Pathogen evolution: **How good bacteria go bad**

Craig Stephens* and William Murray†

Recent findings suggest that dysentry-causing *Shigella* strains have arisen several times from *Escherichia coli* via plasmid acquisition and phenotypic convergence. Similarly, three *Bacillus* strains with distinct pathogenic properties are derivatives of a single species whose behavior is profoundly altered by acquired plasmids.

Addresses*Biology Department, Santa Clara University, 500 El Camino Real, Santa Clara, California 95053, USA. †Department of Biological Sciences, San Jose State University, San Jose, California 95192-0100, USA.
E-mail: cstephens@scu.edu

Current Biology 2001, 11:R53–R56

0960-9822/01/$ – see front matter

Most microorganisms we encounter on a daily basis coexist peacefully with humans. Unfortunately, some of these same microbes have siblings that are notorious for their ability to cause disease. The bacterium *Escherichia coli* is a classic example. It is ordinarily found in our intestines as a harmless commensal, but pathogenic *E. coli* strains can cause a wide spectrum of human diseases, ranging from the simple inconvenience of watery diarrhea to the life-threatening hemolytic-uremic syndrome linked to *E. coli* serotype O157:H7. What can make well-behaved microbes turn nasty?

In many cases, the crucial differences between harmless microbes and pathogens are a few genes encoded on mobile elements such as plasmids, transposons or integrated phage genomes [1]. But if one wants to understand how pathogens evolve, analysis of these mobile genetic elements, though important, is not sufficient. Indeed, the sequences of mobile genes reflect their own evolutionary history, not necessarily that of the cell they reside in. Sequences of highly conserved ribosomal RNA genes are typically used for assessing relationships among microbes across broad evolutionary distances. For examining closely related bacteria, an alternative approach that involves comparing sequences of non-mobile genes encoding metabolic enzymes is gaining popularity [2]. To illustrate this, we will discuss recent work on the evolution of *Shigella* [3] and pathogenic *Bacillus* species [4].

Dysentery has long been a bane of humanity, particularly during wartime or following natural disasters, and in any human habitat where sanitary standards are minimal or compromised [5]. The disease is transmitted by the fecal–oral route, and is characterized by an acute intestinal inflammation resulting in mucosal damage and diarrhea (often bloody), fever and abdominal cramps. Mortality rates vary widely depending on the pathogen involved, the age and health of the patient, and the availability of medical care. *Shigella* species are the major bacterial cause of dysentery, or 'shigellosis'. Millions of cases of shigellosis occur every year, mostly in underdeveloped regions of the world, but not exclusively. In more economically developed nations, outbreaks of shigellosis have occurred in daycare centers and institutions for the mentally disabled, on cruise ships and through food service facilities [5]. As an example, a recent shigellosis outbreak at a San Francisco area restaurant that was attributed to contaminated salsa sickened over 200 customers and killed one.

Several factors contribute to making *Shigella* a successful pathogen (Table 1). The infectious dose necessary to cause illness is exceedingly low — fewer than 100 bacteria consumed orally can produce diarrhea [5]. Shigellae are resistant to gastric acid and bile salts, ordinarily formidable hazards for microbes moving through the gastrointestinal tract. Once they reach the intestines, Shigellae invade the mucosal epithelium and spread from cell to cell in the epithelial layer, causing extensive necrosis while avoiding host defenses [6]. Shigellae also secrete potent cytotoxins which amplify the damage. Shiga toxin, which is produced exclusively by *S. dysenteriae* type 1, is the best known and most harmful *Shigella* toxin. Enterotoxins have been identified in other *Shigella* strains as well [5,7]. As with many bacterial pathogens, an increasing number of *Shigella* isolates are resistant to multiple antibiotics, complicating treatment of the disease.

Four species of *Shigella* are recognized (Table 1), all of which are similar genetically and metabolically to *E. coli*. Indeed, it is generally acknowledged that they could rationally be considered part of the species *E. coli* [7]. The official taxonomic distinction has nevertheless been retained, mostly because shigellosis is a significant disease, and the species designations are useful for diagnostic, treatment and epidemiological purposes. Laboratory differentiation between *Shigella* and *E. coli* isolates relies initially on the lack of motility and inability to ferment lactose exhibited by Shigellae; subsequent characterization is based on serology and a limited number of additional biochemical features (Table 1).

