1 **Title**

2 Extraintestinal pathogenic *Escherichia coli* (ExPEC) are associated with prolonged carriage of

3 extended-spectrum β-lactamase-producing *E. coli* acquired during travel

4

5 **Author list**

6 Boas C.L. van der Putten[1,2]*, Jarne M. van Hattem[1], John Penders[3,4], COMBAT Consortium[#],

7 Daniel R. Mende[1], Constance Schultsz[1,2]

8 [1]Department of Medical Microbiology, Amsterdam UMC, University of Amsterdam, the

9 Netherlands

10 [2]Department of Global Health, Amsterdam Institute for Global Health and Development,

11 Amsterdam UMC, University of Amsterdam, the Netherlands

12 [3]School for Public Health and Primary Care (Caphri), Department of Medical Microbiology,

13 Maastricht University Medical Centre, Maastricht, Netherlands

14 [4]School for Nutrition and Translational Research in Metabolism (NUTRIM), Maastricht University

15 Medical Centre, Maastricht, Netherlands

16 *Corresponding author. Email: boas.vanderputten@amsterdamumc.nl

17 Telephone: +31205664868.

18 [#]Members are listed in the appendixes

19 Article type: Original article

Abstract word count: 246

Main text word count: 2521

Abstract

Objectives Extended-spectrum β-lactamase-producing *Escherichia coli* (ESBL-Ec) are frequently acquired during international travel, contributing to the global spread of antimicrobial resistance. Human-adapted ESBL-Ec are predicted to exhibit increased intestinal carriage duration, resulting in a higher likelihood of onward human-to-human transmission. Yet, bacterial determinants of increased carriage duration are unknown. Previous studies analysed small traveler cohorts, with short follow-up times, or did not employ high-resolution molecular typing, and were thus unable to identify bacterial traits associated with long-term carriage.

Methods In a prospective cohort study of 2001 international travelers, we analysed 160 faecal ESBL-Ec isolates from all 38 travelers who acquired ESBL-Ec during travel and subsequently carried ESBL-Ec for at least 12 months after return, by whole-genome sequencing. For 17 travelers, we confirmed the persistence of ESBL-Ec strains through single nucleotide variant typing. To identify determinants of increased carriage duration, we compared the 17 long-term carriers (≥12 months carriage) with 33 age-, sex- and destination-matched short-term carriers (<1 month carriage). Long-read sequencing was employed to investigate ESBL plasmid persistence.

Results We show that in healthy travelers with very low antibiotic usage, extraintestinal pathogenic lineages of *E. coli* (ExPEC) are significantly more likely to persist than other *E. coli*

lineages. The long-term carriage of *E. coli* from ExPEC lineages is mainly driven by sequence type 131 and phylogroup D *E. coli*.

Conclusions Although ExPEC frequently cause extra-intestinal infections such as bloodstream infections, our results imply that ExPEC are also efficient intestinal colonizers, which potentially contributes to their onward transmission.

Introduction

International travel contributes significantly to the spread of extended-spectrum β-lactamase (ESBL) gene positive *Escherichia coli* (ESBL-Ec).[1,2] Genetically diverse ESBL-Ec are frequently acquired during international travel.[3–5] Whilst travel-acquired ESBL-Ec typically are lost during travel or within the first month after return (Figure 1A),[6] ESBL-Ec and ESBL genes have been detected for more than 12 months after return in a proportion of travelers.[1,3–5,7] ESBL genes can persist through at least two different mechanisms.[8] Bacterial strains that carry ESBL-encoding genes in their chromosome or on a stable plasmid can persist over time as part of the local microbiome (strain persistence, Figure 1B). In addition, ESBL genes can be located on mobile genetic elements (MGEs) which persist in the microbiome by their ability to transfer between different bacterial hosts (MGE persistence, Figure 1C).

ESBL-Ec lineages which are capable of long-term colonization and are adapted to the human intestinal tract are likely to contribute to onward transmission of ESBL-Ec and ESBL genes. However, it is unknown which ESBL-Ec lineages are capable of persistence after return from travel, due to a lack of studies with a sufficiently large sample size and a prospective longitudinal study design. Additionally, high-resolution typing methods such as whole-genome sequencing (WGS) are needed to reliably determine whether the strains have been carried over

a long period. One study employed WGS to investigate persistence in 16 travelers who acquired ESBL-Ec abroad and showed that only one traveler carried a travel-acquired ESBL-Ec strain for at least 7 months.[7] A very recent study analysed data from 11 travellers which were colonised for >3 months.[9] Due to the low sample sizes in both studies, few ESBL-Ec attributes could be identified that were significantly associated with long-term carriage. Armand-Lefèvre et al. reported an association between phylogroups B2/D/F with prolonged carriage duration (p = 0.02).[9] Other studies either did not employ WGS or focused on short-term carriage only.

