# Outbreak Investigation Using High-Throughput Genome Sequencing within a Diagnostic Microbiology Laboratory

Norelle L. Sherry,[a,b] Jessica L. Porter,[c] Torsten Seemann,[d,e] Andrew Watkins,[f] Timothy P. Stinear,[c,d] Benjamin P. Howden[a,b,c,d]

Department of Microbiology, Austin Health, Melbourne, Australia[a]; Austin Centre for Infection Research (ACIR), Infectious Diseases Department, Austin Health, Melbourne, Australia[b]; Department of Microbiology and Immunology, University of Melbourne, Melbourne, Australia[c]; Department of Microbiology, Monash University, Melbourne, Australia[d]; Victorian Bioinformatics Consortium, Monash University, Melbourne, Australia[e]; Department of Paediatrics, Mercy Hospital for Women, Melbourne, Australia[f]

Next-generation sequencing (NGS) of bacterial genomes has recently become more accessible and is now available to the routine diagnostic microbiology laboratory. However, questions remain regarding its feasibility, particularly with respect to data analysis in nonspecialist centers. To test the applicability of NGS to outbreak investigations, Ion Torrent sequencing was used to investigate a putative multidrug-resistant *Escherichia coli* outbreak in the neonatal unit of the Mercy Hospital for Women, Melbourne, Australia. Four suspected outbreak strains and a comparator strain were sequenced. Genome-wide single nucleotide polymorphism (SNP) analysis demonstrated that the four neonatal intensive care unit (NICU) strains were identical and easily differentiated from the comparator strain. Genome sequence data also determined that the NICU strains belonged to multilocus sequence type 131 and carried the $bla_{CTX-M-15}$ extended-spectrum beta-lactamase. Comparison of the outbreak strains to all publicly available complete *E. coli* genome sequences showed that they clustered with neonatal meningitis and uropathogenic isolates. The turnaround time from a positive culture to the completion of sequencing (prior to data analysis) was 5 days, and the cost was approximately $300 per strain (for the reagents only). The main obstacles to a mainstream adoption of NGS technologies in diagnostic microbiology laboratories are currently cost (although this is decreasing), a paucity of user-friendly and clinically focused bioinformatics platforms, and a lack of genomics expertise outside the research environment. Despite these hurdles, NGS technologies provide unparalleled high-resolution genotyping in a short time frame and are likely to be widely implemented in the field of diagnostic microbiology in the next few years, particularly for epidemiological investigations (replacing current typing methods) and the characterization of resistance determinants. Clinical microbiologists need to familiarize themselves with these technologies and their applications.

Research in the field of pathogen biology has been transformed over the last several decades with the introduction of whole-genome sequencing, beginning with the complete sequencing of the *Haemophilus influenzae* genome in 1995 ([1]). This has led to significant developments in the study of molecular epidemiology, virulence, antimicrobial resistance, and vaccinology and in understanding complex microbial communities. More recently, the development of high-throughput (or "next-generation") sequencing technologies has meant, for the first time, that these methods fall within the financial and technical grasp of a medium or large diagnostic microbiology laboratory ([2]).

Next-generation sequencing (NGS) methods are also being used in smaller-scale local projects to determine epidemiology in outbreak settings ([2], [3]), examine the development of resistance mutations during antibiotic use in a single patient ([4], [5]), and identify bacteria in place of using 16S rRNA sequencing ([6]). However, the uptake of these technologies into the diagnostic laboratory setting has been slow. Here, we conducted a pilot project to assess the feasibility and practicability of applying NGS methods to address common clinical questions in a diagnostic microbiology laboratory setting.

**Outbreak description.** An outbreak of extended-spectrum beta-lactamase (ESBL)-producing *Escherichia coli* was suspected in the neonatal intensive care unit (NICU) of the Mercy Hospital for Women, Melbourne, Australia. Multidrug-resistant Gram-negative strains have not been detected previously in this NICU. Subsequently, a screening program was implemented which screened all babies in the NICU for ESBL *E. coli* by the use of rectal swabs in a 48-h period. The rectal swabs were cultured directly onto selective chromogenic medium (chromID ESBL agar, bioMérieux), and positive colonies were further identified and had antibiotic susceptibility testing performed on the Vitek 2 Compact system (bioMérieux). Aside from the index case, three out of 33 neonates (9%) were found to be positive for ESBL *E. coli* carriage by screening.

The first case (*E. coli* isolate BPH0657) was from a blood culture from the index case, a twin male born at 26 weeks gestation, who developed fatal sepsis and meningitis at 16 days postbirth. He was treated empirically with intravenous cefotaxime and gentamicin, according to NICU protocol, and died prior to the culture results becoming available. The second case (*E. coli* isolate BPH0530) was cultured from an eye swab from the twin brother of the index case, who was also positive by rectal swab screening. The first two cases (twins) were born to a paraplegic mother, who was managed by a tertiary spinal unit, and who had a history of recurrent urosepsis.