In an effort to clarify relationships between *Shigella* and *E. coli* strains, Pupo *et al.* [3] obtained DNA sequences of eight housekeeping genes from 46 *Shigella* strains (including representatives of all known serotypes), eight *E. coli*

Table 1

Phenotypic differences between Shigellae, enteroinvasive *E. coli* and commensal *E. coli*.

	Shigella dysenteriae	*Shigella flexneri*	*Shigella boydii*	*Shigella sonnei*	Enteroinvasive *E. coli*	Commensal *E. coli*
Serotypes	12	6	18	1	14	Several hundred
Plasmid	140 MDa	140 MDa	140 MDa	120 MDa	140 MDa	Variable types, if present
Virulence factors						
Gastric acid resistance	High	High	High	High	Moderate	Low
Host cell invasion	+	+	+	+	+	–
Shiga toxin	+	–	–	–	–	–
Other enterotoxins	+	+	+	+	+	–
Lysine decarboxylase, Lactose fermentation	–	–	–	–	Variable	+
Flagellar motility	–	–	–	–	–	Most
Infection consequences	Most severe dysentery	Dysentery	Dysentery	Often watery diarrhea w/o dysentery	Often watery diarrhea w/o dysentery	None
Prevalence in U.S.	Rare	Occasional	Rare	Most common	Rare	Ubiquitous
Prevalence in developing countries	Common	Occasional	Common	Occasional	Rare	Ubiquitous

strains and *Salmonella enterica* LT2, a more distantly related enteric pathogen. Over 7000 total base pairs of DNA sequence from each strain were compared. Sequence alignments were used to assemble phylogenetic trees identifying nearest neighbors for each strain (Figure 1a). The results show three major clusters of closely related *Shigella* strains, and five outlying strains. Most of the *Shigella* branches of the tree are flanked by *E. coli* strains, confirming that the *Shigella* 'genus' as a whole fits comfortably within the *E. coli* taxonomic group. Only one *S. boydii* isolate is so divergent that it seems truly worthy of a separate species designation. Each of the three major clusters of *Shigella* strains contain serotypes previously assigned to different species, suggesting that the *Shigella* species designations currently used do not even accurately reflect common evolutionary descent.

The existence of distinct lineages of *Shigella* strains within the *E. coli* taxonomic group suggests that the pathogenic phenotype characteristic of Shigellae could have arisen independently multiple times. This is not an unreasonable assertion, considering that the most important trait allowing these bacteria to cause disease — the ability to invade and spread between mucosal epithelial cells — is attributable primarily to plasmid-encoded genes [8]. Acquisition of the virulence-determining plasmid would have been a prerequisite for *Shigella*-like pathogenicity. Plasmids are a proficient vehicle for lateral gene transfer by conjugation, and the crowded microcosm of the bowel certainly provides ample opportunity for plasmid-bearing cells to conjugate with commensal *E. coli*, creating new strains.

The other phenotypic properties that characterize *Shigella* species, such as lack of flagellar motility, could have arisen by mutation after virulence plasmid acquisition. Different *Shigella* strains have distinct mutations eliminating flagellar synthesis, consistent with independent evolutionary histories [9]. With the ability to spread intracellularly using actin-based motility, flagellar motility might be dispensable, or even disadvantageous because of its high energetic costs. Additional selective pressures may have driven the convergent loss of other genes in *Shigella* strains. Loss of lysine decarboxylase through genetic deletion appears to be favorable, as the product of the lysine decarboxylase reaction inhibits enterotoxin activity [10]. Deletion of gene encoding lysine decarboxylase has occurred in at least one enteroinvasive *E. coli* strain as well. Enteroinvasive *E. coli* are diarrheagenic and resemble *Shigella* in enough ways (Table 1) to suggest that the invisible hand of natural selection that created *Shigellae* is also guiding the convergent evolution of enteroinvasive *E. coli*.

