceptualisation – KSB, BL, CJ, Data curation – BL, SKJ, SKM, SS, DG, CC, MDS, Formal analysis –
704 KSB, RJB, CC, MDS , Funding acquisition – KSB, AAC, BL, CJ, Investigation – KSB, AAC, BL, KS-M, SS, CJ,
705 RJB, SKJ, SKM, MDS, CC, Methodology – KSB, AAC, BL, KS-M, SKJ, SKM, SS, DG, RJB, Project
706 administration – KSB, CJ, BL, Resources – AAC, BL, CJ, DG, Software – RJB, Supervision – KSB, CJ,
707 Validation – DG, RJB, Visualisation – RJB, Writing original draft – CJ, RJB, BL, KSB Writing review and
708 editing – All.
709
710 **Competing interests statement**
711 The authors declare no competing interests.
712

Tables

Table 1. Phylogenetic and epidemiological features of *Escherichia albertii* Bayesian Analysis of Population Structure (BAPs) clusters for isolates from Great

Britain

	Genomic features		Isolate composition				Statistical support and nomenclature	
BAPS cluster	Congruence with phylogeny	Average Pairwise distance	Total isolates (n)	Human (% of cluster)	Wild bird (% of cluster)	Captive bird (% of cluster)	Proportion humans [Proportion (95% Confidence Interval), two tailed p-value]	Final determination^
1	Monophyletic	79	4	4 (100)	0 (0)	0 (0)	1.00 (0.40 - 1.00), p=0.1232	HAC
2	Monophyletic	28	17	17 (100)	0 (0)	0 (0)	1.00 (0.80 - 1.00), p<0.0001	HAC
3	Monophyletic	132	7	7 (100)	0 (0)	0 (0)	1.00 (0.59 - 1.00), p=0.0156	HAC
4	Monophyletic	25	16	16 (100)	0 (0)	0 (0)	1.00 (0.79 - 1.00), p<0.0001	HAC
5	Monophyletic	167	4	4 (100)	0 (0)	0 (0)	1.00 (0.40 - 1.00), p=0.1232	HAC
6	Monophyletic	967	61	12 (20)	49 (80)	0 (0)	0.20 (0.11 - 0.33), p<0.0001	BAC

7	Monophyletic	73	29	4 (14)	24 (86)	1 (3)	0.17 (0.05 - 0.35), p=0.0003	BAC
8	Polyphyletic	3169	24	19 (79)	1 (4)	4 (17)	0.80 (0.58 - 0.93), p=0.0148	HAC
Total	NA	NA	162	85	74	5		

717   ^BAC = Bird associated cluster, HAC = human associated cluster

718

719

Table 2. Antimicrobial resistance phenotypes of 11 *Escherichia albertii* isolates with varied antimicrobial resistance genotypes

Sequence Read Archive Accession	Genotype [a]	Minimum. Inhibitory Concentration (µg/mL)							
		Cephalosporin		Fosfomycin	Tetracycline	β-lactam	Fluoroquinolone	Chloramphenicol	Rifampicin
		Cefoxitin	Ceftriaxone	Fosfomycin	Tetracycline	Ampicillin	Ciprofloxacin	Chloramphenicol	Rifampicin
SRR12769799	*uhpT_E350Q*, marR_S3N*, parE_I355T**	8	0.047	1	<u>1.5</u>	3	0.016	6	8
SRR12769953	*uhpT_E350Q*, marR_S3N*, parE_I355T**	6	0.047	6	0.5	4	0.008	3	2
SRR13049225	*uhpT_E350Q*, marR_S3N*, parE_I355T**	6	0.047	2	<u>1</u>	4	0.006	4	4
SRR13049237	*uhpT_E350Q*, parE_I355T**	6	0.047	1	<u>1</u>	4	0.012	6	3
SRR11442290	*uhpT_E350Q*, marR_S3N*, parE_I355T*, gyrA_S83L**	4	0.047	1.5	<u>0.75</u>	6	0.125	3	12
SRR15338008	*uhpT_E350Q*, marR_S3N*, gyrA_S83L*, parE_I355T**	1.5	<0.016	1.5	0.38	3	0.032	2	4

SRR8981835	uhpT_E350Q*, marR_S3N*, gyrA_S83L*, parE_I355T*	1.5	<0.016	1.5	0.25	3	0.032	2	4
SRR15338057	blaDHA-1, uhpA_G97D*, uhpT_E350Q*, tet(A), marR_S3N*, blaTEM-1, gyrA_S83L*, parE_I355T*, qnrB4	6	0.064	<u>12</u>	<u>48</u>	<u>>256</u>	0.38	1.5	4
SRR9050433	blaDHA-1, uhpA_G97D*, uhpT_E350Q*, tet(A), marR_S3N*, blaTEM-1, gyrA_S83L*, parE_I355T*, qnrB4	1	0.047	<u>8</u>	<u>48</u>	<u>96</u>	<u>0.5</u>	3	16
SRR11425059	uhpT_E350Q*, tet(A), marR_S3N*, blaTEM-1, gyrA_S83L*, parE_I355T*	12	0.064	2	<u>48</u>	<u>>256</u>	0.19	4	8

