

GENOME WATCH

EPEC: a cocktail of virulence

Teemu Kallonen and Christine J. Boinett

Genomics studies are prompting a re-evaluation of the diversity of *Escherichia coli* pathovars and how this diversity corresponds to virulence.

Enteropathogenic *Escherichia coli* (EPEC) can cause moderate-to-severe diarrhoea, which can sometimes be fatal, and is characterized by the presence of the locus of enterocyte effacement (LEE) pathogenicity island, which encodes a type III secretion system (T3SS) and various cognate effectors. The EPEC pathovar is generally distinguished from enterohaemorrhagic *E. coli* (EHEC) by the absence of shiga toxin (*stx*) and can itself be subdivided into typical EPEC (tEPEC), which has the plasmid-encoded bundle-forming pilus (BFP), and atypical EPEC (aEPEC), which does not. However, two new genomics studies suggest that the diversity of EPEC lineages and the genetic basis of virulence are more complex than is indicated by the tEPEC–aEPEC classification.

Ingle *et al.*¹ sequenced 196 aEPEC isolates from the Global Enteric Multicenter Study² (GEMS) using the Illumina HiSeq platform and analysed these data together with sequences from an *E. coli* reference panel. They identified ten common clonal

groups that accounted for 71% of the isolates; these isolates were widely distributed both geographically and throughout the *E. coli* phylogeny. By analysing the phylogeny of LEE in the context of *de novo*-assembled genomes, the authors found that LEE had inserted into the genome only once in each clonal group, at one of three conserved sites. The LEE phylogeny enabled the classification of the aEPEC isolates into three major lineages. The lineages could be further divided into 30 subtypes, each with a distinct set of non-LEE effectors that are encoded elsewhere in the genome. Four of the clonal groups had isolates with *bfpA* or *stx*, which indicates that some aEPEC lineages may have acquired the BFP-encoding plasmid and/or *stx*. The varied composition of non-LEE virulence genes may therefore influence the virulence of EPEC lineages.

In a second study, Hazen *et al.*³ sequenced 70 EPEC isolates from the GEMS using the Illumina HiSeq platform, including isolates associated with 24 lethal infections, 23 non-lethal but symptomatic infections and 23 asymptomatic infections. Phylogenetic analysis of these isolates together with previously sequenced *E. coli* and *Shigella* spp. isolates revealed a greater diversity within the GEMS isolates than within prototype *E. coli* strains. Using an established multi-locus sequence typing-based phylogeny of six lineages (EPEC1–6), the authors discovered that only 27 of the GEMS isolates clustered in these lineages. They proposed an additional four lineages (EPEC7–10) to describe 29 of the remaining isolates. However, 14 isolates, mostly from asymptomatic infections, did not cluster in any lineage, and almost half of these isolates lacked *bfpA*. Further analysis of this gene revealed no lineage specificity, which suggests that *bfpA* (and probably the plasmid on which it is encoded) may be subject to several independent loss and acquisition events in different

lineages, as was also suggested by the findings of Ingle *et al.*¹.

Using a comparative genomic analysis known as a large-scale BLAST score ratio⁴ analysis, a core genome composed of 1,080 gene clusters was analysed alongside clinical outcomes. Although some gene clusters were more prevalent in lethal and/or symptomatic infections, none of these gene clusters was associated solely with a single clinical outcome, which may suggest that no single gene or gene cluster is responsible for the severity of the clinical outcome. The authors propose that the clinical outcome may instead rely on the concerted function of several genomic regions and/or host factors, which may explain how EPEC from diverse lineages can produce similar clinical outcomes to one another.

Together, these two detailed genomic analyses highlight a possible oversimplification in the traditional division of the EPEC pathovar into aEPEC and tEPEC based on the presence or absence of BFP. Furthermore the analyses suggest that classifying EPEC as distinct from other *E. coli* pathovars on the basis of whether LEE is present in the genome may similarly be an oversimplification.

Teemu Kallonen and Christine J. Boinett are at the Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK.
e-mail: microbes@sanger.ac.uk

doi:10.1038/nrmicro.2016.33
Published online 7 March 2016

1. Ingle, D. J. *et al.* Evolution of atypical enteropathogenic *E. coli* by repeated acquisition of LEE pathogenicity island variants. *Nat. Microbiol.* **1**, 15010 (2016).
2. Kotloff, K. L. *et al.* The Global Enteric Multicenter Study (GEMS) of diarrheal disease in infants and young children in developing countries: epidemiologic and clinical methods of the case/control study. *Clin. Infect. Dis.* **55**, S232–S245 (2012).
3. Hazen, T. H. *et al.* Genomic diversity of EPEC associated with clinical presentations of differing severity. *Nat. Microbiol.* **1**, 15014 (2016).
4. Sahl, J. W. *et al.* The large-scale blast score ratio (LS-BSR) pipeline: a method to rapidly compare genetic content between bacterial genomes. *PeerJ* **2**, e332 (2014).

Competing interests statement
The authors declare no competing interests.



NPG