1 **Title**

2 Extraintestinal pathogenic (ExPEC) lineages explain prolonged carriage of travel-acquired

3 extended-spectrum β-lactamase-producing *Escherichia coli*

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. Telephone:

17 +31205664868.

18 [#]Members are listed in the appendixes

Abstract

Fecally carried extended-spectrum β-lactamase (ESBL)-producing *Escherichia coli* (ESBL-Ec) are frequently acquired during international travel, contributing to global spread of AMR. However, determinants of long-term carriage of travel-acquired ESBL-Ec are unknown.

From a prospective cohort study of 2001 international travelers, we selected all 28 who acquired ESBL-Ec during travel and subsequently carried ESBL-Ec for at least 12 months after return. We sequenced a total of 155 ESBL-Ec isolates from these long-term carriers and 54 age-, sex- and destination-matched short-term carriers (<1 month carriage). We confirmed persistence of ESBL-Ec in long-term carriers using SNP typing and compared ESBL-Ec from long-term and short-term carriers using in silico multi-locus sequence and phylogroup typing. We employed long-read sequencing to investigate ESBL plasmid dissemination.

We show that extraintestinal pathogenic (ExPEC) lineages of *Escherichia coli* are significantly more likely to persist in healthy travelers than other *E. coli* lineages, in the absence of antibiotic usage, mainly driven by sequence type (ST) 131 and phylogroup D *E. coli*. Additionally, we identified two epidemiologically unrelated clonal lineages of ST38 carrying a range of ESBL- and carbapenemase-encoding genes. Using public datasets, we demonstrate the recent parallel emergence of these lineages and their subsequent rapid global dissemination, which has major implications for epidemiological tracking of ST38 strains.

Our findings show that persistence of travel-acquired ESBL-Ec is mediated by a limited number of ExPEC lineages. These findings support recent insights suggesting ExPEC lineages have not necessarily evolved towards virulence, but rather towards efficient intestinal colonization.

Importance

In 2018, an estimated 1.4 billion international trips were undertaken according to the World Tourism Organization. Antibiotic resistant *Escherichia coli* are frequently acquired during travel due to contact with contaminated foodstuff, or fecal-oral, environmental and human-to-human transmission. Resistant *E. coli* that can persist in the travelers' gut for long periods of time after return from travel, are likely to contribute to further transmission. Using highly detailed genomic typing of resistant *E. coli*, isolated from a large prospective cohort of international travelers, we identified bacterial characteristics explaining long-term carriage. Our results provide important information that can be used to estimate the risk of long-term carriage when travelers return with acquired resistant *E. coli*.

Introduction

International travel contributes significantly to the spread of extended-spectrum β-lactamase (ESBL) gene positive *Escherichia coli* (ESBL-Ec).[1,2] ESBL-Ec are acquired often during international travel and typing have shown these ESBL-Ec to be genetically diverse.[3–5] Travel-acquired ESBL-Ec are generally quickly lost within the first month after return from travel, although studies show a fraction of travelers stay positive for ESBL genes long after return.[1,3–6]

Studying these long-term carriers of travel-acquired ESBL-Ec has proven challenging, mainly for two reasons. First, inclusion of a large number of travelers is needed to find enough long-term carriers. Secondly, without costly high-resolution typing methods, such as whole-genome sequencing (WGS), it becomes nearly impossible to reliably differentiate between persistence of

a travel-acquired strain and re-acquisition of a highly similar strain. 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 strain for at least 7 months.[6] Due to the low sample size, no ESBL-Ec attributes could be identified that were associated with long-term carriage. Given that a relatively small fraction of travel-acquired strains persists for more than 6 months after return,[1] a sufficiently large number of included travelers is needed to investigate long-term carriage in detail.

The COMBAT study represents the largest longitudinal study exploring the acquisition of ESBL-positive Enterobacteriaceae (ESBL-E) during international travel to date, including 2001 Dutch travellers.[1] Out of 1847 travelers who were ESBL-E negative before travel, 633 travelers (34.3%) acquired ESBL-E abroad, of whom 38 travelers (6.0%) were colonized for ≥12 months after ESBL-E acquisition, based on ESBL gene typing. Persistence of ESBL-E was linked to *E. coli* carrying CTX-M group 9 ESBL genes.[1] However, WGS analysis was not performed in the early phase of the COMBAT study and hence, persistent colonization of ESBL-producing bacteria could not be demonstrated.

