Genomic analysis uncovers a phenotypically diverse but genetically homogeneous *Escherichia coli* ST131 clone circulating in unrelated urinary tract infections

Gemma Clark¹, Konrad Paszkiewicz², James Hale³, Vivienne Weston⁴, Chrystala Constantinidou⁵, Charles Penn⁵, Mark Achtman³ and Alan McNally¹*

¹Pathogen Research Group, Nottingham Trent University, Nottingham, UK; ²College of Life and Environmental Sciences, University of Exeter, Exeter, UK; ³Environmental Research Institute, University College Cork, Lee Road, Cork, Ireland; ⁴Nottingham University Hospitals NHS Trust, Derby Road, Nottingham, UK; ⁵School of Biosciences, University of Birmingham, Birmingham, UK

*Corresponding author. Tel: +44-115-8483324; Fax: +44-115-8483384; E-mail: alan.mcnelly@ntu.ac.uk

Received 4 October 2011; returned 8 December 2011; revised 16 December 2011; accepted 20 December 2011

Objectives: To determine variation at the genome level in *Escherichia coli* ST131 clinical isolates previously shown to be phenotypically diverse.

Methods: The genomes of 10 ST131 isolates extensively characterized in previous studies were sequenced using combinations of Illumina and 454 sequencing technology. Whole-genome comparisons and phylogenetic comparisons were then performed across the strain set and with other closely related extraintestinal pathogenic *E. coli* (ExPEC) strain types.

Results: *E. coli* ST131 is overrepresented in a collection of clinical isolates, and there is large phenotypic variation amongst isolates. In contrast, genome sequencing of a selection of non-related clinical isolates shows almost no genetic variation between ST131 strains, and *E. coli* ST131 shows evidence of a genetically monomorphic pathogen showing a similar evolutionary trend to hypervirulent *Clostridium difficile*.

Conclusions: A dominant circulating clone of *E. coli* ST131 has been identified in unrelated clinical urine samples in the UK. The clone splits into two distinct subgroups on the basis of antimicrobial resistance levels and carriage of extended-spectrum β-lactamase plasmids. This provides the most comprehensive snapshot to date of the true molecular epidemiology of ST131 clinical isolates.

Keywords: *E. coli* ST131, genomes, phenotypes, phylogeny, ExPEC, ESBLs

Introduction

Urinary tract infections (UTIs) are among the most common bacterial infectious diseases in the world, with an estimated 20% of women over the age of 18 suffering from a UTI in their lifetimes.¹ Of infections among otherwise healthy women, some 80% are caused by extraintestinal pathogenic *Escherichia coli* or ExPEC.² In addition to the disease burden of UTI, ExPEC are also of significant concern due to the levels of antimicrobial resistance observed in isolates. Epidemiological studies show resistance to front-line antibiotics such as ciprofloxacin and trimethoprim in as many as 20%–45% of isolates tested in large cohorts across Europe, North America and South America.¹,³ Of greater concern is the observed level of extended-spectrum β-lactamase (ESBL) gene carriage in ExPEC.⁴,⁵ ESBLs render bacteria resistant to multiple antimicrobials including the cephalosporins, meaning that only carbapenems remain as drugs of choice for the treatment of some ESBL producers.

Molecular epidemiological analysis of ESBL-positive ExPEC isolates by multilocus sequence typing (MLST) has uncovered the emergence of an apparently dominant sequence type (ST) of ExPEC among UTIs and other extraintestinal infections, namely ST131.⁶ This ST is composed of *E. coli* O25b:H4 strains, and has been implicated as the major cause of dissemination of the CTX-M-15 class of ESBL gene.⁶ ST131 isolates also counter the hypothesis that bacteria exhibiting high levels of antimicrobial resistance do so at the expense of a fitness advantage, which results in decreased pathogenesis.¹ ST131 isolates are also associated with high levels of virulence-associated gene (VAG) carriage⁴ and have been implicated in large-scale disease outbreaks,⁹,¹⁰ leading to the hypothesis that ST131 is a pandemic ExPEC clone.¹¹ Previous work by our group investigated the organisms present in polymicrobial and monomicrobial urine samples, and uncovered the presence of *E. coli* exhibiting high levels of antimicrobial resistance and a hyper-invasive phenotype in *in vitro*...
### Materials and methods

**Genome sequencing of ST131 strains**

MLST was performed using the Achtman typing scheme (http://mlst.ucc.ie/mlst/dbs/Ecoli), adhering to the protocols published on the web site. MLST was performed using the Achtman typing scheme (http://mlst.ucc.ie/mlst/dbs/Ecoli). Bionumerics v.6.5 (http://www.applied-maths.com) was used to generate a nucleotide polymorphism (SNP)-based phylogeny showed that the strains are genetically homogeneous and that the isolates sub-cluster according to antimicrobial resistance and ESBL phenotypes. To address the suggestion that ST131 is a monomorphic pathogen, we investigated whether there was variation in phenotype within the ST131 cluster. An F6 screening of cell invasion phenotypes were obtained as described previously.

