

EDITORIAL

***Shigella* and *Escherichia coli* at the crossroads: machiavellian masqueraders or taxonomic treachery?**

Shigellae cause an estimated 150 million cases and 600 000 deaths annually, and can cause disease after ingestion of as few as 10 bacterial cells [1]. They are spread by the faecal–oral route, with food, water, fomites, insects (especially flies) and direct person-to-person contact. *S. dysenteriae* causes brisk and deadly epidemics, particularly in the developing world; *S. flexneri* and *S. sonnei* account for the endemic form of the disease, particularly in industrialised nations; *S. boydii* is rarely encountered [1, 2].

Shigellosis is a locally invasive colitis in which bacteria invade and proliferate within colonocytes and mucosal macrophages, trigger apoptosis of macrophages and spread through the mucosa from cell to cell [1]. Cytokines produced by epithelial cells and macrophages in response to invading bacteria stimulate an intense acute inflammatory response. This limits the extent of bacterial invasion, but also contributes to mucosal injury [1]. Despite their local invasiveness, shigellae rarely cause bacteraemia. However, shigellosis occasionally precipitates the haemolytic-uraemic syndrome (HUS), which is characterised by vaso-occlusive renal failure, consumptive thrombocytopenia, micro-angiopathic haemolytic anaemia and neurological dysfunction. These manifestations result in part from the actions of Shiga toxin, the distinctive cytotoxin produced by *S. dysenteriae* type 1, on neurons and on renal and cerebral endothelial cells [1, 2].

Ever since the initial discovery of the dysentery bacillus (now *S. dysenteriae*) by Kiyoshi Shiga in 1898 [3], *Shigella* spp. have generally been regarded on both microbiological and clinical grounds as distinct from *Escherichia coli*. Whereas most *E. coli* are motile, ferment lactose, are lysine decarboxylase- and indole-positive, and are mostly avirulent, shigellae are non-motile, are typically lactose-, lysine decarboxylase- and indole-negative, and are obligate pathogens [1, 4]. Consequently, the emergence of *E. coli* O157:H7 and other Shiga toxin-producing *E. coli* strains that, like shigellae, cause food- and water-borne outbreaks of haemorrhagic colitis, a dysentery-like illness sometimes complicated by HUS [5, 6], has prompted a reconsideration of the relationship between these

genera. One authority has even proposed that enterohaemorrhagic *E. coli* (EHEC) such as *E. coli* O157:H7 are essentially ‘*Shigella* in a cloak of *E. coli* antigens’ [7].

Shigella-like strains of *E. coli* that cause an invasive, dysenteric diarrhoeal illness were first described in 1971, over a decade before the appearance in 1982 of the new EHEC strains that launched the current wave of interest in the *E. coli*–*Shigella* connection [8]. Termed ‘enteroinvasive *E. coli*’ (EIEC), these strains, like shigellae, were able to invade and proliferate within intestinal epithelial cells, eventually causing cell death [4, 8]. EIEC share with shigellae a *c.* 140-MDa plasmid (pINV) that encodes several outer-membrane proteins involved in invasion of host cells [4, 8]. Like shigellae, EIEC are usually non-motile and lactose-negative. They are also usually lysine decarboxylase-negative, a characteristic which, in both EIEC and shigellae, is due to large deletions (‘black holes’) in the genome that correspond to chromosomal regions present in *E. coli* K-12. The deleted genes encode enzymes involved in metabolic reactions that produce compounds such as cadaverine that interfere with uptake of shigellae and EIEC by host cells. Hence, the deletions may enhance the ability to invade host cells [4]. EIEC and shigellae also exhibit considerable antigenic cross-reactivity [4, 8]. However, EIEC do not produce Shiga toxins and are not known to cause HUS. Perhaps because of this, or because they are not known to have caused large-scale outbreaks, EIEC have received considerably less public health and media attention than the now-notorious EHEC [7].