**TABLE 1** Sequencing and *de novo* genome assembly quality parameters (Ion Torrent PGM)*a*

| Quality parameter | Value for indicated *E. coli* strain | | | | |
| --- | --- | --- | --- | --- | --- |
| | BPH0657 | BPH0530 | BPH0532 | BPH0658 | BPH0659 (comparator strain) |
| Total no. of bases (Mbp) | 392.41 | 216.97 | 257.88 | 413.50 | 317.96 |
| Total no. of Q20 bases (Mbp) | 364.09 | 122.13 | 168.44 | 379.88 | 302.44 |
| Total no. of reads | 3,959,362 | 1,772,017 | 2,060,732 | 4,094,233 | 3,455,119 |
| Mean read length (bp) | 99 | 122 | 125 | 100 | 92 |
| Longest read length (bp) | 365 | 201 | 201 | 203 | 206 |
| % of reference covered | 89% | 88% | 88% | 89% | 84% |
| Avg depth of coverage ($\times$) | 63.43 | 33.37 | 40.91 | 67.71 | 48.48 |

*a* PGM, personal genome machine. Reads for all isolates were mapped to the *E. coli* S88 reference genome.

The two other NICU isolates (*E. coli* BPH0532 and BPH0658) were detected in neonates during rectal swab screening (asymptomatic colonization). The four isolates had the same biochemical profiles and antibiograms, consistent with an ESBL-type pattern, as follows: ampicillin resistant, cefoxitin susceptible, ceftriaxone resistant, ceftazidime resistant, ceftazidime-clavulanate susceptible, gentamicin resistant, tobramycin resistant, amikacin susceptible, ciprofloxacin resistant, and cotrimoxazole susceptible.

Our laboratory was asked to perform clonality testing to determine if this was a clonal outbreak. As we did not have established methods for clonality testing of Gram-negative organisms, we explored the utility of NGS to investigate this putative outbreak. All four NICU isolates were analyzed.

For comparison, another *E. coli* strain (BPH0659) with the same antibiogram was selected for sequencing. This organism was cultured from a fecal sample of an adult patient in an adjacent intensive care unit (ICU), as no other ESBL *E. coli* had previously been isolated from the NICU.

## MATERIALS AND METHODS

**Cultures and DNA extraction.** A single colony of each isolate was selected and cultured in tryptone soya broth (Oxoid) overnight at 37°C on a shaker. DNA extraction was performed on the broth cultures using the DNeasy Kit (Qiagen).

**Genome sequencing and data analysis.** Sequencing was performed using the Ion Torrent personal genome machine (Life Technologies, Guilford, CT) with 316 chips and 100-bp sequencing chemistry. *De novo* genome assembly and read mapping were performed with the CLC Genomics Workbench v5.1 (CLC bio A/S, Denmark), using the fully sequenced uropathogenic *E. coli* strain S88 (GenBank accession no. NC_011742) as the reference (7). Artemis was then used to explore the resulting FASTA files from the contigs of the partially *de novo*-assembled genomes (8).

**Epidemiological analysis.** For epidemiological analysis, another implementation of read mapping was used to compare the four outbreak isolates to all publicly available *E. coli* complete genome sequences (7, 9–26) (see Table S1 in the supplemental material). The reads from all genomes were aligned with the *E. coli* S88 reference using SHRiMP 2.2 (27). Single nucleotide polymorphisms (SNPs) were identified using Nesoni v0.70, which compares the aligned reads of each genome against the reference to construct a tally of putative differences at each position, including substitutions, insertions, and deletions (Victorian Bioinformatics Consortium). Phylogenetic analyses were performed using a distance method, based on pairwise comparisons of conserved SNPs among all strains. Split decomposition analysis was employed using uncorrected pairwise (p) distances with bootstrapping as implemented in SplitsTree4 (28). Published *E. coli* multilocus sequence typing (MLST) primer sequences (www.mlst.net) were used to determine the relevant sequences

from the partially assembled genomes, and the sequences were analyzed using the online MLST database (www.mlst.net).

**Resistance determinants.** No automated method to screen NGS data for a wide range of antimicrobial resistance determinants is freely available. Therefore, we manually explored the partially assembled genomes for signature DNA sequences derived from important Gram-negative resistance determinants (29–43) and plasmid replicon types (44) using the search function in Artemis (8) (see Table S2 in the supplemental material). A list of primers was compiled from the literature with PubMed using the search terms "multiplex PCR," "*Escherichia coli*," and "resistance." Papers with published primer sequences were included. This search was extensive but not exhaustive.