We studied a cohort of 2001 Dutch international travelers (COMBAT cohort).[1] Six hundred and thirty-three travelers acquired ESBL gene-positive Enterobacterales during travel abroad, of whom 38 travelers (6.0%) were colonized for ≥12 months after acquisition of ESBL gene-positive Enterobacterales, all of which were *E. coli*. Here we report on host and bacterial characteristics associated with ESBL genes persistence and the mechanism through which the ESBL genes persisted. We identified an association between extraintestinal pathogenic *E. coli* and long-term carriage.

Materials and methods

Travelers and isolates selection

We included all 38 travelers from the COMBAT cohort (N=2001) who were colonized for ≥12 months with *Escherichia coli* which possessed ESBL genes belonging to a single ESBL gene group at return from travel and at all subsequent time points (1, 3, 6, and 12 months after return), and who tested ESBL-negative before travel (see reference 1 for a detailed description of sampling

83 and microbiological methods).[1] In short, faecal samples were inoculated in tryptic soy broth
84 containing 50 mg/L vancomycin to select for Enterobacterales. After overnight incubation, the
85 broth was subcultured on chromID ESBL agar plates (bioMérieux). Morphologically different
86 colonies were isolated, with a maximum of five isolates per faecal sample. Eighty-five ESBL-Ec
87 isolates sampled at return from travel (T0) and 75 ESBL-Ec isolates sampled 12 months after
88 return from travel (T12) were included in the current study (Figure 2).

89 ***In silico* typing**

90 DNA extraction and library preparation were performed on all available isolates using the
91 Qiagen Blood and Tissue DNA extraction kit (Cat No./ID: 69506) and Kapa HTP library prep kit
92 (Kit Code KK8234), respectively. Whole-genome sequencing was performed on Illumina HiSeq
93 2500 by the Amsterdam UMC Core Facility Genomics.

94 Sequencing data were analysed using a Snakemake v5.7.1[10] pipeline, available at https://
95 github.com/boasvdp/COMBAT (v1.1.0 archived at https://doi.org/10.5281/zenodo.4582689). In
96 short, Illumina sequencing data were trimmed using fastp v0.20.0,[11] assembled using the Shovill
97 wrapper v1.0.9 (https://github.com/tseemann/shovill) for SPAdes,[12] and resistance genes were
98 identified using AMRfinderplus v3.2.3.[13] *E. coli* phylogroups were predicted using EzClermont
99 v0.4.3.[14] Multi-locus sequence typing (MLST) was performed with the mlst script
100 (https://github.com/tseemann/mlst), using the Achtman scheme for *Escherichia coli*.[15]

101 **Analysis of strain persistence**

102 We defined strain persistence as the presence of one or more ESBL-Ec isolates in faecal samples
103 from a single traveler obtained at return from travel (T0), as well as 12 months thereafter (T12),
104 with a maximum of 25 SNPs per 5 Mbp difference in the core genome.[16]

105 To assess strain persistence, we calculated single nucleotide polymorphism differences between
106 isolates to identify which isolates belonged to the same strain. We mapped sequencing reads
107 using Snippy v4.4.5 (https://github.com/tseemann/snippy) on MLST-specific reference genomes
108 selected with ReferenceSeeker v1.6.3[17] to obtain core genome alignments. Reference genomes
109 used for all MLST sequence types, together with core genome alignment lengths, can be found
110 in table S1. For each MLST, IQtree v1.6.12[18] was used to infer a phylogeny from the core
111 genome alignment under the model advised by Modelfinder.[19] Recombination events in the
112 core genome alignment were identified using ClonalFrameML v1.12[20] and masked using maskrc-
113 svg v0.5 (https://github.com/kwongj/maskrc-svg). SNP differences were counted using snp-dists
114 v0.7.0 (https://github.com/tseemann/snp-dists) and alignment lengths were calculated using a
115 modified version of snp-dists v0.7.0 (https://github.com/boasvdp/snp-dists). SNP counts were
116 scaled to 5 Mbp, to approximate the number of genome-wide SNPs.[16]

117 To determine whether ESBL plasmids had persisted independent of bacterial host, we employed
118 long read sequencing (Oxford Nanopore Technologies). We generated Oxford Nanopore
119 Technologies sequencing data according to Van der Putten et al. (2020).[21] In short, strains were
120 grown overnight at 37 ˚C in liquid LB. DNA extraction and library preparation were performed
121 using the Qiagen MagAttract HMW DNA Kit (Cat. No. 67563) and ONT native barcoding kit (Cat.
122 No. EXP-NBD114), respectively. The library was subsequently sequenced on an ONT MinION
123 flowcell. Raw read data was filtered using Filtlong v0.2.0 (https://github.com/rrwick/Filtlong)

124 and assembled with corresponding Illumina data using Unicycler v0.4.8.[22] Quality control was
125 implemented at several steps in the pipeline using FastQC v0.11.8,[23] Quast v4.6.3,[24] and
126 MultiQC v1.6.[25] Plasmid comparison was performed using ANICalculator.[26]

127 **Comparison with short-term carriers**

128 Seventeen long-term carriers were matched by age (range +/- 7 years), sex, and travel
129 destination (United Nations subregions) to thirty-three travelers who were colonized for less
130 than a month using SPSS 26 (Figure 2). Illumina WGS was performed as described before on all
131 ESBL-Ec isolated at return from travel from these short-term carriers (total 41 isolates).