**Open reading frames (ORFs)** were called using a minimum size of 300 bp fragments, and libraries were prepared using the Illumina TruSeq Genomic library preparation kit and multiplexed using 6 bp index sequences into a single lane. They were then sequenced using the Illumina GAIIx platform.

**Susceptibility testing and plasmid carriage**. In combination, our data show the circulation of a dominant ST131 clone among unrelated cases of UTI in elderly patients resident in the East Midlands area (UK), and raise the suggestion that ST131 is a monomorphic pathogen whose selection is being driven by antimicrobial resistance.

---

### Table 1. Strains sequenced as part of this project

<table>
<thead>
<tr>
<th>Strain</th>
<th>ST</th>
<th>Patient source</th>
<th>Antibiotic resistance</th>
<th>Invasion (cfu/mL)</th>
<th>VAG profile</th>
</tr>
</thead>
<tbody>
<tr>
<td>UTI18</td>
<td>131</td>
<td>community</td>
<td>R S T R S S S S R S R</td>
<td>1.20E+03</td>
<td>PAI, fimH, fyuA, iutA, troT, kpsMT II, K5</td>
</tr>
<tr>
<td>UTI188</td>
<td>131</td>
<td>community</td>
<td>R S S S S S S S S R R</td>
<td>3.22E+03</td>
<td>papC, papG allele II, papG II, III, PAI, papA, fimH, troT</td>
</tr>
<tr>
<td>UTI226</td>
<td>131</td>
<td>hospital</td>
<td>S S S S S S S S S T R</td>
<td>9.44E+03</td>
<td>PAI, fimH, iha, fyuA, troT, kpsMT II, K5</td>
</tr>
<tr>
<td>UTI306</td>
<td>131</td>
<td>community</td>
<td>R R R R R S S S S R R</td>
<td>7.22E+04</td>
<td>PAI, papA, fyuA, iutA, troT, kpsMT II, K5</td>
</tr>
<tr>
<td>UTI423</td>
<td>131</td>
<td>community</td>
<td>R S R S S S S S S S R</td>
<td>1.01E+05</td>
<td>PAI, fimH, fyuA, iutA, troT, kpsMT II, K5</td>
</tr>
<tr>
<td>UTI524</td>
<td>131</td>
<td>community</td>
<td>R S S S S S S S S R R</td>
<td>7.20E+01</td>
<td>PAI, fimH, fyuA, iutA, troT, kpsMT II, K5</td>
</tr>
<tr>
<td>UTI570</td>
<td>131</td>
<td>community</td>
<td>R S S S S S S S S R S</td>
<td>7.83E+05</td>
<td>PAI, fimH, fyuA, iutA</td>
</tr>
<tr>
<td>UTI587</td>
<td>131</td>
<td>community</td>
<td>R S R R R S S S R S R</td>
<td>1.34E+05</td>
<td>PAI, fimH, fyuA, iutA, troT, kpsMT II, K5</td>
</tr>
<tr>
<td>UTI62</td>
<td>131</td>
<td>community</td>
<td>R S R R R S S S R S R</td>
<td>1.05E+02</td>
<td>PAI, fimH, fyuA, iutA, troT, kpsMT II, K5</td>
</tr>
<tr>
<td>UTI48</td>
<td>12</td>
<td>community</td>
<td>R S S S S S S S S R R</td>
<td>1.09E+03</td>
<td>PAI, fimH, fyuA, kpsMT II, K5</td>
</tr>
</tbody>
</table>

**Antibiotic resistance**

AMP, ampicillin (32 mg/L); RAD, ceftazidime (32 mg/L); CTX, cefotaxime (1 mg/L); CAZ, ceftazidime (1 mg/L); TAZ, piperacillin/tazobactam (85 mg/L); TMP, trimethoprim (2 mg/L); CIP, ciprofloxacin (4 mg/L); GEN, gentamicin (2 mg/L); AMC, amoxicillin (32 mg/L); NET, nitrofurantoin (32 mg/L); MEM, meropenem (2 mg/L).