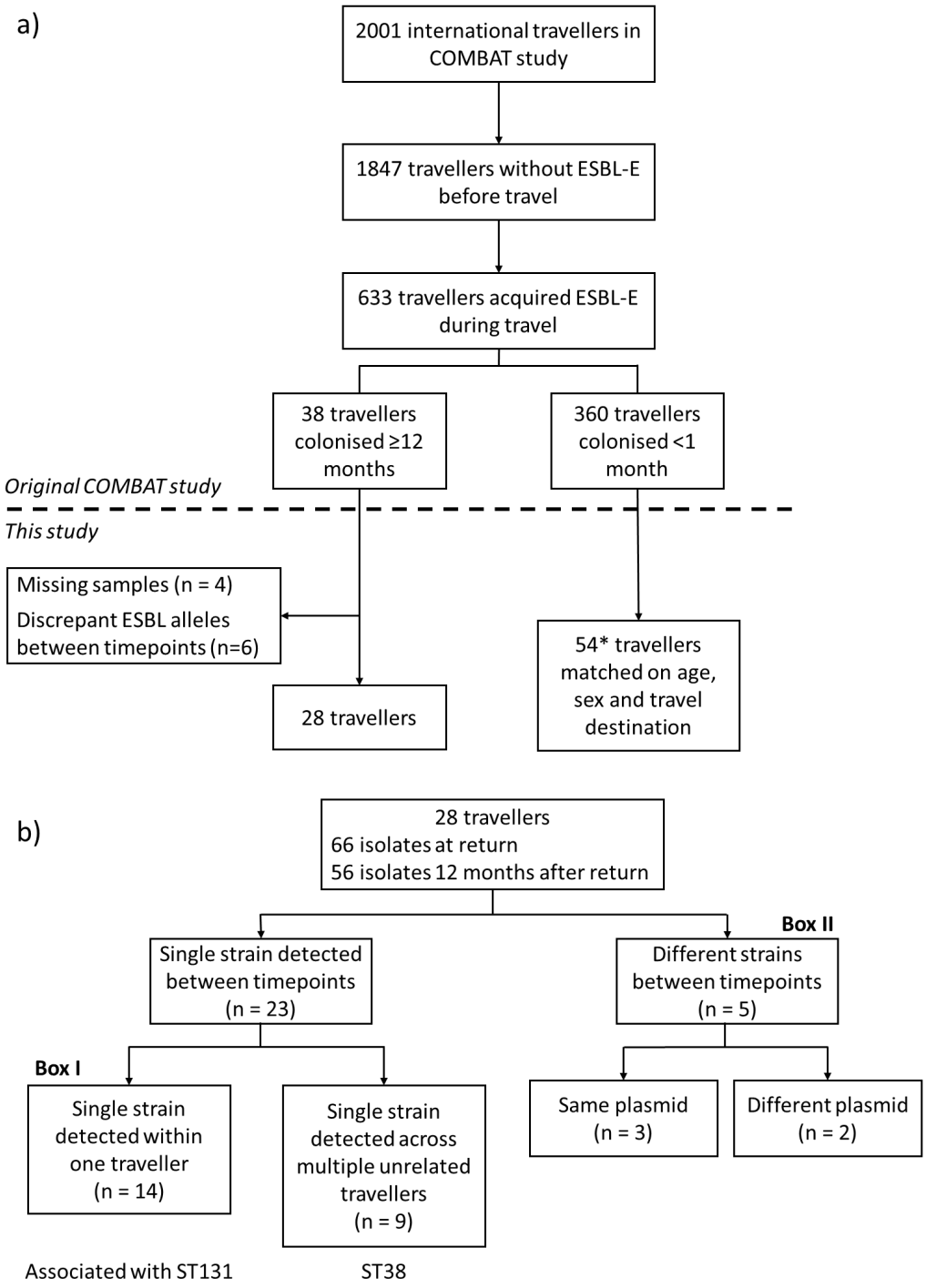


Figure 1. Study design a) Sampling flowchart. *Long-term carriers were matched by age, sex, and travel destination in a 1:2 ratio. For two long-term carriers only one matching short-term carrier could be identified. ESBL-E: extended-spectrum β-lactamase-producing

Enterobacteriaceae. **b)** Flowchart depicting the identification of persistent strains and plasmids. Clonal isolates representing a single strain were defined as isolates fewer than 17 SNPs/Mbp alignment apart. Persistent plasmids were defined as sharing >75% of open reading frames with >99% nucleotide identity. Nine epidemiologically unrelated travelers harbored isolates which were considered identical and shared between travelers, all of which belonged to ST38.

Results

From 2001 Dutch international travelers in the COMBAT study, we included all 38 travelers who acquired ESBL-E abroad and carried ESBL-E for at least 12 months after return.[1] Four travelers were excluded due to missing samples and six travelers were excluded as preliminary WGS analyses indicated they carried ESBL-Ec with different ESBL gene alleles between timepoints (Fig. 1a). These discrepant ESBL alleles between timepoints were not identified in the original COMBAT study since ESBL typing was performed at ESBL group level using microarrays.[1]

These 28 long-term carriers were matched by age, sex, and travel destination with travelers who carried travel-acquired ESBL-Ec for less than one month after return. From these, we randomly selected two short-term carriers per long-term carrier (Fig. 1a). For two long-term carriers, only one matching short-term carrier could be identified yielding 54 short-term carriers.

We performed Illumina WGS on all ESBL-Ec isolates sampled from these travelers on return from travel and 12 months thereafter (long-term carriers only). From the 28 long-term carriers, the final dataset consisted of 66 morphologically different ESBL-Ec strains that were isolated at

return from travel, and 56 ESBL-Ec that were isolated 12 months after return. Additionally, 67 isolates were available from the 54 matched short-term carriers (Fig. 1b).

Antibiotic usage was low before, during, and after travel and similar between long-term and short-term carriers (Table S1). No travelers were admitted to the hospital during travel in either group. International travel within 12 months after return from index travel was common for both groups. However, only 2 out of 28 long-term carriers (and 3 out of 54 short-term carriers) visited the same country as they visited during index travel, indicating reacquisition of the same strain from the same source was very unlikely.

Determination of clonality by SNP comparison between bacterial whole genome sequences

We determined the clonality of isolate pairs based on single nucleotide polymorphism (SNP) distances, by performing an all-versus-all comparison between all isolates in our dataset. All isolate pairs were compared to each other independent of their origin i.e. whether obtained from a single traveler or from unrelated travelers.

SNP distances are a commonly used measure to assess whether a strain has persisted for a certain period of time. In detail, the core genomes are aligned and the number of SNPs between the two isolates calculated.[32] Typically, a threshold of 5-10 SNPs has been used for *E. coli.*[6,33] However, the use of this global threshold suffers from two drawbacks. The first is that strains accumulate (point) mutations over time. The 5-10 SNP threshold is applied in situations where the sampling periods are much shorter than the 12 months representing our sampling period[6,33,34] necessitating the definition of a threshold optimized for a 12-month time frame.

Secondly, the aligned portions of the genomes can differ greatly between pairs of isolates due to the genomic differences between compared strains. In our dataset, the aligned fragments of genomes vary between 0.5 Mbp to 4.0 Mbp. Therefore, we decided to normalize for the number of aligned bases, and thus express differences between isolates as SNPs/aligned Mbp instead of absolute number of SNPs. A similar approach has been used for *Escherichia coli*[35] and for *Klebsiella pneumoniae*[36] isolates. SNP distances between isolates are provided in the appendixes (Table S2).

By identifying isolate pairs from unrelated travelers that were very few SNPs apart, we can also identify clonal lineages that are shared by multiple travelers from different households. Using our approach, we identified two clonal lineages of ST38 that appeared in nine unrelated travelers in our study, and also in recent external datasets from Europe, North America, South America, Oceania, and Asia.[37–55] These isolates harbored identical ESBL alleles and were found to be highly similar (3-16 SNPs/Mbp), even when comparing isolates from unrelated travelers who travelled to different countries or continents. Hence, for ST38, we cannot determine whether strains have persisted within a traveler, or whether different strains from the same ST38 lineage were lost and subsequently acquired. Possibly, to determine epidemiological relatedness, SNP typing of ST38 should be combined with accessory genome typing, for example by measuring differences between isolates in number of genes presence (Table S3 and Figure S1). Here we observed that for five out nine travelers, isolate pairs from the same traveler were more similar to each other than to any other isolate, including external datasets, which suggests persistence of a single strain. Even though our results suggest that accessory genome typing can

144 support strain typing, our inability to confidently determine persistence for the ST38 isolates
145 made us exclude these from further persistence analysis.