Multilocus DNA sequence typing has also been used to explore the evolution of other bacterial pathogens, including spore-formers from the genus *Bacillus*. *Bacillus cereus* is a common soil bacterium, but it is also an opportunistic pathogen associated with food poisoning, trauma-associated eye infections and periodontal disease. *Bacillus thuringiensis* is a virulent insect pathogen that produces crystalline toxins in association with its spores [11]. Genes encoding these toxins of various *B. thuringiensis* strains have been inserted into plant genomes to engineer pest-resistant crops. The final important member of the group, *Bacillus anthracis*, attained a prominent place in the early history of microbiology in the 1870s when Robert Koch conclusively proved that this bacterium causes anthrax. This was the first time a specific microorganism had been proved to cause a disease, and the now-famous 'Koch's postulates' for establishing the aetiological agent of a disease were inspired in part by this work. Anthrax contin-

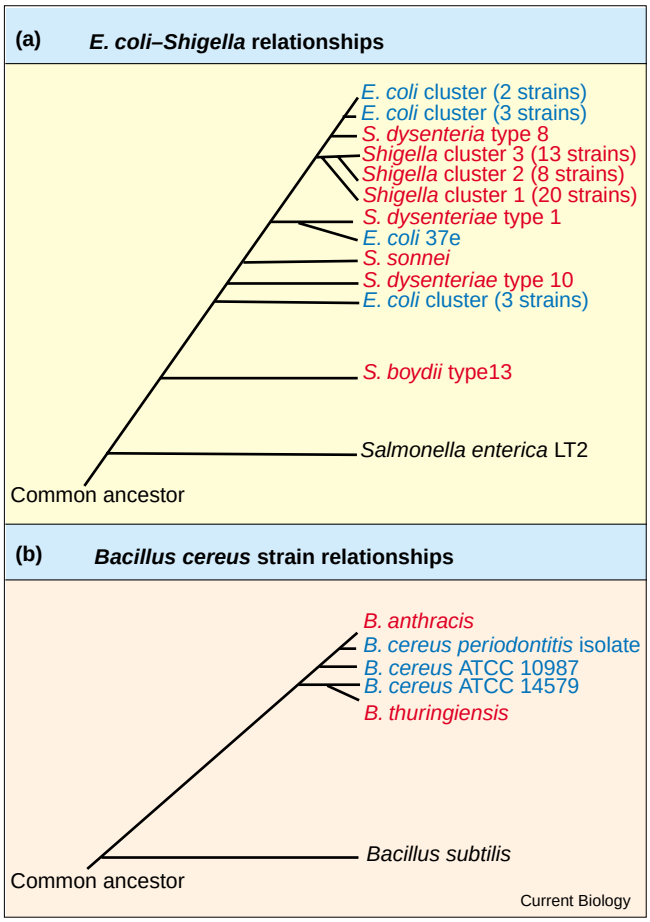
ues to be a global problem for domestic livestock and wild ungulants, but fortunately is increasingly rare as a human disease, at least in developed countries. It nevertheless remains a problem of great interest, as *B. anthracis* spores, which cause a highly lethal form of anthrax when inhaled, could potentially be used as a biological weapon by terrorists or rogue nations [12]. (As a side note, the extremely low infectious dose of *Shigella* also makes it a concern as a bio-terrorism agent.)

What do these three *Bacillus* species exhibiting wildly divergent pathogenic behaviors have in common? In fact, *B. thuringiensis* is virtually indistinguishable from *B. cereus* when the large plasmid that carries the crystalline toxin genes is lost [11]. *B. anthracis* contains two large virulence-determining plasmids, one encoding genes for synthesizing a protective capsule and the other encoding anthrax toxin [13]. Without these plasmids, this organism too acts very much like *B. cereus*. To examine their underlying genetic similarity, Helgason *et al.* [4] obtained DNA sequence data from nine chromosomal genes in three *B. cereus* strains, a mosquito-specific *B. thuringiensis* strain and a *B. anthracis* strain. Pairwise comparisons showed 92–99% sequence identity among all the genes. A tree taking into account all of the sequence comparisons shows *B. anthracis* very closely linked to a periodontitis-causing *B. cereus* strain, and *B. thuringiensis* closely aligned with another reference strain of *B. cereus* (Figure 1b).