**VAG profile**

VAG abbreviations: papC, papG, papA, regions within the pap operon that code for P pil; afadraBC, DR adhesins; PAI, CFT073 PAI marker; fimH, mannose-specific adhesion subunit of type 1 fimbriae; fyuA, yersiniabactin; iutA, aerobactin; troT, serum resistance; kpsMT II, group II capsule synthesis, K5, K5 capsule synthesis.

The ST131 strains selected for sequencing represent the variation within the ST131 study population with regards to antibiotic resistance, CTX-M-15 possession, ability to invade epithelial cells and VAG possession.
102 nt, and BLAST and PFAM scans were run on the resulting ORFs. For 454 sequencing, genomic DNA for each strain was sheared into approximately 8 kb fragments. Paired-end libraries were prepared according to the 454 Sequencing 8 kb Paired End Library Preparation Method Manual (454 Life Sciences/Roche, Branfan, CT, USA). Emulsion PCR were performed for enrichment titration and sequencing according to the manufacturer’s instructions. Titanium sequencing for each library was performed on a 454 GS-FLX apparatus. The reads were assembled and scaffolds were constructed using Newbler (v 2.5). The study information and raw sequence files have been deposited in the Sequence Read Archive at the European Nucleotide Archive (ENA) under accession number ERP001095.

**Comparative genomics**

Genome sequences were manually annotated using Artemis (http://www.sanger.ac.uk/resources/software/artemis) and BLAST functions. Genomes were compared in a pairwise fashion using BRIG.13 To determine levels of SNP variation in the reference genomes for strains belonging to ST95 and ST73, genome sequences were aligned using progressiveMauve (http://asap.ahabs.wisc.edu/software/mauve/overview.html), and the SNP data were exported as a spreadsheet. SNPs were manually curated to remove any ambiguous calls, and to remove insertions.

**Whole-genome-based phylogeny reconstruction**

The phylogeny of ST131 in relation to the UT189 reference genome and the outlier ST12 strain was performed by aligning genome sequences using progressiveMauve,14 and the common core genome was extracted using the stripSubsetLCB script. Bayesian phylogeny was inferred using ClonalFrame15 from the 50% consensus of 10 runs with 10000 iterations. SNP profiling of the Illumina-sequenced strains was performed on a 454 GS-FLX apparatus. The reads were assembled and scaffolds were constructed using Newbler (v 2.5). The study information and raw sequence files have been deposited in the Sequence Read Archive at the European Nucleotide Archive (ENA) under accession number ERP001095.

**Results**

**Phenotypic variation in E. coli ST131 isolated from unrelated clinical UTI cases**

As part of a wider study into the microbial populations of UTIs in elderly patients, 150 E. coli isolates were obtained from 250 unrelated clinical urinary tract samples from patients (aged 70 or over) across the East Midlands (population ~5 million).3 During this study, variation in antimicrobial resistance and epithelial cell invasion was demonstrated within the E. coli isolates.3 To assess the phenotypic variation that existed within the E. coli ST131 population and compare this against other STs, epithelial cell invasion (Figure 1a) and antimicrobial resistance levels (Figure 1b) were overlaid against a minimum spanning tree (MST) of the ExPEC isolates. The overlaid MST showed variation in phenotype within the ST131 isolates. There was variation in the levels of antimicrobial resistance within the group and in the ability to exhibit high cell invasion described previously in this group of isolates.3

**Improved quality draft genome of an ST131 isolate uncovers common ExPEC genomic traits**

In an attempt to further characterize E. coli ST131, one isolate from our strain collection was chosen for high quality draft genome sequencing. Strain UT18 was chosen as it is highly antimicrobial resistant with average invasion levels (Table 1), and was sequenced using a combination of Illumina and 454 sequencing. The genome of UT18 is similar to the recently published NA114 E. coli ST131 genome sequence in that it contains no discernible ‘novel’ regions that would account for increased fitness or pathogenicity when compared with the available genomes of ExPEC isolates (Figure 2). The pathogenicity island (PAI) that encodes cnf, haemolysin and the intact pap operon is absent, as is the sfa fimbrial operon, and there is a transposase insertion in the fimbB gene of the Type I fimbrial operon. UT18 does contain a fully intact high pathogenicity island (HPI) encoding the yersiniabactin locus, and also contains two flagella-encoding regions. The first region is identical to the flagella operons present in other published ExPEC genome sequences, whilst the second is a truncated version of Flag-2 found in the enterogaggregative E. coli O42 genome sequence,16 and in ExPEC strain UMN026, as well as the published E. coli O111 and O26 EPEC genome sequences.