When primer sequences were detected in the interrogated genomes, the sequence between the forward and reverse primers was selected and submitted to a BLAST (Basic Local Alignment Search Tool) nucleotide search on the NCBI database (http://blast.ncbi.nlm.nih.gov/). Results of the BLAST search were used to confirm the presence and exact type of resistance determinants present. In addition to manual searches for primer sequences, a local BLAST database was generated in CLC Genomics Workbench (CLC bio A/S, Denmark) to screen for resistance determinants in the partially assembled genomes.

## RESULTS

**Sequencing and genome assembly results.** There was a 5-day turnaround time from bacterial culture to sequence generation and preparation for analysis. The cost for consumables alone was approximately $300 per strain. Output from the Ion Torrent personal genome sequencer comprised moderate-quality sequencing data for all isolates (Table 1). *De novo* genome assembly and read mapping (using the *E. coli* S88 strain) resulted in 88 to 89% of the reference genome coverage in the outbreak strains and 84% in the comparator strain. Depth of coverage was between 33- and 67-fold (Table 1). The sequence data have been submitted to the NCBI Sequence Read Archive under accession no. SUB142165.

**Epidemiologic analysis.** All four outbreak strains were found to belong to multilocus sequence type (ST) 131, whereas the comparator strain was similar to strains of ST-23. Global SNP analysis revealed 100% homology between the four outbreak isolates, with no SNPs detected, and showed very little homology between those isolates and the local comparator strain (BPH0659) (Fig. 1). This definitively demonstrates that these strains were nosocomially spread within the NICU and that these strains were significantly different from other circulating ESBL *E. coli* strains in our hospital. When the genome sequences of the outbreak strains were compared to all published fully and partially sequenced *E. coli* genomes, they were found to be most closely related to uropathogenic *E. coli* strains (Fig. 1). Of note, the global phylogeny of *E. coli*
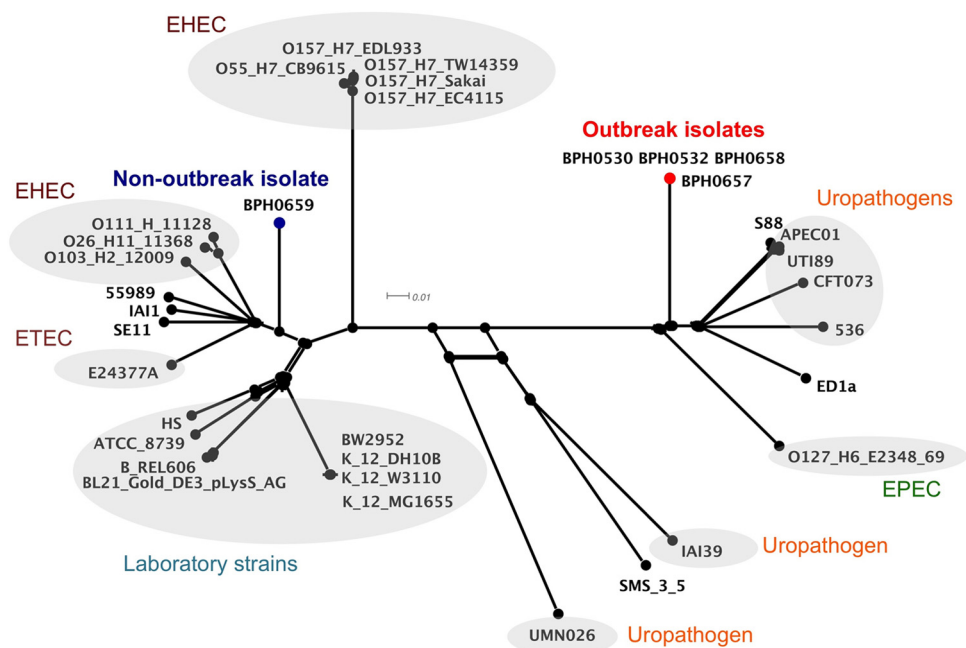
**FIG 1** Phylogenetic analysis of *E. coli* strains. Phylogenetic analysis of the four outbreak *E. coli* strains (BPH0530, BPH0532, BPH0657 [index case], and BPH0658) and local comparator strain (BPH0659) compared to publicly available fully and partially sequenced strains, inferred by split decomposition analysis based on single nucleotide polymorphisms (SNPs). This analysis demonstrates that the four outbreak strains are identical by SNP analysis, cluster with uropathogenic strains of *E. coli* and a neonatal meningitis strain (S88), and differ significantly from the local comparator strain. (See Table S1 in the supplemental material for details about other strains and references). EHEC, enterohemorrhagic *E. coli*; EPEC, enteropathogenic *E. coli*; ETEC, enterotoxigenic *E. coli*.

genome sequences demonstrated clear clustering based on clinical groupings, such as enterohemorrhagic *E. coli* (EHEC) and laboratory strains.