146 Excluding the 9 travelers with ST38, 19 out of the 28 travelers remained and together
147 contributed 96 isolates, of which 50 isolates were cultured at return from travel and 46 isolates
148 12 months after return (Box I and II in Figure 1b). In this data set, a SNP threshold of 17 SNPs per
149 1,000,000 aligned bases was found to be optimal to establish clonality between isolates from a
150 single strain based on F1 score (see Methods). We performed the SNP calling analysis with three
151 reference genomes belonging to ST10, ST73, and ST95 and obtained similar results (Table S4).

152

153 ESBL gene persistence is predominantly mediated by persistence of bacterial strains

154 Fourteen out of the nineteen travelers harbored pairs of ESBL-Ec isolates that were ≤17
155 SNPs/Mbp apart over the course of 12 months, indicating that the travel-acquired strain
156 persisted after return (Figure 1). For the remaining five travelers (Box II in Figure 1b), pairs of *E.*
157 *coli* isolates were not related, indicating loss of the travel-acquired strain. For one of these
158 travelers the isolates were 84.4 SNPs/Mbp apart, while for the four other travelers all isolates
159 were more than 900 SNPs/Mbp apart. The isolate pair which was 84.4 SNPs/Mbp apart
160 belonged to a single ST, which means these isolates would not have been distinguished had we
161 employed only MLST.

162 To approximate the diversity of ESBL-Ec present in the intestine from the five travelers who did
163 not harbor persistent strains, additional ESBL-Ec were isolated from stored fecal samples and
164 whole-genome sequenced (Illumina). Based on previous studies,[6] we chose to aim for a total of

five ESBL-Ec isolates per traveler per time point. This additional sequencing reinforced our earlier hypothesis that these five travelers did not harbor any persistent strains, but carried identical ESBL genes between timepoints.

To determine why five travelers lacking persistent ESBL-Ec strains harbored presumably persistent, travel-acquired ESBL genes we concentrated on ESBL-gene carrying plasmids, which could potentially be transferred between bacterial hosts. Their persistence was studied using Oxford Nanopore Technologies sequencing of one ESBL-Ec isolate per traveler per time point as this technology is able to resolve plasmid structures.

For two out of five travelers, we found significant similarity between the ESBL plasmids, while the *E. coli* strains harboring these plasmids were shown to be unrelated. The plasmids shared >75% of predicted genes with an average nucleotide identity (ANI) of 99.8% (Figure S2 and Table S5). These results indicate transfer and persistence of the ESBL plasmid in different bacterial strains over the course of 12 months.

For three remaining travelers, ESBL-Ec strains and plasmids were different between time points, while the ESBL genes were identical. The mobile elements adjacent to the ESBL genes were also different between plasmids, suggesting a very low probability of the resistance gene being transferred between plasmids through other mobile genetic elements such as transposons (Figure S2).

Taken together, our results indicate that persistence of ESBL-Ec is more commonly explained by persistence of strains (fourteen travelers) than persistence of the ESBL plasmid only (two travelers). For three travelers, we found that both strain and plasmid did not persist. Most

186 likely, these travelers re-acquired different strains with the same resistance allele on different
187 plasmids.

188

189 Phylogroup D and ST131 are associated with persistence

190 Next, we compared strains from 14 long-term carriers (≥12 months carriage of a single travel-
191 acquired strain, Box I in Fig 1b) with those of 27 age, sex, and travel destination matched short-
192 term carriers (<1 month carriage), to assess if there is a bacterial genetic basis for long-term
193 carriage. As described above, we matched each long-term carrier to two short-term carriers, yet
194 for one of the long-term carriers only one matching short-term carrier was available.

195 ST131 (in particular clade C of ST131) was overrepresented in the group of persisting strains.
196 ST131 strains persisted in 4 (28.6%) out of 14 long-term carriers and were absent from 27
197 matched short-term carriers (p-value Fisher's exact test: 0.0099). We note that in earlier smaller
198 studies, acquisition of ST131 *E. coli* during travel was absent[5] or rare.[4,6] We also detected strains
199 which belonged to phylogroup B2, similar to ST131, in short-term carriers but these strains were
200 of different sequence types and clonal complexes (ST1193 twice and ST2346 once). Phylogroup
201 D strains also appeared overrepresented, but this did not reach statistical significance (Figure
202 2a). Phylogroup D strains (ST69, n = 2; ST393, n = 2; ST405, n = 2; and ST449 n = 1) persisted in 7
203 (50%) out of 14 long-term carriers, and were found in 5 (18.5%) out of 27 short-term carriers (p-
204 value Fisher's exact test: 0.068). When 5 travelers harboring ST38 (also part of phylogroup D),
205 which might be persistent based on the accessory gene analysis, were included in the analysis
206 together with their matched controls, the association with phylogroup D was even stronger

(Figure S3, Fisher's exact test p-value: 0.00099). If we adhere to the definition of extraintestinal pathogenic *E. coli* (ExPEC) lineages from a systematic review,[56] we find a strong association between ExPEC carriage and long-term ESBL carriage (Figure 2b, Fisher's exact test p-value: 0.00029).

Based on these results, we conclude there is a bacterial genetic basis for long-term carriage. The short-term carriers acquired genetically diverse ESBL-Ec, as reported in earlier studies,[4–6] whilst only a subset of ESBL-Ec lineages was observed that was shown to efficiently persist.