Comparative analysis of regions outside the accessory virulome of ExPEC highlighted differences in metabolic-pathway-encoding genes between ST131 and the other published ExPEC genome sequences. The idnK and idnDOTR operons, encoding the L-idonate catabolism pathway, are completely missing from ST131. This pathway is a subsidiary pathway for gluconate metabolism in E. coli and is also termed the GntII system. The ancestral asc operon encoding a combined arbutin/salicin/cellobiose uptake and metabolism pathway is also affected by deletions of ascF and ascB, the PTS transporter enzyme and phospho-β-glucosidase enzyme, respectively, which are transcribed from a single promoter.18 Also absent are the putative ABC transporter genes yddA and yddB, and the yrhA and yrhB genes located in a region encoding both the GntI gluconate uptake and metabolism pathway and the GGT small peptide transporter.19

**Illumina sequencing of unrelated ST131 clinical isolates suggests circulation of a genetically homogeneous clone**

In order to confirm that the high quality draft genome sequence strain was representative of our population, a further nine ST131 strains isolated from unrelated clinical samples that displayed the full observed spectrum of phenotypic traits (Table 1) were sequenced using the Illumina GAIIx (8 isolates) or 454 (1 isolate), and draft de novo assemblies were produced. Stepwise BLAST comparisons using BRIG13 were performed of the draft de novo assembled genome sequences, both against our improved quality UT18 genome sequence and against the recently announced NA114 genome sequence of an Indian ST131 isolate.20 These comparisons showed no strain-specific insertions or deletions of accessory mobile islands within our strain set, but did show two regions differing from NA114, which were annotated as fragments of plasmids (Figure 3). This heterogeneity is not observed in ST73 and ST95, where there is variation in the carriage of PAIs between strains within the complex.

SNP profiling of the Illumina-sequenced strains was performed against the published UT189 reference genome sequence, as well as the genome sequence of UT18, an ST12 isolate from our strain collection. SNP profiling shows that the ST131 strains are genetically homogeneous. A total of 15060
Figure 1. Minimum spanning trees of ExPEC isolated from our previous studies, with phenotype overlays for in vitro epithelial cell invasion (a) and antimicrobial resistance (b). This figure appears in colour in the online version of JAC and in black and white in the print version of JAC.
SNPs are conserved between the ST131 strains compared with UTI89, with 1324 SNPs between the ST131 strains, 371 of which are non-synonymous. Strain UTI226 was the most divergent amongst our cohort but had only 460 strain-specific SNPs, with the remaining strains having only 10–60 strain-specific SNPs. Such low-level SNP variation has not been reported in *E. coli* and is rare in Enterobacteriaceae in general, and is more akin to monomorphic highly pathogenic and host-restricted subsets of species such as *Salmonella enterica* serotype Typhi. To ascertain if this monomorphic observation was common across *E. coli* ST complexes, the level of SNP variation was determined in ST95 and ST73 using the published genome sequences of strains from these complexes (Table 2). ProgressiveMauve alignments were performed, and the extracted SNP file manually curated to remove deletions and ambiguous SNP calls. The results showed 14413 SNPs between the three ST95 strains, and 9059 SNPs between the two ST73 strains. Mapping of the ST131-specific SNPs against the UTI18 genome showed that the SNPs were not randomly distributed, suggesting that recombination has played a significant role in the emergence of our ST131 clone (Figure 4). The metabolic operons *gic, glp, ytf* and *tre* are all ST131 SNP hotspots, as is the *fim* operon. Conversely, both flagella operons and the HPI show no SNPs at all.

Whole genome alignments were performed on our 10 ST131 isolates, NA114, the ST12 outlier strain UTI48 and the reference genome UTI19, and the phylogeny was constructed using Clonalframe. When strain phenotypes were mapped against the resulting phylogenetic tree (Figure 5) there was a split between CTX-M-15 plasmid-positive isolates and CTX-M-negative strains, which also mirrored levels of antimicrobial resistance observed in the isolates. In addition, the CTX-M-positive strains also had identical VAG profiles using a multiplex PCR detection method. There was no correlation with invasive phenotype, community or hospital acquisition, or clinical recurrence of UTI in patients from whom the original strains had been isolated.