EHEC and shigellae exhibit a number of clinical and pathogenetic differences. The prototypical EHEC-associated syndrome is haemorrhagic colitis—passage of large-volume, grossly bloody stools without obvious pus or mucus, accompanied by abdominal pain, but not tenesmus or fever [5, 6]. This differs from classical dysentery, and probably reflects the pancolitis of EHEC, in contrast to the more focal proctocolitis of shigellosis and EIEC infection, and the toxigenic but non-invasive nature of EHEC disease, in contrast to the mucosal invasion of shigellosis. EHEC share with *S. dysenteriae* type 1 the production of bacteriophage-

encoded Shiga toxin(s) and expression of a type III secretion apparatus [1, 9]. However, in contrast to shigellae, EHEC produce a unique enterohaemolysin and exhibit 'attaching and effacing' adherence to enterocytes similar to that exhibited by enteropathogenic *E. coli* (EPEC) [9]. Furthermore, whereas in shigellae the type III secretion apparatus is encoded on the 140-MDa virulence plasmid [1], in both EHEC and EPEC it is encoded on a chromosomal pathogenicity-associated island termed 'locus of enterocyte effacement'. Moreover, the virulence plasmid of EHEC is distinct from that of *Shigella* spp. and EIEC (as well as from that of EPEC) [8–10]. Thus, EHEC, EIEC and EPEC all share with one another, and with shigellae, one or more important clinical or pathogenetic features. These perplexing similarities and differences are best understood in the context of the population structure of *E. coli* and *Shigella* spp.

The first direct genetic insights into the evolutionary relationships between *Shigella* spp. and *E. coli* were provided by DNA–DNA re-association studies which showed that shigellae exhibit >75% nucleotide similarity to *E. coli*—within the range observed among different *E. coli* strains [11]. Although the authors of this report proposed only that shigellae could be part of the genus *Escherichia*, the data actually justify placing *Shigella* spp. within *E. coli* proper.

Comparative DNA sequence analysis and multilocus enzyme electrophoresis (MLEE) subsequently showed unequivocally that, from a population genetics standpoint, shigellae were better regarded as pathotypes or clones of *E. coli* [12–16]. These data showed that the four '*Shigella*' species constituted two related clusters within the larger *E. coli* population, one comprising representatives of three of the four putative *Shigella* species, the other comprising representatives of all four species [12, 16]. Within each cluster, different *Shigella* species intermingled in an overlapping fashion in adjacent lineages, but in some instances were actually placed together in the same lineage [16]. This indicated that the conventional nomenclature for *Shigella* spp. could be justified only on historical and clinical grounds [13].

Over the past decade, understanding of the population structure of *E. coli* has been refined [17, 18]. The precise position of *Shigella* strains within the *E. coli* phylogenetic tree was clarified by MLEE analysis of diverse strains of shigellae, EHEC, EIEC, EPEC and enterotoxigenic *E. coli* (ETEC), in comparison with control strains from an *E. coli* reference collection [15]. Shigellae were found to occur primarily within *E. coli* phylogenetic group A, in a single cluster that also included one of the non-pathogenic reference strains. *S. sonnei* strains occupied their own (single) lineage, whereas strains of *S. flexneri*, *S. boydii* and *S. dysenteriae* were intermingled in a subcluster adjacent to the *S. sonnei* lineage. Other *S. flexneri* strains were

assigned to two adjacent lineages within the (distant) phylogenetic group D. These data confirmed the genetic diversity within *Shigella* spp., the intermingling of *Shigella* spp. and *E. coli*, and discrepancies between conventional *Shigella* 'species' designations and observed genetic relationships. They also localised *Shigella* strains specifically to *E. coli* phylogenetic groups A and D [15]. Sequence analysis of the housekeeping gene *mdh* and ribotyping yielded largely similar results, but showed somewhat different distributions of the *Shigella* spp. among the *E. coli* phylogenetic groups [15, 19].

The data of Pupo *et al.* [15] showed that EIEC, EPEC and ETEC were distributed across multiple *E. coli* lineages within phylogenetic groups A and B1, with little clustering according to pathogenic type. No diarrhoeal strains were assigned to phylogenetic group B2, which is the source of most *E. coli* strains that cause urinary tract infections, bacteraemia and neonatal meningitis [20]. Of the diarrhoeagenic types, only EHEC (of which all representatives studied were *E. coli* O157:H7) appeared among the non-aligned lineages of reference strains, where they constituted a single cluster [15]. These data contradicted the hypothesis that O157:H7 EHEC are derived from either shigellae or EIEC, and revealed that EIEC, EPEC and ETEC have evolved independently several times in different *E. coli* lineages, presumably through horizontal acquisition of the requisite virulence traits [15].