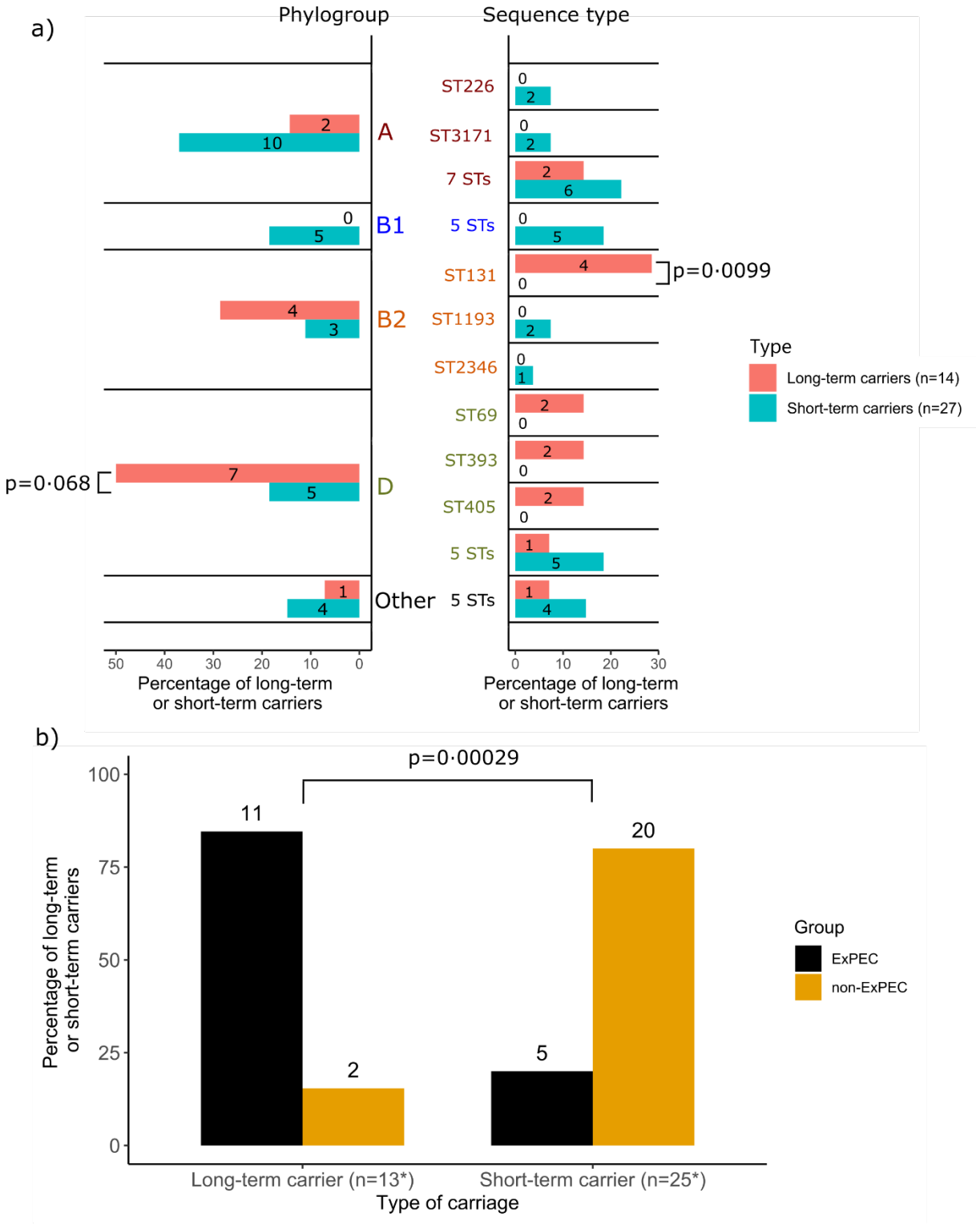
The bacterial lineages we found to be linked to persistence (ST131 and phylogroup D) have been described to harbor ESBL genes and cause extraintestinal disease frequently, in particular the pandemic and extraintestinal pathogenic ST131 lineage.[57,58] Several studies have suggested that ST131 is a successful colonizer of the human gut, besides being a pathogenic clade.[59–61] Our study shows that ST131 is frequently acquired by healthy travelers and is able to persist for more than a year after acquisition abroad in this population, independent of antibiotic usage.

From phylogroup D, we found sequence types 69, 393, 405, and 449 to be persistent in our collection. ST449 belongs to clonal complex 31, as does ST393. ST69, ST393, and ST405 are generally regarded as high-risk clones, not only due to their acquired ESBL resistance genes but also their capability to cause extraintestinal disease.[56,62–64] Due to the clonality of ST38 in our dataset, we cannot definitively assess whether strains of this sequence type have persisted. If some of these strains have actually persisted, as suggested by results of the accessory genome analysis, this will add to the population of phylogroup D strains that are capable of persisting after acquisition abroad.

228

229

230



231

232 **Figure 2. a)** Bar plots of phylogroup and sequence types of isolates that persisted in long-term

233 carriers or were isolated from matched short-term carriers. Bar lengths indicate the percentage

of long-term carriers (N=14) or short-term carriers (N=27) that carried that particular phylogroup/sequence type. Bar numbers indicate the absolute number of strains. Corresponding colors indicate which sequence types belong to which phylogroups. **b)** Association between ExPEC status and term of carriage. *The persistent isolate of one long-term carrier could not be assigned to a sequence type, and thus not to an ExPEC sequence type. This single long-term carrier and its two matched short-term carriers were thus excluded from this analysis.

Intercontinental dissemination of clonal ST38 lineages

Given the global dissemination of two clonal lineages of ESBL-positive ST38 as observed in our study, we further analyzed the ST38 population structure including isolates prevalent in public data. We downloaded all 1784 *E. coli* genomes from Enterobase (https://enterobase.warwick.ac.uk)[65] belonging to clonal complex 38 (CC38), isolated from 1979 onwards. ST38 comprises approximately 75% of the isolates in CC38, although other abundant STs are also present (e.g. ST963). We constructed a core genome phylogeny from these publicly available strains and our 21 ST38 strains and determined clusters within the population. This revealed the presence of several clonal expansions within CC38, two of which corresponding to the clonal lineages from our collection (clusters 1 and 6 in figure 3). These two lineages are represented by 78 (cluster 1) and 111 (cluster 6) genomes which within their respective clusters are separated by a median of 15.7 (range: 0-377) and 13.5 SNPs (range: 0-97) per Mbp alignment, respectively (Figure 4). Strains from both lineages have spread globally, and have been isolated on all continents except Africa and Antarctica.[37–55] Within clusters 1 and 6, 74 and 107 isolates share fewer than 17 SNPs/Mbp with another isolate in a core gene alignment,

respectively, similar to our findings. Strains from both clusters have been isolated from healthy individuals, as well as patients diagnosed with urinary tract infection or bacteremia. All isolates from the two clonal lineages were isolated recently (2012-now) while none of the 111 CC38 isolates sampled before 2012 belonged to these lineages. This might indicate that both lineages have emerged recently. ST38 is capable of acquiring different ESBL genes, illustrated by the fact that 64% of CC38 strains in Enterobase harbor various $bla_{CTX\text{-}M}$ genes. Additionally, CC38 is able to acquire and harbor carbapenemase genes such as $bla_{OXA\text{-}48}$ or $bla_{OXA\text{-}244}$ conferring resistance to last resort antimicrobials carbapenems.[66,67] Taken together, these findings highlight the recent parallel emergence and intercontinental spread of antimicrobial drug resistant ST38 clones.