**Resistance determinants.** $bla_{CTX-M-15}$ was detected in the four outbreak strains, which is consistent with the ESBL phenotype, and $bla_{TEM-1}$ was detected as well. In addition, the genes for aminoglycoside resistance (*aadA1*), tetracycline resistance (*tetA*), and low-level trimethoprim resistance (*dfrA1*) were detected. Point mutations associated with quinolone resistance were detected in *gyrA* ($\Delta$S83L and $\Delta$D87N) and *parC* ($\Delta$S80I and $\Delta$E84V) in all four strains. Plasmid type IncFIA was detected in the outbreak isolates.

Despite having the same antibiogram as the outbreak strains, the comparator strain (BPH0659) carried antimicrobial resistance genes that differed significantly from those in the outbreak strains. Although all strains possessed $bla_{CTX-M-15}$, $bla_{TEM-1}$, *tetA*, and *dfrA1*, the comparator strain also harbored $bla_{OXA-1}$, *qnrS1* (plasmid-mediated quinolone resistance determinant), *aacIb-cr*, and *sul1*, which were not present in the outbreak strains.

## DISCUSSION

Here, we have conducted a pilot project to determine if NGS technology might be applied to a clinical infection control question, and we identified the challenges that need to be overcome before application of this technology in a diagnostic microbiology laboratory becomes routine. We have successfully conducted a local outbreak investigation, assessed our strains in the context of worldwide epidemiology, and characterized the resistance genes of our strains using a single test methodology. In our hands, we estimate the sequencing cost to be approximately $300 per strain (excluding analysis), compared to a locally available pulsed-field

gel electrophoresis (PFGE) cost of approximately $150 per strain and the multilocus sequence typing (MLST) cost of approximately $120 per strain. However, with multiplexing, sequencing costs of less than $100 per isolate are possible with platforms such as the MiSeq benchtop sequencer (Illumina Technologies). We also estimate our real-world turnaround time to be as little as 5 days from a positive culture to sequence completion (prior to data analysis, depending on the clinical question being investigated). This time might decrease to as little as 24 h with the advancement of NGS technologies.

Using SNP analysis of the *E. coli* core genome, we have confirmed our four outbreak isolates to be identical, substantiating the hypothesis of a secondary spread of this ESBL *E. coli* strain within the NICU. The isolates have also been identified as ST-131, a successful *E. coli* clone commonly associated with multidrug resistance, particularly due to the presence of $bla_{CTX-M}$ genes and especially $bla_{CTX-M-15}$. *E. coli* ST-131 strains have been recently described as the worldwide pandemic clone, most commonly causing community-onset antimicrobial-resistant infections, particularly urinary tract infections (45). This clone has only recently been reported in our region, and this information adds to other epidemiologic information regarding the spread of this strain in Australia; the rapid recognition of this clone in the NICU further enhanced concern regarding the outbreak.

As our NICU has had a stringent policy of restricted antimicrobial use, this outbreak was the first time that such a resistant enteric Gram-negative organism had been isolated in the unit. Together with the increasing prevalence of these organisms, these cases demonstrate the increasing need to consider maternal risk factors for colonization with resistant *Enterobacteriaceae* (includ-

**TABLE 2** Potential current and future applications of NGS in the diagnostic microbiology laboratory, and limitations to be addressed before widespread NGS implementation

Potential applications of NGS in the diagnostic microbiology laboratory
- Outbreak investigations
- Epidemiological typing
- Characterization of resistance determinants
- Unknown organism identification
- Detection of known virulence factors in clinically severe disease (e.g., presence of toxins)

Current limitations
- Cost of personal genome sequencing platforms
- Speed of data analysis
- Limited user-friendly bioinformatics platforms available for sequence assembly and data analysis
- Education because of poor knowledge of genomics and bioinformatics among diagnostic microbiologists
- Limited reference sequence data for many species
- Difficulties with DNA extraction directly from clinical samples

Future directions
- Routine sequencing of important nosocomial pathogens (e.g., MRSA, VRE, MR-GNB)
- Metagenomics for complex microbial communities (e.g., chronic sinusitis, leg ulcers, altered gut, or vaginal flora)
- Discovery of new pathogens, resistance mechanisms, and virulence factors in clinical isolates

*a* MRSA, methicillin-resistant *Staphylococcus aureus*; VRE, vancomycin-resistant enterococcus; MR-GNB, multiresistant Gram-negative bacteria.

ing medical history, prior antibiotic therapy, and travel history) when managing a septic neonate.