If O157:H7 EHEC did not evolve from shigellae or EIEC, they may have arisen instead from a pre-existing lineage of O157 antigen-positive non-EHEC through horizontal acquisition of EHEC-specific virulence genes. However, this hypothesis was, in turn, contradicted by analysis of the population structure of O157 antigen-positive *E. coli* strains representing diverse H types, clinical sources and host species, which showed that O157:H7 EHEC occupy their own lineage, far removed from non-EHEC, non-H7 O157 strains [21]. Thus, O157:H7 EHEC appeared to have acquired the O157 antigen by horizontal transfer from an unknown distant donor rather than by vertical inheritance from a clonal ancestor [21].

The ancestral clonal group within which O157:H7 EHEC did evolve was revealed by MLEE data showing that the nearest neighbours of O157:H7 EHEC are EPEC strains of serotype O55:H7 [22]. The O157:H7 EHEC and O55:H7 EPEC strains constitute a cluster that is quite distant from lineages containing other diarrhoeagenic *E. coli* types, including non-H7 EPEC strains of serogroup O55 [22]. The phylogenetic similarity between O157:H7 EHEC and O55:H7 EPEC, together with the observation that such strains also have in common the 'locus of enterocyte enhancement' pathogenicity island, suggests that EHEC O157:H7 probably evolved from an O55:H7 EPEC ancestor by

acquiring, at least: the O157 antigen in place of the O55 antigen; a Shiga toxin-encoding transforming bacteriophage (present in *S. dysenteriae* type 1); and the EHEC enterohaemolysin virulence plasmid [23]. This suggests that the occurrence in remote evolutionary lineages of EHEC strains with O:H serotypes other than O157:H7 [9] probably reflects the horizontal transfer of multiple virulence determinants from O157:H7 EHEC (or other undefined donors) into distantly related recipients, with consequent conversion of previously non-O157 strains into non-O157 EHEC.

Thus, although there are similarities between EHEC and *Shigella* spp. with respect to clinical behaviour, virulence mechanisms and phylogenetic background, certain other *E. coli* strains (e.g., EIEC) are, in many respects, more similar to shigellae than are EHEC. Furthermore, valid comparisons between *Shigella* spp. and *E. coli* require an understanding that shigellae are actually pathotypes of *E. coli*. However, this understanding is not yet generally prevalent, as shown by the suggestion that EHEC are like shigellae masquerading as *E. coli* [7], and by the surprise generated by the similar ribosomal DNA polymorphisms of *Shigella* spp. and of certain *E. coli* strains [24].

Biotypes and clinical behaviour are unreliable as indicators of phylogenetic relationships, as horizontal transfer of virulence genes can convert distantly related organisms into pathogens that behave in a similar fashion. The same process (as well as gene deletions) can create striking and potentially misleading phenotypic differences between members of a clone or species that are otherwise similar. Clinicians and microbiologists should avoid the mental straitjacket of traditional, phenotypically derived taxonomies and the seduction of drawing false phylogenetic conclusions from observed phenotypes and clinical behaviours.

This work was supported in part by VA Merit Review and National Institutes of Health grant DK-47504. Ann Emery helped to prepare the manuscript.

JAMES R. JOHNSON

VA Medical Center and Department of Medicine, University of Minnesota, Minneapolis, Minnesota MN 55417, USA
(e-mail: johns007@tc.umn.edu)

References

1. Sansonetti PI. *Shigella* plays dangerous games. *ASM News* 1999; **65**: 611–617.
2. Keusch GT. The rediscovery of Shiga toxin and its role in human illness. *Infect Dis Clin Pract* 1997; **6**: 533–536.
3. Trofa AF, Ueno-Olsen H, Oiwa R, Yoshikawa M. Dr. Kiyoshi Shiga: discoverer of the dysentery bacillus. *Clin Infect Dis* 1999; **29**: 1303–1306.
4. Maurelli AT, Fernández RE, Bloch CA, Rode CK, Fasano A. "Black holes" and bacterial pathogenicity: a large genomic

