



GENOME WATCH

Demystifying *Escherichia coli* pathovars

Kate S. Baker

This month's Genome Watch compares and contrasts two studies that used a common genomic approach to understand different pathogenic forms of *Escherichia coli*.

Escherichia coli is a Gram-negative bacterium that can be a harmless commensal in the intestine or cause a variety of pathological infections. The diverse disease types caused by *E. coli* have resulted in their classification into distinct pathovars such as enterohaemorrhagic *E. coli* (EHEC), uropathogenic *E. coli* (UPEC) and enteroinvasive *E. coli* (EIEC)¹. The evolutionary relationships between and within these different pathovars can elucidate how they have emerged and how they behave epidemiologically. Furthermore, understanding the molecular mechanisms underpinning the pathogenesis of the different pathovars helps to predict clinical phenotypes and in disease control. In two new studies, researchers sequenced whole genomes of clinical isolates on the Illumina HiSeq 2000 platform to determine the population structure of *E. coli* pathovars and to investigate genes important for their pathogenic profiles.

Salipante *et al.*² sequenced more than 300 clinical isolates of extraintestinal pathogenic *E. coli* (ExPEC), which are an important cause of sporadic infectious pathology outside of the intestine. Isolates from urinary tract ($n=277$) and bloodstream ($n=92$) infections, including serial sampling of individuals, were collected over 3 years from a single North American hospital system and the population structure was defined by mapping the sequences from the clinical isolates to an *E. coli* reference genome. This showed that serial samples from urinary tract infections in a single patient were often genetically distinct, whereas sequential

blood infection isolates were typically almost identical. However, no phylogenetic clustering of isolates according to the site of infection was observed, suggesting that specific ExPEC lineages are not restricted to the bloodstream or the urinary tract. Nonetheless, several genes encoding virulence factors correlated with blood infection, including genes involved in invasion (*traJ*) and adhesion (*papA* and *papG*), and the toxin-encoding genes *sat* and *tosA*. The authors also did a gene association study to identify genes responsible for antimicrobial resistance phenotypes and, although known antimicrobial resistance genes were recovered, no novel resistance genes were identified.

von Mentzer *et al.*³ sequenced 362 clinical isolates of globally sourced enterotoxigenic *E. coli* (ETEC) collected between 1980 and 2011. ETEC causes nearly 400,000 deaths each year in low-income nations and is characterized by the production of enterotoxins and other colonization factors, which are mostly located on plasmids that are usually horizontally transmitted and thought to be sufficient for the ETEC phenotype^{3,4}. The authors determined the phylogenetic relationships of the collection (based on 1,429 gene sequences) and found that ETEC lineages were interspersed among other *E. coli* reference genomes. Using Bayesian clustering, the authors defined 21 ETEC lineages, which frequently contained strains from disparate global regions, indicating the presence of multiple distinct, globally distributed ETEC lineages. By doing gene complement analysis, the authors also showed that most of the lineages had stable colonization factor and toxin profiles, suggesting that these factors may be less mobile than previously thought.

Although isolates from distinct pathovars collected with different sampling strategies were investigated

in these studies, a similar genomic approach was applied. Both studies determined the phylogenetic population structure and correlated this with isolate metadata. In this way, the ETEC analysis identified recently evolved lineages that were globally disseminated, and the ExPEC study determined that there was no correlation of phylogeny with infection site. Similarly, both studies examined specific genes that are important for determining the disease phenotype and correlated this with phylogeny. In doing so, the ExPEC study determined that although virulence factors were highly correlated with population structure, some were independently associated with the site of infection, and the ETEC analysis revealed lineage-specific colonization factor and enterotoxin profiles. Additionally, both studies explored other relevant gene content in the data, including antibiotic resistance determinants in the ExPEC study and plasmids in the ETEC analysis. In each case, a genomic approach furthered the understanding of important disease pathovars and demonstrated the importance of studying both vertically and horizontally transmitted genetic traits in pathogenic bacteria.

Kate S. Baker is at the Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK.

e-mail: microbes@sanger.co.uk

doi:10.1038/nrmicro3411

Published online 8 December 2014

1. Croxen, M. A. & Finlay, B. B. Molecular mechanism of *Escherichia coli* pathogenicity. *Nature Reviews Microbiol.* **8**, 26–38 (2010).
2. Salipante, S. J. *et al.* Large-scale genomic sequencing of extraintestinal pathogenic *Escherichia coli* strains. *Genome Research* <http://dx.doi.org/10.1101/gr.180190.114> (2014).
3. von Mentzer, A. *et al.* Identification of enterotoxigenic *Escherichia coli* (ETEC) clades with long-term global distribution. *Nature Genet.* <http://dx.doi.org/10.1038/ng.3145> (2014).
4. Turner, S. M. *et al.* Phylogenetic comparisons reveal multiple acquisitions of the toxin genes by enterotoxigenic *Escherichia coli* strains of different evolutionary lineages. *J. Clin. Microbiol.* **44**, 4528–4536 (2006).

Competing interests statement

The authors declare no competing interests.