132 **Plotting and statistical analysis**

133 Data were plotted using ggplot2 v3.1.1,[27] ggthemes v2.4.0,[28] and patchwork[29] in R v3.5.1.
134 Tabular data were analysed using Pandas v0.24.2[30] in Python v3.6.7. Statistical analysis was
135 performed using Fisher's exact test as implemented in the Python library SciPy.[31]

136

137 # Results

138 **Data characteristics**

139 Out of 2001 Dutch international travelers, we included 34 travelers with samples available, out
140 of all 38 travelers whose faecal samples were positive for *E. coli* harbouring the same ESBL
141 group gene at return (T0), one month after return from travel (T1) and 12 months after return
142 from travel (T12), but were ESBL-negative before travel (Figure 2).[1] We included a median of 1
143 ESBL-Ec isolate per traveler per timepoint (range: 1-5 isolates). SNP typing showed that acquired

144 ESBL-Ec isolates were genetically diverse, with some travelers acquiring up to four distinct
145 strains.

146 **ESBL-Ec strain persistence**

147 *SNP analysis*

148 Twenty-two out of 34 travelers harboured persistent strains based on SNP typing (Figure 2). Of
149 the 12 travelers who did not harbour persistent strains, ten carried strains with differing MLST
150 profiles between T0 and T12, and two carried strains from a single ST but with the number of
151 core genome-wide SNP differences exceeding the threshold (25 SNPs for a 5 Mbp genome).[16]

152 *Clonal ST38 lineages*

153 Non-persistent, yet highly conserved strains that are widely disseminated within the human
154 population could potentially be misidentified as persistent when using SNV analyses. Hence, we
155 compared SNV distances between isolates obtained from unrelated travelers (i.e., belonging to
156 different households) to identify strains shared across our study population. Conserved strains
157 belonging to two ST38 lineages were identified in five unrelated travelers. Lineage ST38-$bla_{CTX\text{-}M\text{-}}$
158 $_{27}$ was identified in three unrelated travelers and lineage ST38-$bla_{CTX\text{-}M\text{-}14}$ in two other, unrelated
159 travelers. All five unrelated travelers returned from different countries and continents,
160 suggesting these two lineages have disseminated widely. Whole genomes from lineages ST38-
161 $bla_{CTX\text{-}M\text{-}27}$ and ST38-$bla_{CTX\text{-}M\text{-}14}$ were also abundantly present in public data and a phylogenetic
162 analysis of their core genomes confirmed their high similarity (Figure S1). Hence, we could not
163 conclusively determine whether these highly clonal strains had persisted in the five travelers
164 concerned, or whether these strains were re-acquired from an unknown source. Therefore, we

165 excluded the travelers harbouring these clonal ST38 strains from further analyses to identify
166 lineages potentially associated with persistence. ST38 strains colonizing four additional travelers
167 were not related to the ST38-$bla_{CTX\text{-}M\text{-}27}$ and ST38-$bla_{CTX\text{-}M\text{-}14}$ lineages and were thus not excluded
168 from further analyses.

169 We identified two related travelers who harboured highly similar ST69 isolates (9 genome-wide
170 SNPs). The most likely explanation of this observation is that the travellers acquired this strain
171 from the same source or one from the other, and that it has persisted in both travellers since.

172 *Plasmid analysis*

173 For six travelers, we detected identical ESBL genes from isolates sampled at return from travel
174 and twelve months thereafter, but we could exclude ESBL gene persistence in persistent isolates
175 based on SNP typing (Table S1). To determine whether the plasmid carrying these ESBL genes
176 had persisted independent of bacterial host (Figure 1B), we employed long read sequencing. For
177 one traveler, we detected a pair of ESBL-Ec isolates belonging to different MLST types in which
178 an almost identical plasmid (99.8% nucleotide identity) was identified at T0 and T12, suggesting
179 ESBL gene persistence by plasmid transfer between different *E. coli* hosts.

180 In summary, we identified persistent ESBL-Ec strains in 17 out of 34 travelers with prolonged
181 ESBL-Ec carriage and persistent ESBL-plasmid in one traveler (Table S1). Although we included
182 multiple ESBL-Ec isolates for some travelers, we did not observe multiple persistent strains
183 within any single traveler.