We have also characterized the resistance genes for this outbreak strain, yielding a $bla_{\text{CTX-M-15}}$ ESBL gene (the most common ESBL gene worldwide) (46), as well as fluoroquinolone resistance mutations and the genes encoding aminoglycoside, tetracycline, and low-level trimethoprim resistance. Sequencing has also been used to demonstrate the presence of significantly different antimicrobial resistance genes to discriminate between two strains with the same antibiograms.

As the costs of NGS continue to fall, personal (benchtop) genome sequencers (compact and relatively low-cost platforms suitable for the diagnostic laboratory setting) become more widely available, and turnaround times become more rapid, whole-genome sequencing of pathogens will very soon be within the reach of the routine diagnostic laboratory (2). However, as noted by others, the manipulation and interpretation of data are more likely to be the rate-limiting steps in NGS application than is genome sequencing (2, 47, 48). Although the situation is improving, there are currently few user-friendly bioinformatics software platforms available for use by diagnostic microbiology scientists and doctors, who might not have extensive knowledge of genomics and bioinformatics (47, 49). There is also an urgent need for scientists and clinical microbiologists to increase their understanding of genomics before these technologies can be applied to clinical questions, and before sequencing results and limitations can be accurately conveyed to clinicians (49) (Table 2).

Perhaps the most promising application for NGS technologies is in molecular epidemiology, offering the ultimate in high-resolution genomic epidemiology. It has the potential to offer real-time, portable, digital, and clinically relevant molecular typing of isolates in outbreak investigations, at costs that will very soon approach those of the older, more labor-intensive typing methods (2, 50). However, further studies are required to examine the rates and modes of genetic evolution of different pathogens before large-scale application is available in this area. In the medium term, it is important to ensure that sequencing data are backwards compatible with current typing methods, such as with MLST in the case of *E. coli* (51). There is also a need to collect more sequence data on less-common organisms, which otherwise might be neglected in sequencing studies (2).

Of course, not every multidrug-resistant organism or outbreak will require the use of NGS, especially in the short term before this technology becomes more commonplace in the diagnostic laboratory. However, we have demonstrated here its potential utility in a common clinical scenario and have identified some of the challenges that we face as a community of scientists and clinicians before its widespread implementation. Strong partnerships between experts in the fields of sequencing, genome assembly and annotation, molecular epidemiology, and bioinformatics will be required to create user-friendly, streamlined workflows before NGS can be successfully applied in diagnostic laboratories.

## REFERENCES

1. **Brown TA.** 2002. Genomes, 2nd ed. Wiley-Liss, Oxford, United Kingdom.
2. **Pallen MJ, Loman NJ, Penn CW.** 2010. High-throughput sequencing and clinical microbiology: progress, opportunities and challenges. Curr. Opin. Microbiol. **13**:625–631.
3. **Köser CU, Holden MT, Ellington MJ, Cartwright EJ, Brown NM, Ogilvy-Stuart AL, Hsu LY, Chewapreecha C, Croucher NJ, Harris SR, Sanders M, Enright MC, Dougan G, Bentley SD, Parkhill J, Fraser LJ, Betley JR, Schulz-Trieglaff OB, Smith GP, Peacock SJ.** 2012. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. N. Engl. J. Med. **366**:2267–2275.
4. **Feng J, Lupien A, Gingras H, Wasserscheid J, Dewar K, Légaré D, Ouellette M.** 2009. Genome sequencing of linezolid-resistant *Streptococcus pneumoniae* mutants reveals novel mechanisms of resistance. Genome Res. **19**:1214–1223.
5. **Hornsey M, Loman N, Wareham DW, Ellington MJ, Pallen MJ, Turton JF, Underwood A, Gaulton T, Thomas CP, Doumith M, Livermore DM, Woodford N.** 2011. Whole-genome comparison of two *Acinetobacter baumannii* isolates from a single patient, where resistance developed during tigecycline therapy. J. Antimicrob. Chemother. **66**:1499–1503.
6. **Petrosino JF, Highlander S, Luna RA, Gibbs RA, Versalovic J.** 2009. Metagenomic pyrosequencing and microbial identification. Clin. Chem. **55**:856–866.
7. **Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S, Bidet P, Bingen E, Bonacorsi S, Bouchier C, Bouvet O, Calteau A, Chiapello H, Clermont O, Cruveiller S, Danchin A, Diard M, Dossat C, Karoui ME, Frapy E, Garry L, Ghigo JM, Gilles AM, Johnson J, Le Bouguénec C, Lescat M, Mangenot S, Martinez-Jéhanne V, Matic I, Nassif X, Oztas S, Petit MA, Pichon C, Rouy Z, Ruf CS, Schneider D, Tourret J, Vacherie B, Vallenet D, Médigue C, Rocha EP, Denamur E.** 2009. Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths. PLoS Genet. **5**:e1000344. doi:10.1371/journal.pgen.1000344.
8. **Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B.** 2000. Artemis: sequence visualization and annotation. Bioinformatics **16**:944–945.