**Discussion**

Previous work by our group showed variation in phenotypic characteristics among ExPEC isolated from elderly patients. Molecular epidemiology of this group of strains uncovered a large
proportion of ST131 isolates within the population, exhibiting variation in VAG carriage. In this study we further investigated this apparent variation in the phenotype of ST131 by mapping phenotypic traits against a minimum-spanning tree of our ExPEC population. Our data corroborates the current ST131 literature reporting significant increases of isolation of the organism from extraintestinal infections, with ST131 as the most common ST isolated in our ExPEC population. Our data also shows variation in phenotype observed within our ST131 population, correlating with our earlier observation of variation in VAG carriage.

Table 2. Published reference genomes used in this study

<table>
<thead>
<tr>
<th>Strain</th>
<th>ST</th>
<th>Strain history</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>UTI89</td>
<td>95</td>
<td>uncomplicated cystitis</td>
<td></td>
</tr>
<tr>
<td>CFT073</td>
<td>73</td>
<td>acute pyelonephritis</td>
<td></td>
</tr>
<tr>
<td>ABU83972</td>
<td>73</td>
<td>asymptomatic bacteriuria</td>
<td></td>
</tr>
<tr>
<td>Apec01</td>
<td>95</td>
<td>poultry colibacillosis</td>
<td></td>
</tr>
<tr>
<td>E. coli 536</td>
<td>92</td>
<td>acute pyelonephritis</td>
<td></td>
</tr>
<tr>
<td>E. coli HS</td>
<td>46</td>
<td>human commensal</td>
<td></td>
</tr>
<tr>
<td>IA139</td>
<td>62</td>
<td>UTI</td>
<td><a href="http://www.genoscope.cns.fr/spip/Escherichia-fergusonii-coli-.html">http://www.genoscope.cns.fr/spip/Escherichia-fergusonii-coli-.html</a></td>
</tr>
<tr>
<td>S88</td>
<td>95</td>
<td>neonatal meningitis</td>
<td><a href="http://www.genoscope.cns.fr/spip/Escherichia-fergusonii-coli-.html">http://www.genoscope.cns.fr/spip/Escherichia-fergusonii-coli-.html</a></td>
</tr>
<tr>
<td>UMNO26</td>
<td>597</td>
<td>UTI</td>
<td><a href="http://www.genoscope.cns.fr/spip/Escherichia-fergusonii-coli-.html">http://www.genoscope.cns.fr/spip/Escherichia-fergusonii-coli-.html</a></td>
</tr>
</tbody>
</table>

Figure 3. BRIG alignment of the nine ST131 genomes sequenced using Illumina GAIIx. The comparisons are made relative to the Indian ST131 strain NA114, which lacks the plasmid DNA fragments found in our ST131 isolates annotated on the circular diagram. This figure appears in colour in the online version of JAC and in black and white in the print version of JAC.
carriage within the cohort. Most reports of ST131 populations have focused on the likelihood of an emerging clone, and focus on the ST131 isolates carrying CTX-M variants; however, our previous work, combined with data presented here, show that clinical ST131 isolates are phenotypically heterogeneous, and that this cannot be attributed to simple variation in carriage of the CTX-M-encoding plasmids.

In order to investigate if this phenotypic variation was mirrored by genotypic variation we genome-sequenced 10 ST131 strains isolated from unrelated clinical episodes in elderly patients living in a catchment area of approximately 5 million people. The strains were chosen to represent the wide spectrum in phenotypic and VAG carriage profiles observed in our population. In addition, our data was compared with the recently announced NA114 genome, an ST131 strain isolated in India. The striking observation from our data is the lack of variation across the genomes of the ST131 strains isolated. Previous ExPEC genome studies have shown heterogeneity in genome architecture and content among strains, including between strains of the same ST, as exemplified by UT189, APEC01 and S88 (all ST95), and by ABU83972 and CFT073 (both ST73). In contrast, all 10 of our ST131 isolates show characteristics of being genetically monomorphic, with no variation in accessory genome content beyond the carriage of antimicrobial resistance genes and associated plasmids. This would suggest the ST131 circulating in our population is not participating in accessory genome flux and that it is a stable clone. Similarly, there were no obvious discriminatory genomic signatures, such as novel or unusual PAIs or VAGs, although the absence of the sfa and pap fimbrial operons and deletion in the fimB gene of all isolates merits further study for biological relevance. Previous work by our group highlighted the fact that both the sfa and pap operons were statistically less frequently found in ExPEC strains exhibiting an increased virulence phenotype. The absence of P fimbriae in our clinical ST131-sequenced isolates and the insertion in fimB raises questions as to the true virulence nature of our ST131 isolates. The relevance of these mutations and the true virulence of our ST131 strains is the focus of current work in our group.