184 **Comparison of ESBL-Ec from long-term and short-term carriers**

For each long-term carrier (≥12 months carriage) harbouring a persistent strain, two short-term carriers (<1 month carriage) were matched by age, sex and travel destination. For one long-term carrier, only a single matching short-term carrier could be identified, resulting in a comparison of 17 isolates from 17 long-term carriers and 42 isolates from 33 matched short-term carriers, which were sequenced.

Antibiotic usage was low before, during, and after travel and similar between long-term and short-term carriers (Table 1). None of the travelers were admitted to the hospital during or after travel in either group. One single traveler returned to the same country as visited during index travel, within 12 months after return from index travel.

Following the recent definition of common extraintestinal pathogenic *E. coli* (ExPEC) lineages according to Manges *et al*. (2019),[32] persistent strains belonged significantly more often to ExPEC lineages than non-persistent strains. In 15 out of 17 long-term carriers, ExPEC were identified as the persistent strain while in only 7 out of 33 short-term carriers ExPEC were detected (odds ratio 27.86, 95% confidence interval: 5.11-151.74). This difference appears to be driven mostly by ST131 and phylogroup D strains.

Discussion

We demonstrated long-term carriage of travel-acquired ESBL-positive *Escherichia coli* in 17 travelers out of a cohort of 2001 travelers, which was driven by persistence of ESBL-Ec belonging to ExPEC lineages. In a single traveler, the persistence of ESBL-Ec seemed to be due to an ESBL plasmid which shifted between bacterial hosts after colonising the traveler's gut.

206 Our study strengthens the finding of Armand-Lefèvre et al.[9] that ESBL-Ec lineage is associated
207 with persistent carriage after acquisition during travel. Interestingly, we come to similar
208 conclusions although we have focused on long-term carriage (≥12 months) as compared to the
209 previous study (>3 month carriage). Earlier work could not make strong assertions about the
210 persistence of strains, either due to the limited number of included travelers, because the
211 typing methods employed were insufficiently discriminating between isolates, because of a
212 limited duration of follow-up, or because the study population was not representative of
213 community exposure.[3–5,7] In a recent study in Laos, a group of travelers attending a course at
214 local hospitals acquired a very high diversity of ESBL-Ec *E. coli* immediately upon arrival and it
215 was shown that this acquisition of resistant *E. coli* during travel is highly dynamic.[6] However, the
216 short follow-up in this study did not allow analysis of long-term outcomes of the acquisition of
217 ESBL-Ec.

218 Previous studies have shown that a limited number of *E. coli* lineages contribute to a large
219 fraction of *E. coli*-mediated extraintestinal disease, including urinary tract infections and
220 bloodstream infections.[32] These extraintestinal pathogenic *E. coli* are commonly referred to as
221 ExPEC. While ExPEC display pathogenic potential, epidemiological studies often find ExPEC
222 colonizing the human gut in the absence of symptoms.[33,34] In fact, it has been proposed that
223 ExPEC have evolved towards particularly efficient intestinal colonization and their
224 extraintestinal virulence is an evolutionary "byproduct".[35] The long-term persistence of ExPEC
225 lineages, acquired after relatively short travel duration as observed in our study (median 20
226 days, interquartile range: 17 – 29 days) suggests that ExPEC lineages have spread globally
227 successfully due to their adaptation to the human intestinal tract. It should however be noted

228 that we restricted our analysis to ESBL-producing *E. coli* and duration of persistence of
229 susceptible ExPEC in this cohort is unknown.

230 Whole-genome sequencing allowed us to detect two clonal ST38 lineages which were shared
231 between unrelated travelers. By additional analysis of publicly available WGS data, these ST38
232 lineages were shown to have spread globally. The extremely high degree of similarity within
233 these ST38 lineages interfered with reliable strain identification based on core genome DNA
234 sequence analysis. Future studies should explore the application of SNP typing using the
235 pangenome rather than only the core genome, for example through software like Pandora[36].
236 These novel methods however currently display a higher error rate than the core genome
237 analysis approach used in the present study[36]. Awareness of the circulation of ST38 or other
238 rapidly expanding lineages that are indistinguishable by their core genome is important, for
239 example for management of suspected outbreaks of ST38 in hospital settings.

240 Our current study focused solely on ESBL-producing *E. coli* and found an association between
241 ExPEC carriage and increased carriage duration. Our approach could also be applied to non-
242 ESBL-producing ExPEC, to determine whether acquisition of ExPEC in general, independent of
243 their antibiotic susceptibility profile, is more likely to results in increased carriage duration after
244 travel. The frequency of sampling can be considered a limitation of this study. However, when
245 comparing between two timepoints, twelve months apart, we could still identify very closely
246 related isolates in samples from the same traveler, indicating strain persistence. Additionally,
247 we only had isolates available for household members of four long-term carriers. Including
248 more household members in future studies might allow us to estimate the likelihood of onward
249 transmission following long-term carriage.