9. **Chen SL, Hung CS, Xu J, Reigstad CS, Magrini V, Sabo A, Blasiar D, Bieri T, Meyer RR, Ozersky P, Armstrong JR, Fulton RS, Latreille JP, Spieth J, Hooton TM, Mardis ER, Hultgren SJ, Gordon JI.** 2006. Identification of genes subject to positive selection in uropathogenic strains of *Escherichia coli*: a comparative genomics approach. Proc. Natl. Acad. Sci. U. S. A. **103**:5977–5982.

10. **Durfee T, Nelson R, Baldwin S, Plunkett G 3rd, Burland V, Mau B, Petrosino JF, Qin X, Muzny DM, Ayele M, Gibbs RA, Csörgo B, Pósfai G, Weinstock GM, Blattner FR.** 2008. The complete genome sequence of *Escherichia coli* DH10B: insights into the biology of a laboratory workhorse. J. Bacteriol. **190**:2597–2606.

11. **Eppinger M, Mammel MK, Leclerc JE, Ravel J, Cebula TA.** 2011. Genomic anatomy of *Escherichia coli* O157:H7 outbreaks. Proc. Natl. Acad. Sci. U. S. A. **108**:20142–20147.

12. **Ferenci T, Zhou Z, Betteridge T, Ren Y, Liu Y, Feng L, Reeves PR, Wang L.** 2009. Genomic sequencing reveals regulatory mutations and recombinational events in the widely used MC4100 lineage of *Escherichia coli* K-12. J. Bacteriol. **191**:4025–4029.

13. **Fricke WF, Wright MS, Lindell AH, Harkins DM, Baker-Austin C, Ravel J, Stepanauskas R.** 2008. Insights into the environmental resistance gene pool from the genome sequence of the multidrug-resistant environmental isolate *Escherichia coli* SMS-3-5. J. Bacteriol. **190**:6779–6794.

14. **Hayashi T, Makino K, Ohnishi M, Kurokawa K, Ishii K, Yokoyama K, Han CG, Ohtsubo E, Nakayama K, Murata T, Tanaka M, Tobe T, Iida T, Takami H, Honda T, Sasakawa C, Ogasawara N, Yasunaga T, Kuhara S, Shiba T, Hattori M, Shinagawa H.** 2001. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. DNA Res. **8**:11–22.

15. **Hochhut B, Wilde C, Balling G, Middendorf B, Dobrindt U, Brzuszkiewicz E, Gottschalk G, Carniel E, Hacker J.** 2006. Role of pathogenicity island-associated integrases in the genome plasticity of uropathogenic *Escherichia coli* strain 536. Mol. Microbiol. **61**:584–595.

16. **Iguchi A, Thomson NR, Ogura Y, Saunders D, Ooka T, Henderson IR, Harris D, Asadulghani M, Kurokawa K, Dean P, Kenny B, Quail MA, Thurston S, Dougan G, Hayashi T, Parkhill J, Frankel G.** 2009. Complete genome sequence and comparative genome analysis of enteropathogenic *Escherichia coli* O127:H6 strain E2348/69. J. Bacteriol. **191**:347–354.

17. **Jeong H, Barbe V, Lee CH, Vallenet D, Yu DS, Choi SH, Couloux A, Lee SW, Yoon SH, Cattolico L, Hur CG, Park HS, Ségurens B, Kim SC, Oh TK, Lenski RE, Studier FW, Daegelen P, Kim JF.** 2009. Genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3). J. Mol. Biol. **394**:644–652.

18. **Johnson TJ, Kariyawasam S, Wannemuehler Y, Mangiamele P, Johnson SJ, Doetkott C, Skyberg JA, Lynne AM, Johnson JR, Nolan LK.** 2007. The genome sequence of avian pathogenic *Escherichia coli* strain O1: K1:H7 shares strong similarities with human extraintestinal pathogenic *E. coli* genomes. J. Bacteriol. **189**:3228–3236.

19. **Kulasekara BR, Jacobs M, Zhou Y, Wu Z, Sims E, Saenphimmachak C, Rohmer L, Ritchie JM, Radey M, McKevitt M, Freeman TL, Hayden H, Haugen E, Gillett W, Fong C, Chang J, Beskhlebnaya V, Waldor MK, Samadpour M, Whittam TS, Kaul R, Brittnacher M, Miller SI.** 2009. Analysis of the genome of the *Escherichia coli* O157:H7 2006 spinach-associated outbreak isolate indicates candidate genes that may enhance virulence. Infect. Immun. **77**:3713–3721.

20. **Ogura Y, Ooka T, Iguchi A, Toh H, Asadulghani M, Oshima K, Kodama T, Abe H, Nakayama K, Kurokawa K, Tobe T, Hattori M, Hayashi T.** 2009. Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. **106**:17939–17944.