24

	Genotype[a]								
SRR3574322	uhpT_E350Q*, tet(A), marR_S3N*, blaTEM-1, gyrA_S83L*, parE_I355T*	3	<0.016	1	32	96	0.023	4	4
Genotype associated with resistance [b]		blaDHA-1		uhpA_G97D*, uhpT_E350Q*	tet(A), tet(B), marR_S3N*	blaTEM-1, blaTEM-135, marR_S3N*	gyrA_S83L*, parC_S57T*, parE_I355T*, qnrB19, qnrB4, qnrS13, marR_S3N*	marR_S3N*	marR_S3N*

[a] Genes and point mutations (*) found present in isolates, in this study
[b] Genes and point mutations (*) associated with resistance to antimicrobial classes tested in this study
[c] Underlined MIC determination results highlight MIC breakpoints (⌈gmL^{-1}) classed as resistant according to EUCAST guidelines
(https://www.eucast.org/fileadmin/src/media/PDFs/EUCAST_files/Breakpoint_tables/v_11.0_Breakpoint_Table

25

726 **Figure 1.** Midpoint rooted maximum likelihood phylogenetic tree of *Escherichia albertii* isolates from
727 Great Britain showing human demographic features and antimicrobial resistance genes (ARGs). The
728 scale bar is shown in single nucleotide polymorphisms (SNPs). Isolate metadata are displayed in the
729 adjacent tracks on the right according to the inlaid keys on the left (BAPS = Bayesian Analysis of
730 Population Structure). Tracks in the centre panel shows presence of ARGs grouped by antimicrobial
731 class, with the *gryA* S83L point mutation highlighted in bold and indicated with an asterisk.
732 Phylogenetic branches highlighted in red indicate nodes with low bootstrap support between 50 and
733 70%.
734
735 **Figure 2.** Midpoint rooted maximum likelihood phylogenetic tree of *Escherichia albertii* isolates from
736 Great Britain showing bird host characteristics and virulence-associated genes. The scale bar is shown
737 in single nucleotide polymorphisms (SNPs). Isolate metadata are displayed on the adjacent tracks
738 according to the inlaid key. Phylogenetic branches highlighted in red indicate nodes with low
739 bootstrap support between 50 and 70%. BAPS = Bayesian Analysis of Population Structure.
740
741 **Figure 3.** Occurrence of antimicrobial resistance (AMR) among *Escherichia albertii* isolates from Great
742 Britain. (**A**) Stacked barplot demonstrates the number of isolates from birds and humans carrying
743 known AMR genetic determinants. Genetic determinants highlighted with asterisks represent point
744 mutations and different antimicrobial drug classes shown in alternating coloured text. (**B**) UpSet plot
745 illustrates the prevalence of AMR genotypic profile among human isolates. The combination matrix in
746 the centre panel shows the various genotypic AMR profiles, in which each column represents a unique
747 profile, and each black dot represents presence of a genetic determinant conferring
748 resistance/reduced susceptibility to a drug class (displayed on the left). Vertical barplot above the
749 matrix shows the number of isolates with a particular genotype, the number above each bar shows
750 the exact number of isolates with the genotype.

751 **Figure 4.** Phylogenetic tree comprising of 162 *Escherichia albertii* isolates from the current study and
752 an additional 475 isolates retrieved from EnteroBase. Tree was constructed based on MLST sequences.
753 Circles at tree tips highlight *E. albertii* isolates from Great Britain under investigation in this study, and
754 the colour of the circles represents the BAPS clusters identified earlier in the study. The thicker inner
755 ring, demonstrates the source of the isolates and the thinner outer ring demonstrates the isolate
756 country of origin, all of which are labelled according to the inlaid keys displayed on the left.

757

ARGs

Legend:

BAPS cluster
1, 2, 3, 4, 5, 6, 7, 8

Host
Avian, Human

Age
Infant (<2), Children (2-15), Adult (16-60), Elderly (>60)

Travel
Asia, Americas (Mexico), Europe, Africa

Captive/Wild
Captive, Wild

Column labels: BAPS cluster, Host, Age, Travel destination, Captive / Wild, aadA1, aadA2, aadA12, strA, strB, blaDHA-1, blaTEM-1, blaTEM-135, dfrA5, dfrA7, dfrA12, dfrA14, dfrA15, dfrA17, Inc(F), Inc(G), mph(A), qnrB4, qnrB19, qnrS13, gyrA S83L*, sul1, sul2, sul3, tet(A), tet(B)

266 SNPs

A

AMR genetic determinants

B

Human *E. albertii* AMR genotype profiles

BAPS cluster
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8

Source niche
- Human
- Avian (non-poultry)
- Avian (poultry)
- Avian (not defined)
- Mammal (livestock)
- Mammal (wild species)
- Mammal (companion species)
- Food
- Water
- Unknown

Country
- United States
- Canada
- Mexico
- Gambia
- Guinea-Bissau
- United Kingdom
- Netherlands
- Denmark
- Czech Republic
- France
- Luxembourg
- Germany
- China
- Japan
- South Korea
- India
- Bangladesh
- Australia

500 alleles