Applying genomic epidemiology to a large traveler cohort, we have shown that ESBL-positive *E. coli* acquired during travel are able to persist for more than a year. The strains that showed the longest carriage duration belonged predominantly to pathogenic ExPEC lineages. Our data imply that long-term carriage of resistant *E. coli* is governed by bacterial characteristics which are associated with lineage. This finding possibly allows a more precise risk assessment for international travelers returning with travel-acquired resistant *E. coli*.

Acknowledgements

The authors would like to thank Arie van der Ende for the helpful discussions. We thank SURFsara (https://www.surfsara.nl) for the support in using the Lisa Compute Cluster.

Funding information

The COMBAT study was funded by Netherlands Organization for Health, Research and Development (ZonMw; 50-51700-98-120) and EU-H2020 programme (COMPARE, 643476). BP was funded through an internal grant of the Amsterdam UMC ("flexibele OiO beurs").

Declaration of interests

We declare no competing interests.

Data and code availability

All Illumina and Oxford Nanopore Technologies sequencing data used in this study are currently available free of restrictions at NCBI under project accession number PRJEB40103. Metadata linking isolates to travelers, required to reproduce our analyses, are currently available free of

restrictions at the GitHub repository of this project (https://www.github.com/boasvdp/COMBAT, v1.1.0 archived through Zenodo at https://doi.org/10.5281/zenodo.4582689).

All code is available free of restrictions under the MIT license at https://www.github.com/boasvdp/COMBAT (v1.1.0 archived through Zenodo at https://doi.org/10.5281/zenodo.4582689).

Ethical statement

The COMBAT study was approved by the Medical Research Ethics Committee, Maastricht University Medical Centre (METC 12-4-093). All participants provided written informed consent.

Appendices

Figure S1: Phylogeny of clonal complex 38 genomes with continent of origin.

Table S1: All SNP comparisons between isolates with relevant metadata (169 rows).

Supplemental information: Members of the COMBAT consortium.

References

1. Arcilla MS, van Hattem JM, Haverkate MR, et al. Import and spread of extended-spectrum β-lactamase-producing Enterobacteriaceae by international travellers (COMBAT study): a prospective, multicentre cohort study. *Lancet Infect Dis*. 2017;17(1):78-85. doi:10.1016/S1473-3099(16)30319-X

2. Ruppé E, Andremont A, Armand-Lefèvre L. Digestive tract colonization by multidrug-resistant Enterobacteriaceae in travellers: An update. *Travel Med Infect Dis*. 2018;21:28-35. doi:10.1016/j.tmaid.2017.11.007

3. Ruppé E, Armand-Lefèvre L, Estellat C, et al. High Rate of Acquisition but Short Duration of Carriage of Multidrug-Resistant Enterobacteriaceae After Travel to the Tropics. *Clin Infect Dis Off Publ Infect Dis Soc Am*. 2015;61(4):593-600. doi:10.1093/cid/civ333

296 4. Paltansing S, Vlot JA, Kraakman MEM, et al. Extended-spectrum β-lactamase-producing
297 enterobacteriaceae among travelers from the Netherlands. *Emerg Infect Dis*. 2013;19(8):1206-
298 1213. doi:10.3201/eid.1908.130257

299 5. Pires J, Kuenzli E, Kasraian S, et al. Polyclonal Intestinal Colonization with Extended-Spectrum
300 Cephalosporin-Resistant Enterobacteriaceae upon Traveling to India. *Front Microbiol*. 2016;7:1069.
301 doi:10.3389/fmicb.2016.01069

302 6. Kantele A, Kuenzli E, Dunn SJ, et al. Dynamics of intestinal multidrug-resistant bacteria colonisation
303 contracted by visitors to a high-endemic setting: a prospective, daily, real-time sampling study.
304 *Lancet Microbe*. 2021;0(0). doi:10.1016/S2666-5247(20)30224-X

305 7. Bevan ER, McNally A, Thomas CM, Piddock LJV, Hawkey PM. Acquisition and Loss of CTX-M-
306 Producing and Non-Producing Escherichia coli in the Fecal Microbiome of Travelers to South Asia.
307 *mBio*. 2018;9(6). doi:10.1128/mBio.02408-18

308 8. Sheppard AE, Stoesser N, Wilson DJ, et al. Nested Russian Doll-Like Genetic Mobility Drives Rapid
309 Dissemination of the Carbapenem Resistance Gene blaKPC. *Antimicrob Agents Chemother*.
310 2016;60(6):3767-3778. doi:10.1128/AAC.00464-16

311 9. Armand-Lefèvre L, Rondinaud E, Desvillechabrol D, et al. Dynamics of extended-spectrum beta-
312 lactamase-producing Enterobacterales colonization in long-term carriers following travel abroad.
313 *Microb Genomics*. 2021;7(7). doi:10.1099/mgen.0.000576