21. **Oshima K, Toh H, Ogura Y, Sasamoto H, Morita H, Park SH, Ooka T, Iyoda S, Taylor TD, Hayashi T, Itoh K, Hattori M.** 2008. Complete genome sequence and comparative analysis of the wild-type commensal *Escherichia coli* strain SE11 isolated from a healthy adult. DNA Res. **15**: 375–386.

22. **Perna NT, Plunkett G 3rd, Burland V, Mau B, Glasner JD, Rose DJ, Mayhew GF, Evans PS, Gregor J, Kirkpatrick HA, Pósfai G, Hackett J, Klink S, Boutin A, Shao Y, Miller L, Grotbeck EJ, Davis NW, Lim A, Dimalanta ET, Potamousis KD, Apodaca J, Anantharaman TS, Lin J, Yen G, Schwartz DC, Welch RA, Blattner FR.** 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature **409**:529–533.

23. **Rasko DA, Rosovitz MJ, Myers GS, Mongodin EF, Fricke WF, Gajer P, Crabtree J, Sebaihia M, Thomson NR, Chaudhuri R, Henderson IR, Sperandio V, Ravel J.** 2008. The pangenome structure of *Escherichia coli*:

24. comparative genomic analysis of *E. coli* commensal and pathogenic isolates. J. Bacteriol. **190**:6881–6893.

24. **Riley M, Abe T, Arnaud MB, Berlyn MK, Blattner FR, Chaudhuri RR, Glasner JD, Horiuchi T, Keseler IM, Kosuge T, Mori H, Perna NT, GPlunkett 3rd, Rudd KE, Serres MH, Thomas GH, Thomson NR, Wishart D, Wanner BL.** 2006. *Escherichia coli* K-12: a cooperatively developed annotation snapshot–2005. Nucleic Acids Res. **34**:1–9.

25. **Welch RA, Burland V, Plunkett G III, Redford P, Roesch P, Rasko D, Buckles EL, Liou SR, Boutin A, Hackett J, Stroud D, Mayhew GF, Rose DJ, Zhou S, Schwartz DC, Perna NT, Mobley HL, Donnenberg MS, Blattner FR.** 2002. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. **99**:17020–17024.

26. **Zhou Z, Li X, Liu B, Beutin L, Xu J, Ren Y, Feng L, Lan R, Reeves PR, Wang L.** 2010. Derivation of *Escherichia coli* O157:H7 from its O55:H7 precursor. PLoS One **5**:e8700. doi:10.1371/journal.pone.0008700.

27. **Rumble SM, Lacroute P, Dalca AV, Fiume M, Sidow A, Brudno M.** 2009. SHRiMP: accurate mapping of short color-space reads. PLoS Comp. Biol. **5**:e1000386. doi:10.1371/journal.pcbi.1000386.

28. **Huson DH, Bryant D.** 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. **23**:254–267.

29. **Boyd DA, Tyler S, Christianson S, McGeer A, Muller MP, Willey BM, Bryce E, Gardam M, Nordmann P, Mulvey MR.** 2004. Complete nucleotide sequence of a 92-kilobase plasmid harboring the CTX-M-15 extended-spectrum beta-lactamase involved in an outbreak in long-term-care facilities in Toronto, Canada. Antimicrob. Agents Chemother. **48**:3758–3764.

30. **Dallenne C, Da Costa A, Decré D, Favier C, Arlet G.** 2010. Development of a set of multiplex PCR assays for the detection of genes encoding important beta-lactamases in Enterobacteriaceae. J. Antimicrob. Chemother. **65**:490–495.

31. **Ellem J, Partridge SR, Iredell JR.** 2011. Efficient direct ESBL detection by multiplex real-time PCR: accurate assignment of phenotype by use of a limited set of genetic markers. J. Clin. Microbiol. **49**:3074–3077.

32. **Ellington MJ, Kistler J, Livermore DM, Woodford N.** 2007. Multiplex PCR for rapid detection of genes encoding acquired metallo-beta-lactamases. J. Antimicrob. Chemother. **59**:321–322.

33. **Espedido BA, Partridge SR, Iredell JR.** 2008. *bla*(IMP-4) in different genetic contexts in Enterobacteriaceae isolates from Australia. Antimicrob. Agents Chemother. **52**:2984–2987.

34. **Grape M, Motakefi A, Pavuluri S, Kahlmeter G.** 2007. Standard and real-time multiplex PCR methods for detection of trimethoprim resistance dfr genes in large collections of bacteria. Clin. Microbiol. Infect. **13**:1112–1118.

35. **Kim HB, Park CH, Kim CJ, Kim EC, Jacoby GA, Hooper DC.** 2009. Prevalence of plasmid-mediated quinolone resistance determinants over a 9-year period. Antimicrob. Agents Chemother. **53**:639–645.