Figure 4. Circular diagram showing the location of ST131-specific SNPs relative to the UTI18 genome. The innermost ring is GC content; the two outermost rings are CDS found on the coding and complementary strand. Red marks illustrate the positions of ST131 SNPs; SNP hotspot regions are annotated with arrows; the two regions completely free of SNPs are marked by rectangles outside of the circular diagram. This figure appears in colour in the online version of JAC and in black and white in the print version of JAC.
The genetically monomorphic nature of ST131 was further confirmed when phylogenetic analyses were performed based on whole-genome data. SNP analysis of the 10 ST131 genome sequences showed low-level polymorphism of 1324 SNPs between strains (typically 10–60 strain-specific SNPs, with one strain containing 386), in contrast to the 14413 SNPs between ST95 genome-sequenced strains and the 9059 between ST73 genome-sequenced strains. Indeed the levels of variation between our ST131 strains are similar to those observed in intra-strain variation during human bladder passage using ABU83972, where some 29 SNPs occurred, accompanied by one large deletion and four smaller deletions.29 Such low levels of variation are only seen in monomorphic, highly niche-restricted and pathogenic subsets of species, such as Salmonella Typhi and hypervirulent Clostridium difficile O27, where inter-strain SNP variation levels of 196421 and 187430 SNPs, respectively, have been reported. Both these organisms are subtypes that have independently evolved into highly pathogenic variants, and, in the case of Salmonella Typhi, are accompanied by gene loss and niche restriction. The inclusion of the Indian NA114 isolate in the middle of our phylogenetic tree raises the possibility that ST131 is a globally disseminated monophyletic clone that is evolving into subclades on the basis of antimicrobial resistance.

Together the data from our study provide evidence of the circulation of a genetically monomorphic E. coli ST131 clone as a dominant strain isolated from unrelated clinical cases. To our knowledge, this is the first time such a phenomenon has been reported for an ST of E. coli; most studies focus on pathotypes encompassing diverse STs. In order to elucidate the emergence of ST131 from a common environment to a dominant human pathogen a full genome-level investigation of a contemporaneous strain set separated geographically, temporally and by source reservoir is required, in conjunction with comparative studies of closely related strain types and more distant ExPEC relatives. This would allow detailed Bayesian analysis of clonal expansion of ST131 with accurate dating and provide clues as to the triggers for the evolution of pathogenic lineages of E. coli, particularly the role of antimicrobial resistance and ESBL carriage, in driving evolutionary selection of ST131. Such informative clues would be of great value in understanding not just the

Figure 5. Phylogenetic tree of the 10 ST131 isolates sequenced in this study with the outlier ST12 strain (UTI48) and the reference strain used to assemble sequences and call SNPs (UTI89). The number of discriminatory SNPs are shown. The VAG carriage profile of the isolates is also presented by the presence (filled squares) or absence (open squares) of genes, as determined by PCR in a previous study.17 Strain characteristics are mapped onto the tree according to the key. This figure appears in colour in the online version of JAC and in black and white in the print version of JAC.
emergence of ST131, but also how new dominant pathogenic variants of *E. coli*, such as that of the recent O104 epidemic, arise.

**Acknowledgements**

We would like to thank Nicholas Thomson and Thomas Connor of the Wellcome Trust Sanger Institute for advice and guidance during genome and phylogenetic analysis, and appraisal of the manuscript, and Xavier Didelot for advice with the stripSubsetLCBs algorithm.

**Funding**

This work was supported by a Dowager Countess Eleanor Peel Trust grant awarded to A. M. and grant 05/FE1/8882 from the Science Foundation Ireland to M. A.

**Transparency declarations**

The authors declare no competing or financial interests in this work.

**References**

32. Johnson TJ, Kariyawasam S, Wannemuehler Y et al. The genome sequence of avian pathogenic *Escherichia coli* strain O1:K1:H7 shares...