314 10. Köster J, Rahmann S. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics*.
315 2012;28(19):2520-2522. doi:10.1093/bioinformatics/bts480

316 11. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*.
317 2018;34(17):i884-i890. doi:10.1093/bioinformatics/bty560

318 12. Bankevich A, Nurk S, Antipov D, et al. SPAdes: A New Genome Assembly Algorithm and Its
319 Applications to Single-Cell Sequencing. *J Comput Biol*. 2012;19(5):455-477.
320 doi:10.1089/cmb.2012.0021

321 13. Feldgarden M, Brover V, Haft DH, et al. Validating the AMRFinder Tool and Resistance Gene
322 Database by Using Antimicrobial Resistance Genotype-Phenotype Correlations in a Collection of
323 Isolates. *Antimicrob Agents Chemother*. 2019;63(11). doi:10.1128/AAC.00483-19

324 14. Waters NR, Abram F, Brennan F, Holmes A, Pritchard L. Easy phylotyping of Escherichia coli via the
325 EzClermont web app and command-line tool. *Access Microbiol*. Published online 2020.
326 doi:10.1099/acmi.0.000143

327 15. Wirth T, Falush D, Lan R, et al. Sex and virulence in Escherichia coli: an evolutionary perspective.
328 *Mol Microbiol*. 2006;60(5):1136-1151. doi:10.1111/j.1365-2958.2006.05172.x

329 16. Gorrie CL, Silva AGD, Ingle DJ, et al. Systematic analysis of key parameters for genomics-based real-
330 time detection and tracking of multidrug-resistant bacteria. *bioRxiv*. Published online September
331 25, 2020:2020.09.24.310821. doi:10.1101/2020.09.24.310821

332 17. Schwengers O, Hain T, Chakraborty T, Goesmann A. ReferenceSeeker: rapid determination of
333 appropriate reference genomes. *J Open Source Softw*. 2020;5(46):1994. doi:10.21105/joss.01994

334 18. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective Stochastic
335 Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol Biol Evol*. 2015;32(1):268-274.
336 doi:10.1093/molbev/msu300

337 19. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model
338 selection for accurate phylogenetic estimates. *Nat Methods*. 2017;14(6):587-589.
339 doi:10.1038/nmeth.4285

340 20. Didelot X, Wilson DJ. ClonalFrameML: Efficient Inference of Recombination in Whole Bacterial
341 Genomes. *PLOS Comput Biol*. 2015;11(2):e1004041. doi:10.1371/journal.pcbi.1004041

342 21. van der Putten BCL van der, Roodsant TJ, Haagmans MA, Schultsz C, Ark KCH van der. Five
343 Complete Genome Sequences Spanning the Dutch Streptococcus suis Serotype 2 and Serotype 9
344 Populations. *Microbiol Resour Announc*. 2020;9(6). doi:10.1128/MRA.01439-19

345 22. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short
346 and long sequencing reads. *PLOS Comput Biol*. 2017;13(6):e1005595.
347 doi:10.1371/journal.pcbi.1005595

348 23. Andrews S, others. *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Babraham
349 Bioinformatics, Babraham Institute, Cambridge, United Kingdom; 2010.

350 24. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies.
351 *Bioinformatics*. 2013;29(8):1072-1075. doi:10.1093/bioinformatics/btt086

352 25. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools
353 and samples in a single report. *Bioinformatics*. 2016;32(19):3047-3048.
354 doi:10.1093/bioinformatics/btw354

355 26. Varghese NJ, Mukherjee S, Ivanova N, et al. Microbial species delineation using whole genome
356 sequences. *Nucleic Acids Res*. 2015;43(14):6761-6771. doi:10.1093/nar/gkv657

357 27. Wickham H. *Ggplot2: Elegant Graphics for Data Analysis*. Springer; 2016.

358 28. Arnold JB. ggthemes: Extra Themes, Scales and Geoms for "ggplot2." *R Package Version*. 2017;3(0).

359 29. Pedersen TL. patchwork: The Composer of ggplots. *R Package Version 00*. 2017;1.

360 30. team T pandas development. *Pandas-Dev/Pandas: Pandas*. Zenodo; 2020.
361 doi:10.5281/zenodo.3509134

362 31. Jones E, Oliphant T, Peterson P, others. *SciPy: Open Source Scientific Tools for Python*.; 2001.
363 http://www.scipy.org/

364 32. Manges AR, Geum HM, Guo A, Edens TJ, Fibke CD, Pitout JDD. Global Extraintestinal Pathogenic
365 Escherichia coli (ExPEC) Lineages. *Clin Microbiol Rev*. 2019;32(3). doi:10.1128/CMR.00135-18

366 33. Denamur E, Clermont O, Bonacorsi S, Gordon D. The population genetics of pathogenic Escherichia
367 coli. *Nat Rev Microbiol*. 2021;19(1):37-54. doi:10.1038/s41579-020-0416-x