These relationships are supported by analysis of electrophoretic mobility patterns for 13 enzymes from over 200 strains of *B. anthracis*, *B. thuringiensis* and *B. cereus* [4]. A neighbor-joining tree based on the complete allozyme data set shows *B. cereus* and *B. thuringiensis* strains intermingled. The *B. anthracis* strains are all tightly clustered with two groups of *B. cereus* strains, one of which was isolated from humans and another that was isolated from soil in the vicinity of an anthrax outbreak. The latter strains have a chromosomal marker previously identified as indicative of *B. anthracis*, but lacked the two *B. anthracis* virulence plasmids. These strains thus appear to be degenerate *B. anthracis* that, by virtue of having lost their plasmids, have reverted to something like their ancestral state, plasmid-free *B. cereus*. It is not much of a stretch to envision proto-*B. anthracis* and *B. thuringiensis* strains being continually created by plasmid acquisition, as the virulence plasmids of *B. anthracis* and *B. thuringiensis* are readily transferable by conjugation. It is not as clear as with *Shigella*, though, what sort of environment might facilitate such genetic exchange.

Is it possible to estimate when contemporary strains of bacterial pathogens originated? Pupo *et al.* [3] used mutation frequencies to calculate divergence times for the major *Shigella* clusters from *E. coli* relatives on the basis of a range of published molecular clock rates, obtaining estimates of between 35,000 and 270,000 years ago. This dating calls into question the notion that 'crowd diseases' such as shigellosis — which are human-specific, rely heavily on person-to-person transmission and are short in duration — were unlikely to arise until humans developed agriculturally based societies with reasonably large population centers [3,14]. Such conditions have apparently

Figure 1



Phylogenetic trees showing relationships among bacterial pathogens determined by multilocus DNA sequence comparisons. The distance between branches of each tree reflects genetic divergence between strains. The trees shown here are merely meant to be illustrative. Interested readers should see [3,4] for the actual sequence data, original dendrograms and the methods by which they were generated. **(a)** Relationships between *E. coli* and *Shigella* strains (highlighted in red) [3]. The type designation after a strain refers to the serotype. *Shigella* cluster 1 contains nine *S. dysenteria*, nine *S. boydii* and two *S. flexneri* serotypes. Cluster 2 contains seven *S. boydii* and one *S. dysenteriae* serotype. Cluster 3 contains twelve *S. flexneri* and one *S. boydii* serotype. For perspective, the genera *Salmonella* and *Escherichia* are estimated to have had a common ancestor approximately 140 million years ago [18], while the three major *Shigella* clusters are estimated to have diverged from each other 35,000–160,000 years ago [3]. **(b)** Relationships between *Bacillus* strains [4]. DNA sequences from three strains of *B. cereus* were compared to one strain of *B. anthracis* and one strain of *B. thuringiensis v. kurstaki* (the latter two strains are highlighted in red).

occurred only in the last approximately 10,000 years. How *Shigella* strains might have been maintained in smaller, more isolated groups of humans is an interesting question worthy of further thought.

Anthrax-causing strains tell a different story. Anthrax isolates from around the world show so little phenotypic or genetic heterogeneity that it has taken a great deal of effort to identify genetic markers useful for distinguishing *B. anthracis* strains for epidemiological investigations [15]. Does that mean that anthrax is a 'new' disease? Probably not, as its animal hosts have been around much longer than modern humans. Population bottlenecks can produce genetic homogeneity, but why would *B. anthracis* be affected by this more than any other pathogen? Although not a complete answer, it is worth pointing out that the biology of endospore-forming bacteria can slow the rate at which mutations accumulate. *Bacillus* spores are inert, non-replicating cells in which the genome is vigilantly protected from damage. Spores can survive for long periods of time — to cite an extreme case, revival of *Bacillus* spores after 250 million years has recently been reported [16]. If *B. anthracis* and its relatives spend a significant time between infections as inert spores in soil, mutation frequencies cannot be readily used to calculate the real time elapsed since different strains diverged.