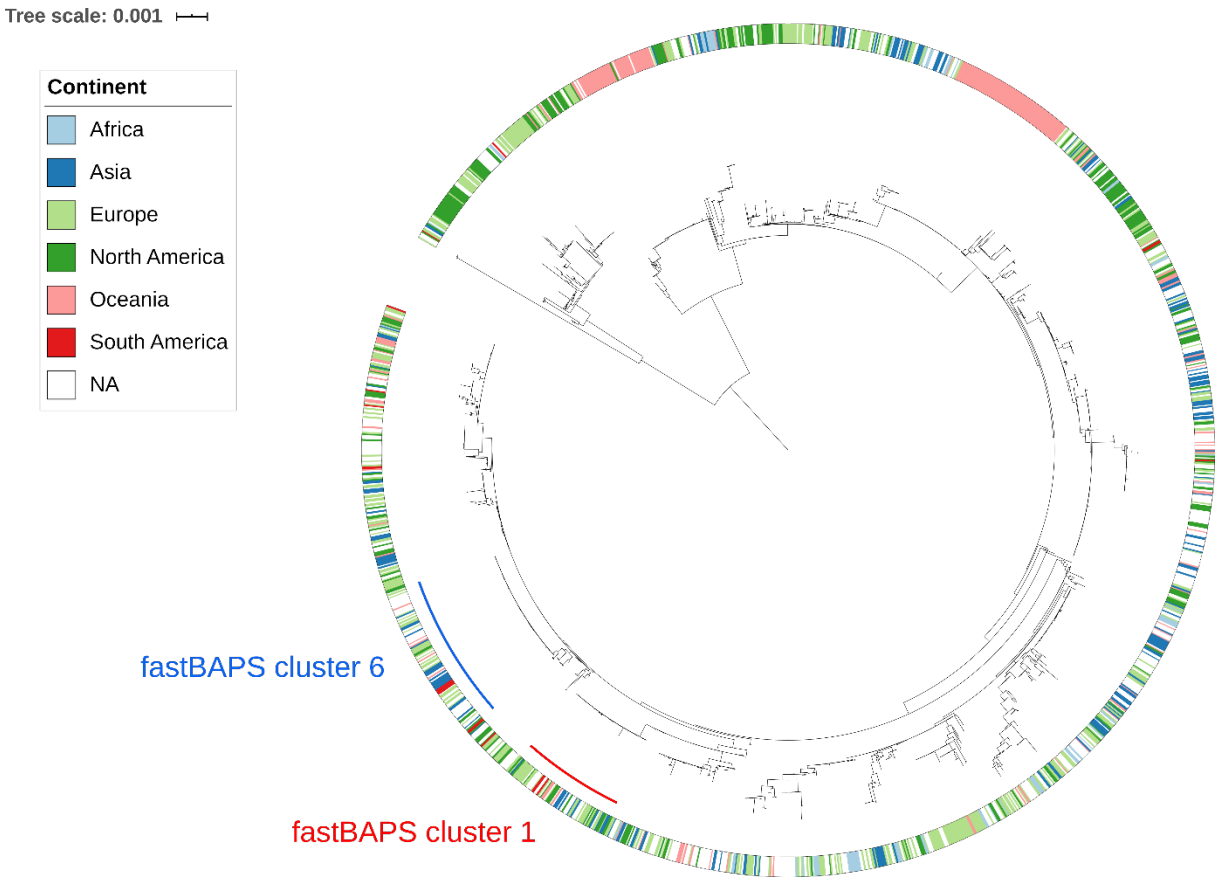


Figure 3. Core genome phylogeny of 1805 clonal complex 38 *E. coli* strains. Outer ring indicates continent on which the strain was isolated. The two clonal ST38 lineages present in the COMBAT collection are marked in red and blue. Tree and metadata available through iTOL:[68] https://itol.embl.de/tree/21312783472680715892106 57.

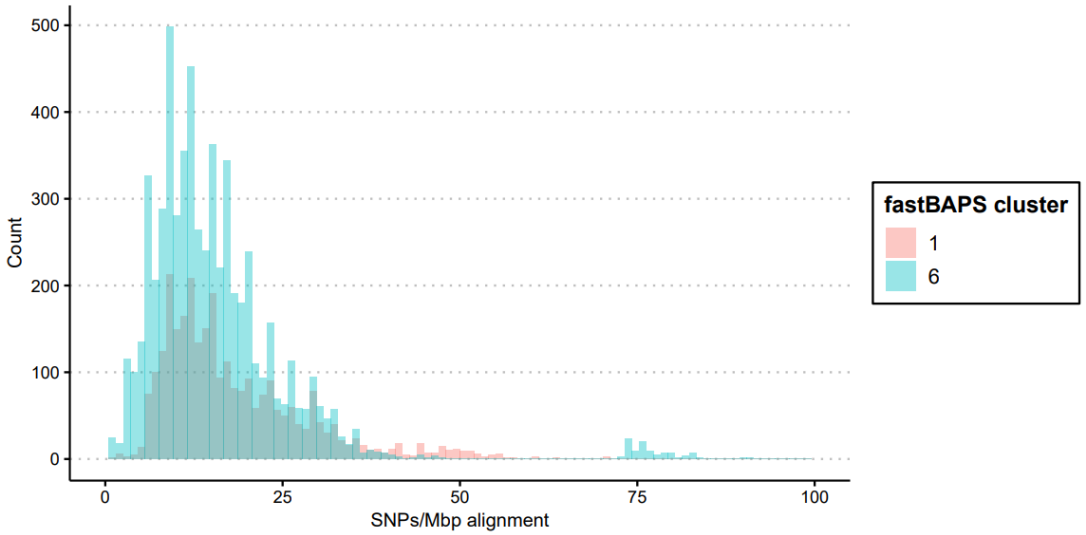


Figure 4. SNP histogram of 189 publicly available strains from two clonal ST38 lineages (fastBAPS clusters 1 and 6). The SNP threshold used in the main analysis to differentiate strains was 17 SNPs/Mbp alignment.

Discussion

Studying long-term carriage of travel-acquired ESBL-Ec necessitates large traveler cohorts and high-resolution typing methods. Previous work could not make strong assertions about the persistence of strains, either due to the number of included travelers or the typing methods employed.[3–6] By employing WGS to analyze strains isolated in a large traveler cohort, we were able to address most of the issues regarding sample size and typing methods. However, even with WGS data available, not all strains could be discerned with certainty. Two clonal ST38 lineages were present in our collection, and strains from these two lineages could not be reliably differentiated due to the within-lineage genetic similarity.