368 34. Verschuuren TD, van Hout D, Arredondo-Alonso S, et al. Comparative genomics of ESBL-producing
369 Escherichia coli (ESBL-Ec) reveals a similar distribution of the 10 most prevalent ESBL-Ec clones and
370 ESBL genes among human community faecal and extra-intestinal infection isolates in the
371 Netherlands (2014–17). *J Antimicrob Chemother*. 2021;76(4):901-908. doi:10.1093/jac/dkaa534

372 35. Le Gall T, Clermont O, Gouriou S, et al. Extraintestinal Virulence Is a Coincidental By-Product of
373 Commensalism in B2 Phylogenetic Group Escherichia coli Strains. *Mol Biol Evol*. 2007;24(11):2373-
374 2384. doi:10.1093/molbev/msm172

375 36. Colquhoun RM, Hall MB, Lima L, et al. Nucleotide-resolution bacterial pan-genomics with reference
376 graphs. *bioRxiv*. Published online November 24, 2020:2020.11.12.380378.
377 doi:10.1101/2020.11.12.380378

378

379 **Table 1.** Characteristics of long-term and short-term carriers. Matching was performed on sex,
380 age and travel destination (United Nations subregions). These characteristics are depicted in
381 italic. Data are presented as number (%) for all characteristics except age and travel duration,
382 which is presented as median (IQR).

		Long-term carriers (n = 17)	**Short-term carriers (n = 33)**
Male		4 (23.5%)	8 (24.2%)
Age (years)		50.2 (41.7-59.4)	51.4 (36.5-58.1)
Continent visited during index travel	*Asia*	16 (94.1%)	31 (93.9%)
	North and South America	1 (5.9%)	2 (6.1%)
Travel duration (days)		20 (17-28)	19 (14-21)
Admitted to hospital during index travel		0 (0%)	0 (0%)
Travel to the same country as index travel, within 12 months after return from index travel		1 (5.9%)	0 (0%)
Antibiotic usage	Within 3 months before index travel	0 (0%)	0 (0%)
	During index travel	3 (17.6%)	2 (6.1%)
	Within 1 month after return from index travel	0 (0%)	0 (0%)
	Within 1 to 3 months after return from index travel	1 (5.9%)	2 (6.7%)
	Within 3 to 6 months after return from index travel	0 (0%)	2 (7.1%)
	Within 6 to 12 months after return from index travel	3 (17.6%)	2 (7.4%)
Traveler's diarrhoea		6 (35.3%)	9 (27.3%)