Few biologists would dispute the notion that taxonomy should reflect true evolutionary relationships, but this has been a continuing challenge with microorganisms. Only in the last 20 years have molecular sequence methods become the gold standard for establishing phylogenetic relationships among prokaryotes [17]. As PCR amplification and DNA sequencing become easier, faster and cheaper, such methods are seeing wider use in diagnostic, epidemiological, and evolutionary studies. Enteroinvasive *E. coli* infections, for example, are confirmed by genetic analysis [7]. The new work discussed here illustrates nicely how such tools can reveal underlying genetic relationships in microorganisms despite diverse phenotypic properties. Not coincidentally, such studies also emphasize the role of horizontal gene transfer in the evolution and diversification of pathogenic microbes, something we would do well to appreciate at a time when the global battle against infectious diseases seems to have reached a standstill.

References

1. Finlay BB, Falkow S: Common themes in microbial pathogenicity revisited. *Micro Mol Biol Rev* 1997, 61:136-169.
2. Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, *et al.*: Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci USA* 1998, 95:3140-3145.
3. Pupo GM, Lan R, Reeves P: Multiple independent origins of ***Shigella*** clones of ***Escherichia coli*** and convergent evolution of many of their characteristics. *Proc Natl Acad Sci USA* 2000, 97:10567-10572.
4. Helgason E, Okstad OA, Caugant DA, Johansen HA, Fouet A, Mock M, Hegna I, Kolsto: ***Bacillus*** anthracis, ***Bacillus*** cereus, and ***Bacillus*** thuringiensis — one species on the basis of genetic evidence. *Appl Environ Microbiol* 2000, 66:2627-2630.
5. Keusch GT, Bennish ML: Shigellosis. In *Bacterial Infections of Humans.* 3rd ed. Edited by Evans AS, Brachman P. New York: Plenum Medical; 1998:631-656.
6. Sansonetti PJ, Tran Van Nhieu G, Egile C: Rupture of the intestinal epithelial barrier and mucosal invasion by ***Shigella*** flexneri. *Clin Infect Dis* 1999, 28:466-475.
7. Bopp CA, Brenner F, Wells JG, Strockbine NA: ***Escherichia***, ***Shigella***, and ***Salmonella***. In *Manual of Clinical Microbiology.* 7th edition. Washington DC: ASM Press; 1999:459-474.
8. Hale T: Genetic basis of virulence in ***Shigella*** species. *Microbiol Rev* 1991, 55:206-224.
9. Al Mamun AA, Tominaga A, Enomoto M: Cloning and characterization of the region III flagellar operons of the four ***Shigella*** subgroups: genetic defects that cause loss of flagella of ***Shigella*** boydii and ***Shigella*** sonnei. *J Bacteriol* 1997, 179:4493-4500.
10. Maurelli AT, Fernandez 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.
11. Schnepf E, Crickmore N, Van Rie J, Lereclus D, Baum J, Feitelson J, Zeigler DR, Dean DH: ***Bacillus thuringiensis*** and its pesticidal crystal proteins. *Microbiol Mol Biol Rev* 1998, 62:775-806.
12. Zilinskas RA: Terrorism and biological weapons: inevitable alliance? *Perspec Biol Med* 1990, 34:44-72.
13. Okinaka R, Cloud K, Hampton O, Hoffmaster A, Hill K, Keim P, Koehler T, Lamke G, Kumano S, Manter D, *et al.*: Sequence, assembly and analysis of pX01 and pX02. *J Appl Microbiol* 1999, 87:261-262.
14. McKeown T: *The Origins of Human Disease.* Oxford: Blackwell Publishing; 1998.
15. Keim P, Price LB, Klevytska AM, Smith KL, Schupp JM, Okinaka R, Jackson PJ, Hugh-Jones ME: Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within ***Bacillus*** anthracis. *J Bacteriol* 2000, 182:2928-2936
16. Vreeland RH, Rosenzweig WD, Powers DW: Isolation of a 250 million-year-old halotolerant bacterium from a primary salt crystal. *Nature* 2000, 407:897-900.
17. Woese CR: There must be a prokaryote somewhere: Microbiology's search for itself. *Microbiol Rev* 1994, 58:1-9.
18. Selander RK, Li J, Nelson K: Evolutionary genetics of ***Salmonella enterica***. In *Escherichia coli and Salmonella: Cellular and Molecular Biology.* 2nd Edition. Washington DC: ASM Press; 1996:2691-2707.