Our approach can be extended to other bacterial species, or other types of antimicrobial resistance. In the primary COMBAT study, 100 non-*E. coli* isolates were isolated from travelers who acquired ESBL genes abroad[1]. Similarly, other studies investigating ESBL gene acquisition have found ESBL-positive *E. coli* at high rates, but also non-*E. coli* species carrying ESBL genes after travel. Additionally, resistance genes other than ESBL genes can be investigated using our approach and software. Of heightened interest for such analyses are the emerging resistances against carbapenems (e.g. bla_{NDM}, bla_{KPC} or bla_{OXA} genes) or colistin (*mcr* genes).

In conclusion, we found that specific bacterial lineages are associated with long-term carriage of ESBL-Ec. Many of these lineages generally are regarded as ExPEC (ST131, ST393, ST405, and ST69).[56] Since the likelihood of onward transmission of travel-acquired ESBL-E depends on carriage duration after return, these persistent lineages may be more likely to be transmitted and cause infection. Additionally, we detected ST38 lineages that are abundant in public data and have recently spread across continents with little genetic differences. Our findings emphasize the risk of AMR transmission through travel, and provide information about which *E. coli* lineages in particular require our attention from both a public health and clinical medicine perspective.

Materials and methods

We included all 38 travelers from the COMBAT cohort who were colonized for ≥12 months with Enterobacteriaceae positive for the same ESBL gene group (microarray) at return from travel and at all subsequent time points (1, 3, 6, and 12 months after return from travel).[1] Whole-genome sequencing (Illumina HiSeq 2500) was performed on all ESBL-Ec strains available from

samples isolated at return from travel and 12 months after return from travel. Inclusion and exclusion are summarized in Figure 1. DNA extraction and library preparation were performed using the Qiagen Blood and Tissue DNA extraction kit and Kapa HTP library prep kit, respectively.

122 strains from 28 long-term colonized travelers were included, with 66 strains isolated at return from travel and 56 strains isolated 12 months after return. The strains were analyzed using a Snakemake v5.7.1[7] pipeline, available at https:// github.com/boasvdp/COMBAT. In short, Illumina sequencing data was trimmed using fastp v0.20.0,[8] assembled using the Shovill wrapper v1.0.9 (https://github.com/tseemann/shovill) for SPAdes,[9] and resistance genes were identified using AMRfinderplus v3.2.3.[10] *E. coli* phylogroups were predicted using EzClermont v0.4.3.[11] A core genome alignment of all strains from long-term carriers was made by mapping onto a common reference genome (*E. coli* ATCC 25922, Genbank: CP009072) using Snippy v4.4.5 (https://github.com/tseemann/snippy). SNP analysis was repeated with two different reference genomes (*E. coli* K-12 MG1655 and DSM 30083^T, Genbank: NC_000913, and NZ_CP033092, respectively). IQtree v1.6.12[12] was used to infer a phylogeny under a transversion model with equal base frequencies and a FreeRate model with 3 categories, as advised by Modelfinder[13] from the core genome alignment, using a count of constant sites from snp-sites v2.5.1.[14] Recombination events in the core genome alignment were identified using ClonalFrameML v1.12[15] and masked using maskrc-svg v0.5 (https://github.com/kwongj/maskrc-svg). SNP comparisons were made using snp-dists v0.7.0 (https://github.com/tseemann/snp-dists) and a modified version of snp-dists v0.7.0 which calculates alignment lengths (https://github.com/boasvdp/snp-dists). Comparisons were expressed as SNPs/Mbp alignment

between pairs. F1 score was calculated for SNP thresholds ranging from 1 to 1000 SNPs/Mbp by taking the harmonic mean of precision and recall. As there are no reference tests available against which our approach can be benchmarked, we calculated recall as the number of travelers in whom we were able to identify isolate pairs which were fewer SNPs/Mbp apart than the threshold, divided by the total of 19 travelers. False positives were defined as travelers from different households who carried isolate pairs between them that shared fewer SNPs/Mbp than the tested threshold. Data were plotted using ggplot2 v3.1.1,[16] ggthemes v2.4.0,[17] and patchwork[18] in R v3.5.1.[19] Tabular data was analyzed using Pandas v0.24.2[20] in Python v3.6.7.[21]

For the accessory genome typing of ST38, all 189 genomes from this study and Enterobase which are part of fastBAPS clusters 1 and 6 of CC38 as defined in this study were included. A pangenome was constructed using Roary v3.13.0[22] and differences in gene presence between genomes were calculated using snp-dists (https://github.com/tseemann/snp-dists). We quantified the differences between COMBAT genomes isolated from the same traveler, but at different timepoints. We also quantified differences with any other genome (COMBAT and Enterobase data).

Next, 54 travelers who were colonized for <1 month were matched using SPSS 26 by age (range +/- 7 years), sex, and travel destination (United Nations subregions) to 28 long-term carriers (Fig. 1a). After SNP analysis and exclusion of travelers carrying ST38, 14 long-term carriers remained together with their 27 matched short-term carriers. Illumina WGS was performed as described before on all ESBL-Ec strains isolated at return from travel from the remaining short-term carriers (33 strains, median 1 strain per traveler).

352 To determine why five travelers lacking persistent ESBL-Ec strains harbored presumably
353 persistent, travel-acquired ESBL genes, we concentrated on ESBL-gene carrying plasmids, which
354 could potentially be transferred between bacterial hosts. We generated Oxford Nanopore
355 Technologies sequencing data for 10 strains from these 5 travelers according to Van der Putten
356 et al. (2020)[23]. In short, strains were grown overnight at 37 ˚C in liquid LB. DNA extraction and
357 library preparation were performed using the Qiagen MagAttract HMW DNA Kit (Cat. No.
358 67563) and ONT native barcoding kit (Cat. No. EXP-NBD114), respectively. The library was
359 subsequently sequenced on an ONT MinION flowcell. Raw read data was filtered using Filtlong
360 v0.2.0 (https://github.com/rrwick/Filtlong) and assembled with corresponding Illumina data
361 using Unicycler v0.4.8.[24] Quality control was implemented at several steps in the pipeline using
362 FastQC v0.11.8,[25] Quast v4.6.3,[26] and MultiQC v1.6.[27] Plasmid comparison was performed using
363 ANICalculator.[28]

364 1784 genome assemblies marked as ST38 Complex were downloaded from Enterobase and
365 supplemented with 21 ST38 genomes from this study. Genomes were annotated using prokka
366 v1.14,[29] a core genome was defined using Roary v3.13.0[22] and 2678 core genes were
367 subsequently aligned with MAFFT v7.307.[30] IQtree v1.6.6[12] was used to infer the phylogeny
368 under a general time reversible model with base frequencies estimated from the data and a
369 FreeRate heterogeneity model with three categories (GTR+F+R3) advised by ModelFinder.[13]
370 Metadata were downloaded from Enterobase and supplemented with metadata available in
371 NCBI BioSample. Phylogenetic clusters were defined using fastBAPS v1.0.0[31] and AMRfinderplus
372 v3.2.3[10] was used to detect resistance determinants from the genomes.