383

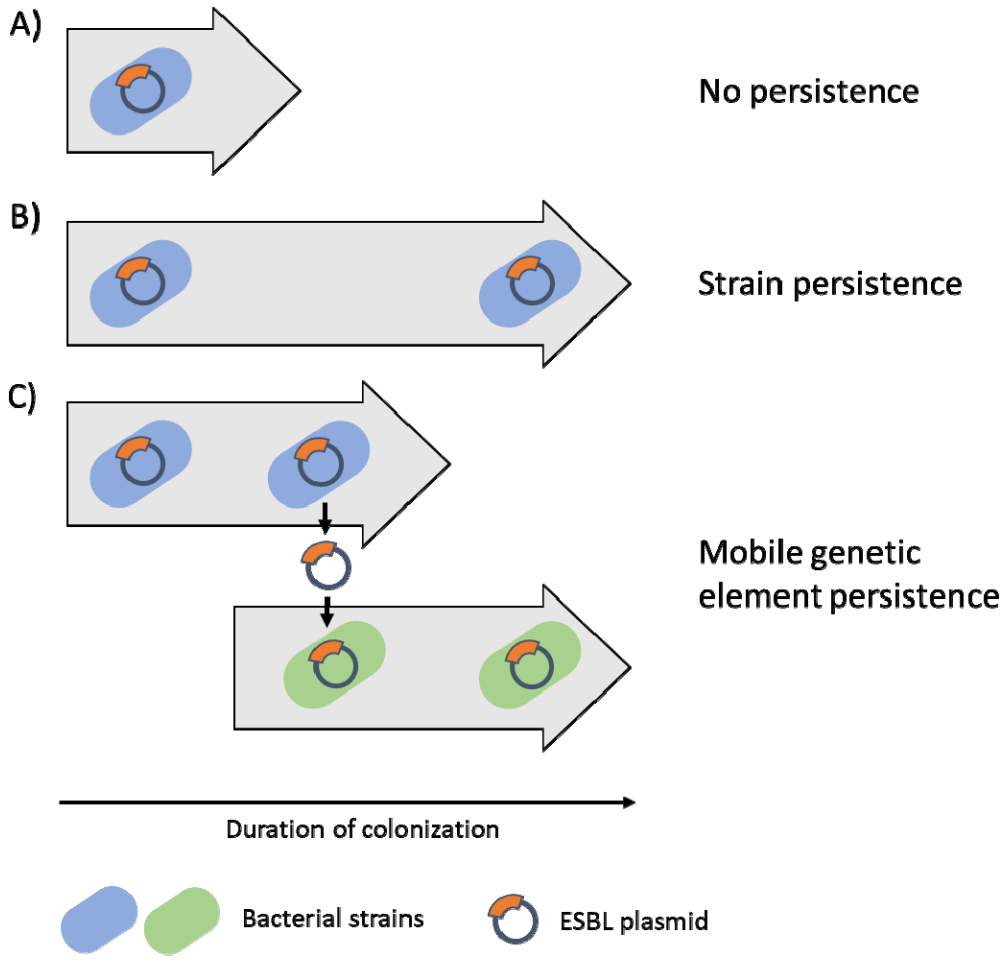

Figure 1. Schematic overview of main persistence mechanism of ESBL genes in the gut of returning travelers. Gray arrows indicate carriage duration over time. A) No persistence, where the colonizing strain is lost during the follow-up period. B) Strain persistence, where a bacterial strain harboring an ESBL gene is continuously present. C) Persistence through mobile genetic elements, where the original strain harboring the ESBL gene has been lost, but the ESBL gene (located on a mobile genetic element) has been passed on to another strain.

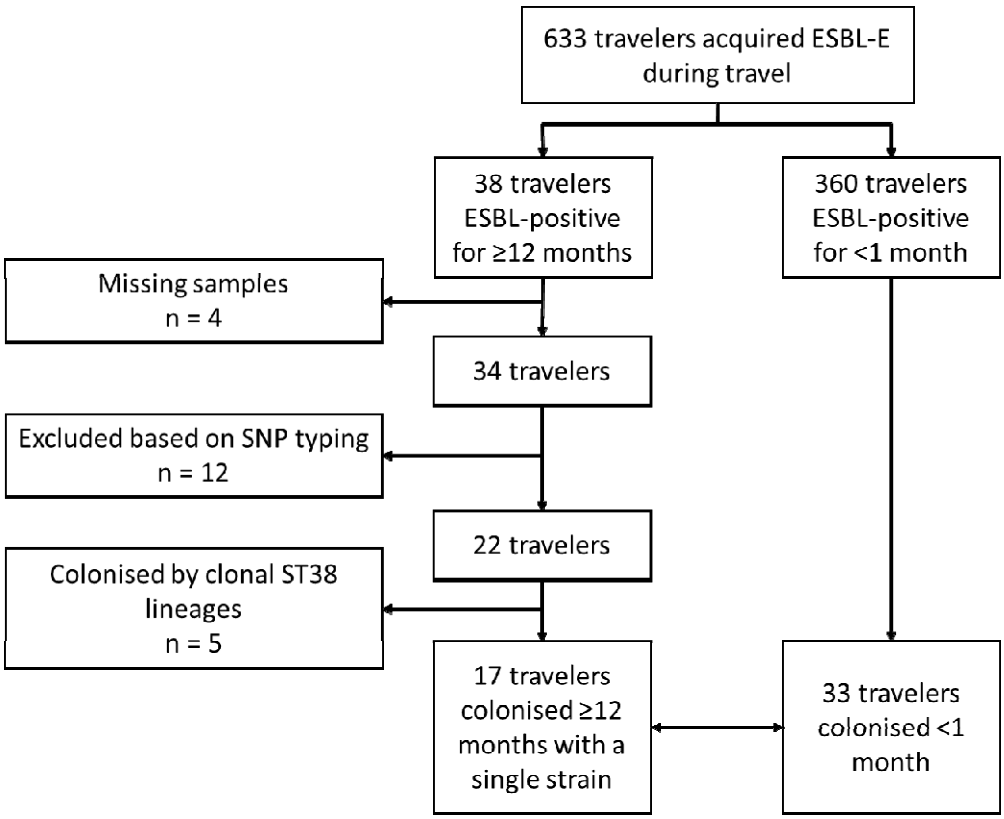

Figure 2. Study flowchart. Strain persistence was defined as isolates from a single traveler, sample twelve months apart with identical ESBL alleles, identical MLST profiles and 25 or fewer genome-wide SNPs difference. Clonal ST38 isolates were excluded. Seventeen long-term carriers (colonized ≥12 months with a single strain) were matched on sex, age and travel destination to thirty-three short-term carriers (colonized <1 month).

397 **Figure S1**. Core genome phylogeny of 1805 clonal complex 38 *E. coli* strains. Outer ring indicates

398 continent on which the strain was isolated. The two clonal ST38 lineages present in the COMBAT

399 collection are marked in red and blue. Tree and metadata available through iTOL:

400 https://itol.embl.de/tree/2131278347268071589210657.