- deletion that enhances the virulence of *Shigella* spp. and enteroinvasive *Escherichia coli*. *Proc Natl Acad Sci USA* 1998; **95**: 3943–3948.
5. Boyce TG, Swerdlow DL, Griffin PM. *Escherichia coli* O157:H7 and the hemolytic-uremic syndrome. *N Engl J Med* 1995; **333**: 364–368.
6. Slutsker L, Ries AA, Greene KD, Wells JG, Hutwagner L, Griffin PM. *Escherichia coli* O157:H7 diarrhea in the United States: clinical and epidemiologic features. *Ann Intern Med* 1997; **126**: 505–513.
7. Lederberg J. Parasites face a perpetual dilemma. *ASM News* 1999; **65**: 77–80.
8. Levine MM. *Escherichia coli* that cause diarrhea: enterotoxigenic, enteropathogenic, enteroinvasive, enterohemorrhagic, and enteroadherent. *J Infect Dis* 1987; **155**: 377–389.
9. Schmidt H, Geitz C, Tarr PI, Frosch M, Karch H. Non-O157:H7 pathogenic Shiga toxin-producing *Escherichia coli*: phenotypic and genetic profiling of virulence traits and evidence for clonality. *J Infect Dis* 1999; **179**: 115–123.
10. Fasano A, Kay BA, Russell DRJ, Maneval MM. Enterotoxin and cytotoxin production by enteroinvasive *Escherichia coli*. *Infect Immun* 1990; **58**: 3717–3723.
11. Brenner DJ, Fanning GR, Skerman FJ, Falkow S. Polynucleotide sequence divergence among strains of *Escherichia coli* and closely related organisms. *J Bacteriol* 1972; **109**: 953–965.
12. Whittam TS, Ochman H, Selander RK. Multilocus genetic structure in natural populations of *Escherichia coli*. *Proc Natl Acad Sci USA* 1983; **80**: 1751–1755.
13. Selander RK, Caugant DA, Whittam TS. Genetic structure and variation in natural populations of *Escherichia coli*. In: Neidhardt FC, Ingraham KL, Magasanik B, Low KB, Schaechter M, Umberger HE (eds) *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol 2. Washington, DC, American Society for Microbiology. 1987: 1625–1648.
14. Hartl DL, Dykhuizen DE. The population genetics of *Escherichia coli*. *Annu Rev Genet* 1984; **18**: 31–68.
15. Pupo GM, Karaolis DKR, Lan R, Reeves PR. Evolutionary relationships among pathogenic and nonpathogenic *Escherichia coli* strains inferred from multilocus enzyme electrophoresis and *mdh* sequence studies. *Infect Immun* 1997; **65**: 2685–2692.
16. Ochman H, Whittam TS, Caugant DA, Selander RK. Enzyme polymorphism and genetic population structure in *Escherichia coli* and *Shigella*. *J Gen Microbiol* 1983; **129**: 2715–2726.
17. Lecointre G, Rachdi L, Darlu P, Denamur E. *Escherichia coli* molecular phylogeny using the incongruence length difference test. *Mol Biol Evol* 1998; **15**: 1685–1695.
18. Herzer PJ, Inouye S, Inouye M, Whittam TS. Phylogenetic distribution of branched RNA-linked multicopy single-stranded DNA among natural isolates of *Escherichia coli*. *J Bacteriol* 1990; **172**: 6175–6181.
19. Rolland K, Lambert-Zechovsky N, Picard B, Denamur E. *Shigella* and enteroinvasive *Escherichia coli* strains are derived from distinct ancestral strains of *E. coli*. *Microbiology* 1998; **144**: 2667–2672.
20. Picard B, Sevali Garcia J, Gouriou S *et al.* The link between phylogeny and virulence in *Escherichia coli* extraintestinal infection. *Infect Immun* 1999; **67**: 546–553.
21. Whittam TS, Wilson RA. Genetic relationships among pathogenic *Escherichia coli* of serogroup O157. *Infect Immun* 1988; **56**: 2467–2473.
22. Wang G, Whittam TS, Berg CM, Berg DE. RAPD (arbitrary primer) PCR is more sensitive than multilocus enzyme electrophoresis for distinguishing related bacterial strains. *Nucleic Acids Res* 1993; **21**: 5930–5933.
23. Whittam TS. Evolution of *Escherichia coli* O157:H7 and other Shiga toxin-producing *E. coli* strains. In: Kaper JB, O'Brien AD (eds) *Escherichia coli* O157:H7 and other Shiga toxin-producing *E. coli* strains. Washington, DC, American Society for Microbiology. 1998: 195–209.
24. Widjoatmodjo MN, Fluit AC, Verhoef J. Molecular identification of bacteria by fluorescence-based PCR-single-strand conformation polymorphism analysis of the 16S rRNA gene. *J Clin Microbiol* 1995; **33**: 2601–2606.