We provide a Snakemake pipeline in which researchers can estimate the optimal SNP threshold in their own datasets on the basis of a core genome analysis (https://github.com/boasvdp/SNP-distance-analysis). The funders had no role in study design, collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the paper for publication.

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). BCLP 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, v1.0.0 archived through Zenodo at https://doi.org/10.5281/zenodo.4046159). Other traveler-related metadata will not be made available due to privacy concerns.

All code is available free of restrictions under the MIT license at https://www.github.com/boasvdp/COMBAT (v1.0.0 archived through Zenodo at https://doi.org/10.5281/zenodo.4046159). Researchers can analyze their own data using the Snakemake pipeline available at https://github.com/boasvdp/SNP-distance-analysis.

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.

Appendixes

Figure S1: Accessory gene typing of ST38 isolates.

Figure S2: Alignment of plasmids isolated twelve months apart from the same travelers.

Figure S3: Phylogroup comparison including accessory genome-typed ST38 isolates.

Table S1: Characteristics of long-term and short-term travelers included in this study.

Table S2: SNP comparisons between strains with associated metadata (7382 rows).

Table S3: Accessory gene typing of ST38 isolates.

Table S4: Comparison between different reference genomes in the SNP calling pipeline (1000 rows).

Table S5: Alignment metrics for plasmids isolated twelve months apart from the same travelers.

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

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

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

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

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

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

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

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

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

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

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

14. Page AJ, Taylor B, Delaney AJ, et al. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb Genomics*. 2016;2(4). doi:10.1099/mgen.0.000056

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

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

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

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

19. Team RC, others. *R: A Language and Environment for Statistical Computing*. Vienna, Austria; 2013.

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

21. Van Rossum G, Drake FL. *Python 3 Reference Manual*. CreateSpace; 2009.

22. Page AJ, Cummins CA, Hunt M, et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics*. 2015;31(22):3691-3693. doi:10.1093/bioinformatics/btv421

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

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

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

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

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

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

29. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014;30(14):2068-2069. doi:10.1093/bioinformatics/btu153

30. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol*. 2013;30(4):772-780. doi:10.1093/molbev/mst010

31. Tonkin-Hill G, Lees JA, Bentley SD, Frost SDW, Corander J. Fast hierarchical Bayesian analysis of population structure. *Nucleic Acids Res*. 2019;47(11):5539-5549. doi:10.1093/nar/gkz361

32. Baker S, Hanage WP, Holt KE. Navigating the future of bacterial molecular epidemiology. *Curr Opin Microbiol*. 2010;13(5):640-645. doi:10.1016/j.mib.2010.08.002

33. Jenkins C, Dallman TJ, Launders N, et al. Public Health Investigation of Two Outbreaks of Shiga Toxin-Producing Escherichia coli O157 Associated with Consumption of Watercress. *Appl Environ Microbiol*. 2015;81(12):3946-3952. doi:10.1128/AEM.04188-14

34. Dallman TJ, Byrne L, Ashton PM, et al. Whole-Genome Sequencing for National Surveillance of Shiga Toxin–Producing Escherichia coli O157. *Clin Infect Dis*. 2015;61(3):305-312. doi:10.1093/cid/civ318

35. de Been M, Lanza VF, de Toro M, et al. Dissemination of cephalosporin resistance genes between Escherichia coli strains from farm animals and humans by specific plasmid lineages. *PLoS Genet*. 2014;10(12):e1004776. doi:10.1371/journal.pgen.1004776

36. Gorrie CL, Mirčeta M, Wick RR, et al. Gastrointestinal Carriage Is a Major Reservoir of Klebsiella pneumoniae Infection in Intensive Care Patients. *Clin Infect Dis*. 2017;65(2):208-215. doi:10.1093/cid/cix270

37. Bergh MFQK den, Rossen JWA, Bruijning-Verhagen PCJ, et al. Whole-Genome Multilocus Sequence Typing of Extended-Spectrum-Beta-Lactamase-Producing Enterobacteriaceae. *J Clin Microbiol*. 2016;54(12):2919-2927. doi:10.1128/JCM.01648-16

38. Birgy A, Madhi F, Jung C, et al. Diversity and trends in population structure of ESBL-producing Enterobacteriaceae in febrile urinary tract infections in children in France from 2014 to 2017. *J Antimicrob Chemother*. 2020;75(1):96-105. doi:10.1093/jac/dkz423

39. Campos ACC, Andrade NL, Ferdous M, et al. Comprehensive Molecular Characterization of Escherichia coli Isolates from Urine Samples of Hospitalized Patients in Rio de Janeiro, Brazil. *Front Microbiol*. 2018;9. doi:10.3389/fmicb.2018.00243

40. Glaize A, Gutierrez-Rodriguez E, Hanning I, et al. Transmission of antimicrobial resistant non-O157 Escherichia coli at the interface of animal-fresh produce in sustainable farming environments. *Int J Food Microbiol*. 2020;319:108472. doi:10.1016/j.ijfoodmicro.2019.108472

41. Harris PNA, Ben Zakour NL, Roberts LW, et al. Whole genome analysis of cephalosporin-resistant Escherichia coli from bloodstream infections in Australia, New Zealand and Singapore: high prevalence of CMY-2 producers and ST131 carrying blaCTX-M-15 and blaCTX-M-27. *J Antimicrob Chemother*. 2018;73(3):634-642. doi:10.1093/jac/dkx466

42. Hastak P, Cummins ML, Gottlieb T, et al. Genomic profiling of Escherichia coli isolates from bacteraemia patients: a 3-year cohort study of isolates collected at a Sydney teaching hospital. *Microb Genomics*. 2020;6(5):e000371. doi:10.1099/mgen.0.000371

43. Liu CM, Stegger M, Aziz M, et al. Escherichia coli ST131-H22 as a Foodborne Uropathogen. *mBio*. 2018;9(4). doi:10.1128/mBio.00470-18

44. MacFadden DR, Melano RG, Coburn B, Tijet N, Hanage WP, Daneman N. Comparing Patient Risk Factor-, Sequence Type-, and Resistance Locus Identification-Based Approaches for Predicting Antibiotic Resistance in Escherichia coli Bloodstream Infections. *J Clin Microbiol*. 2019;57(6). doi:10.1128/JCM.01780-18