36. **Minarini LA, Poirel L, Cattoir V, Darini AL, Nordmann P.** 2008. Plasmid-mediated quinolone resistance determinants among enterobacterial isolates from outpatients in Brazil. J. Antimicrob. Chemother. **62**: 474–478.

37. **Monstein HJ, Ostholm-Balkhed A, Nilsson MV, Nilsson M, Dornbusch K, Nilsson LE.** 2007. Multiplex PCR amplification assay for the detection of *bla*SHV, *bla*TEM and *bla*CTX-M genes in Enterobacteriaceae. APMIS **115**:1400–1408.

38. **Paterson DL, Hujer KM, Hujer AM, Yeiser B, Bonomo MD, Rice LB, Bonomo RA, International Klebsiella Study Group.** 2003. Extended-spectrum beta-lactamases in *Klebsiella pneumoniae* bloodstream isolates from seven countries: dominance and widespread prevalence of SHV- and CTX-M-type beta-lactamases. Antimicrob. Agents Chemother. **47**:3554–3560.

39. **Pazhani GP, Chakraborty S, Fujihara K, Yamasaki S, Ghosh A, Nair GB, Ramamurthy T.** 2011. QRDR mutations, efflux system & antimicrobial resistance genes in enterotoxigenic *Escherichia coli* isolated from an outbreak of diarrhoea in Ahmedabad, India. Indian J. Med. Res. **134**:214–223.

40. **Pérez-Pérez FJ, Hanson ND.** 2002. Detection of plasmid-mediated AmpC beta-lactamase genes in clinical isolates by using multiplex PCR. J. Clin. Microbiol. **40**:2153–2162.

41. **Sáenz Y, Briñas L, Domínguez E, Ruiz J, Zarazaga M, Vila J, Torres C.** 2004. Mechanisms of resistance in multiple-antibiotic-resistant Escherichia coli strains of human, animal, and food origins. Antimicrob. Agents Chemother. **48**:3996–4001.

42. **Shin SY, Kwon KC, Park JW, Song JH, Ko YH, Sung JY, Shin HW, Koo**

**SH.** 2009. Characteristics of aac(6′)-Ib-cr gene in extended-spectrum beta-lactamase-producing *Escherichia coli* and *Klebsiella pneumoniae* isolated from Chungnam area. Korean J. Lab. Med. **29**:541–550.

43. **Hannecart-Pokorni E, Depuydt F, de wit L, van Bossuyt E, Content J, Vanhoof R.** 1997. Characterization of the 6′-*N*-aminoglycoside acetyl-transferase gene aac(6′)-Im [corrected] associated with a sulI-type inte-gron. Antimicrob. Agents Chemother. **41**:314–318.

44. **Carattoli A, Bertini A, Villa L, Falbo V, Hopkins KL, Threlfall EJ.** 2005. Identification of plasmids by PCR-based replicon typing. J. Microbiol. Methods **63**:219–228.

45. **Rogers BA, Sidjabat HE, Paterson DL.** 2011. *Escherichia coli* O25b-ST131: a pandemic, multiresistant, community-associated strain. J. Anti-microb. Chemother. **66**:1–14.

46. **Peirano G, Pitout JD.** 2010. Molecular epidemiology of *Escherichia coli* producing CTX-M beta-lactamases: the worldwide emergence of clone ST131 O25:H4. Int. J. Antimicrob. Agents **35**:316–321.

47. **Kumar K, Desai V, Cheng L, Khitrov M, Grover D, Satya RV, Yu C,**

**Zavaljevski N, Reifman J.** 2011. AGeS: a software system for microbial genome sequence annotation. PLoS One **6**:e17469. doi:10.1371/journal.pone.0017469.

48. **Su Z, Ning B, Fang H, Hong H, Perkins R, Tong W, Shi L.** 2011. Next-generation sequencing and its applications in molecular diagnostics. Expert Rev. Mol. Diagn. **11**:333–343.

49. **Tan TW, Lim SJ, Khan AM, Ranganathan S.** 2009. A proposed mini-mum skill set for university graduates to meet the informatics needs and challenges of the "-omics" era. BMC Genomics **10**(Suppl 3):S36. doi:10.1186/1471-2164-10-S3-S36.

50. **Hyytiä-Trees EK, Cooper K, Ribot EM, Gerner-Smidt P.** 2007. Recent developments and future prospects in subtyping of foodborne bacterial pathogens. Future Microbiol. **2**:175–185.

51. **Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, Jelsbak L, Sicheritz-Pontén T, Ussery DW, Aarestrup FM, Lund O.** 2012. Multilocus sequence typing of total-genome-sequenced bacteria. J. Clin. Microbiol. **50**:1355–1361.