45. Marsh JW, Mustapha MM, Griffith MP, et al. Evolution of Outbreak-Causing Carbapenem-Resistant Klebsiella pneumoniae ST258 at a Tertiary Care Hospital over 8 Years. *mBio*. 2019;10(5). doi:10.1128/mBio.01945-19

46. Mellmann A, Bletz S, Böking T, et al. Real-Time Genome Sequencing of Resistant Bacteria Provides Precision Infection Control in an Institutional Setting. *J Clin Microbiol*. 2016;54(12):2874-2881. doi:10.1128/JCM.00790-16

47. Mukerji S, Stegger M, Truswell AV, et al. Resistance to critically important antimicrobials in Australian silver gulls (Chroicocephalus novaehollandiae) and evidence of anthropogenic origins. *J Antimicrob Chemother*. 2019;74(9):2566-2574. doi:10.1093/jac/dkz242

48. Pecora N, Zhao X, Nudel K, et al. Diverse Vectors and Mechanisms Spread New Delhi Metallo-β-Lactamases among Carbapenem-Resistant Enterobacteriaceae in the Greater Boston Area. *Antimicrob Agents Chemother*. 2019;63(2). doi:10.1128/AAC.02040-18

49. Peña-Gonzalez A, Soto-Girón MJ, Smith S, et al. Metagenomic Signatures of Gut Infections Caused by Different Escherichia coli Pathotypes. *Appl Environ Microbiol*. 2019;85(24). doi:10.1128/AEM.01820-19

50. Raven KE, Ludden C, Gouliouris T, et al. Genomic surveillance of Escherichia coli in municipal wastewater treatment plants as an indicator of clinically relevant pathogens and their resistance genes. *Microb Genomics*. 2019;5(5). doi:10.1099/mgen.0.000267

51. Roberts LW, Forde BM, Henderson A, et al. Intensive infection control responses and whole genome sequencing to interrupt and resolve widespread transmission of OXA-181 Escherichia coli in a hospital setting. *bioRxiv*. Published online November 30, 2019:850628. doi:10.1101/850628

52. Sütterlin S, Téllez-Castillo CJ, Anselem L, Yin H, Bray JE, Maiden MCJ. Heavy Metal Susceptibility of Escherichia coli Isolated from Urine Samples from Sweden, Germany, and Spain. *Antimicrob Agents Chemother*. 2018;62(5). doi:10.1128/AAC.00209-18

53. Syre H, Hetland MAK, Bernhoff E, et al. Microbial risk factors for treatment failure of pivmecillinam in community-acquired urinary tract infections caused by ESBL-producing Escherichia coli. *APMIS*. 2020;128(3):232-241. doi:10.1111/apm.13013

54. Thanh Duy P, Thi Nguyen TN, Vu Thuy D, et al. Commensal Escherichia coli are a reservoir for the transfer of XDR plasmids into epidemic fluoroquinolone-resistant Shigella sonnei. *Nat Microbiol*. 2020;5(2):256-264. doi:10.1038/s41564-019-0645-9

55. Zhou X, García-Cobos S, Ruijs GJHM, et al. Epidemiology of Extended-Spectrum β-Lactamase-Producing E. coli and Vancomycin-Resistant Enterococci in the Northern Dutch–German Cross-Border Region. *Front Microbiol*. 2017;8. doi:10.3389/fmicb.2017.01914

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

57. Nicolas-Chanoine M-H, Blanco J, Leflon-Guibout V, et al. Intercontinental emergence of Escherichia coli clone O25:H4-ST131 producing CTX-M-15. *J Antimicrob Chemother*. 2008;61(2):273-281. doi:10.1093/jac/dkm464

58. Nicolas-Chanoine M-H, Bertrand X, Madec J-Y. Escherichia coli ST131, an Intriguing Clonal Group. *Clin Microbiol Rev*. 2014;27(3):543-574. doi:10.1128/CMR.00125-13

59. McNally A, Kallonen T, Connor C, et al. Diversification of Colonization Factors in a Multidrug-Resistant Escherichia coli Lineage Evolving under Negative Frequency-Dependent Selection. *mBio*. 2019;10(2). doi:10.1128/mBio.00644-19

60. Sarkar S, Hutton ML, Vagenas D, et al. Intestinal Colonization Traits of Pandemic Multidrug-Resistant Escherichia coli ST131. *J Infect Dis*. 2018;218(6):979-990. doi:10.1093/infdis/jiy031

61. Vimont S, Boyd A, Bleibtreu A, et al. The CTX-M-15-Producing Escherichia coli Clone O25b: H4-ST131 Has High Intestine Colonization and Urinary Tract Infection Abilities. *PLOS ONE*. 2012;7(9):e46547. doi:10.1371/journal.pone.0046547

62. Cagnacci S, Gualco L, Debbia E, Schito GC, Marchese A. European Emergence of Ciprofloxacin-Resistant Escherichia coli Clonal Groups O25:H4-ST 131 and O15:K52:H1 Causing Community-Acquired Uncomplicated Cystitis. *J Clin Microbiol*. 2008;46(8):2605-2612. doi:10.1128/JCM.00640-08

63. Lee MY, Choi HJ, Choi JY, et al. Dissemination of ST131 and ST393 community-onset, ciprofloxacin-resistant Escherichia coli clones causing urinary tract infections in Korea. *J Infect*. 2010;60(2):146-153. doi:10.1016/j.jinf.2009.11.004

64. Woodford N, Turton JF, Livermore DM. Multiresistant Gram-negative bacteria: the role of high-risk clones in the dissemination of antibiotic resistance. *FEMS Microbiol Rev*. 2011;35(5):736-755. doi:10.1111/j.1574-6976.2011.00268.x

65. Zhou Z, Alikhan N-F, Mohamed K, et al. The EnteroBase user's guide, with case studies on Salmonella transmissions, Yersinia pestis phylogeny, and Escherichia core genomic diversity. *Genome Res*. 2020;30(1):138-152. doi:10.1101/gr.251678.119

66. Pitout JDD, Peirano G, Kock MM, Strydom K-A, Matsumura Y. The Global Ascendency of OXA-48-Type Carbapenemases. *Clin Microbiol Rev*. 2019;33(1). doi:10.1128/CMR.00102-19

67. van Hattem JM, Arcilla MS, Bootsma MC, et al. Prolonged carriage and potential onward transmission of carbapenemase-producing Enterobacteriaceae in Dutch travelers. *Future Microbiol*. 2016;11(7):857-864. doi:10.2217/fmb.16.18

68. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res*. 2016;44(W1):W242-W245. doi:10.1093/nar/gkw290

596