Published in final edited form as:

Mol Microbiol. 2006 February ; 59(3): 975–988.

Clonal analysis reveals high rate of structural mutations in fimbrial adhesins of extraintestinal pathogenic *Escherichia coli*

Scott J. Weissman¹, Sujay Chattopadhyay², Pavel Aprikian², Mana Obata-Yasuoka³, Yuliya Yarova-Yarovaya⁴, Ann Stapleton⁴, William Ba-Thein⁵, Daniel Dykhuizen⁶, James R. Johnson^{7,8}, and Evgeni V. Sokurenko^{2,*}

¹Division of Infectious Disease, Immunology and Rheumatology, Department of Pediatrics, University of Washington, Seattle, WA, USA.

²Department of Microbiology, University of Washington, Seattle, WA, USA.

³Department of Obstetrics/Gynecology, University of Tsukuba, Tsukuba, Japan.

⁴Division of Allergy and Infectious Disease, Department of Medicine, University of Washington, Seattle, WA, USA.

⁵Department of Microbiology/Immunology, Thammasit University, Pathum-Thani, Thailand.

⁶Department of Evolutionary Biology, State University of New York, Stony Brook, NY, USA.

⁷Mucosal and Vaccine Research Center, Minneapolis VA Medical Center, University of Minnesota, Minneapolis, MN, USA.

⁸Department of Medicine, University of Minnesota, Minneapolis, MN, USA.

Summary

Type 1 fimbriae of *Escherichia coli* mediate mannose-specific adhesion to host epithelial surfaces and consist of a major, antigenically variable pilin subunit, FimA, and a minor, structurally conserved adhesive subunit, FimH, located on the fimbrial tip. We have analysed the variability of *fimA* and *fimH* in strains of vaginal and other origin that belong to one of the most prominent clonal groups of extraintestinal pathogenic *E. coli*, comprised of O1:K1-, O2:K1- and O18:K1-based serotypes. Multiple locus sequence typing (MLST) of this group revealed that the strains have identical (at all but one nucleotide position) eight housekeeping loci around the genome and belong to the ST95 complex defined by the publicly available *E. coli* MLST database. Multiple highly diverse *fimA* alleles have been introduced into the ST95 clonal complex via horizontal transfer, at a frequency comparable to that of genes defining the major O- and H-antigens. However, no further significant FimA diversification has occurred via point mutation after the transfers. In contrast, while *fimH* alleles also move horizontally (along with the *fimA* loci), they acquire point amino acid replacements at a higher rate than either housekeeping genes or *fimA*. These FimH mutations enhance binding to monomannose receptors and bacterial tropism for human vaginal epithelium. A similar pattern of rapid within-clonal structural evolution of the adhesive, but not pilin, subunit is also seen, respectively, in *papG* and *papA* alleles of the di-galactose-specific P-fimbriae. Thus, while structurally diverse pilin subunits of *E. coli* fimbriae are under selective pressure for frequent horizontal transfer between clones, the adhesive subunits of extraintestinal *E. coli* are under strong positive selection ($D_n/D_s > 1$ for *fimH* and *papG*) for functionally adaptive amino acid replacements.

*For correspondence. E-mail evs@u.washington.edu; Tel. (+1) 206 685 2162; Fax (+1) 206 543 8297..

Introduction

Evolution of bacterial virulence may be viewed as a process of adaptation that enables a pathogenic lifestyle, often involving acquisition of the ability to spread from a primary niche (e.g. a commensal or environmental reservoir) to a novel niche (e.g. a normally sterile or protected host compartment). At the genetic level, virulence evolution may occur either by horizontal transfer of novel genes encoding virulence factors or by functional modification of pre-existing genes (Sokurenko *et al.*, 1999). Mutations producing functional modifications are called pathogenicity-adaptive, or pathoadaptive, mutations and are often single nucleotide polymorphisms (SNPs) producing amino acid replacements in proteins essential for a pathogen's success. One important group of such proteins is the bacterial adhesins, which mediate receptor-specific interactions with target cells or tissues and are generally necessary for the successful colonization and infection of the host. In this study, we show that two major adhesins of extraintestinal pathogenic *Escherichia coli* (ExPEC) – type 1 and P fimbrial adhesins – acquire structural SNPs at a rapid rate, and this adaptation constitutes a major factor in the pathoadaptive microevolution and genetic diversification of ExPEC clonal groups.

Escherichia coli are common inhabitants of the large intestine of healthy humans and other vertebrates, constituting the species' primary habitat. *E. coli* are also associated with infections of a variety of extraintestinal sites that include, but are not limited to, genitourinary compartments such as the vaginal introitus, periurethral area and urinary bladder (Johnson and Russo, 2002). These can be considered as secondary habitats for the bacteria, colonization of which is relatively short-term. ExPEC also cause severe invasive diseases, such as urosepsis and meningitis, and most isolates recovered in such extraintestinal syndromes belong to phylogenetic group B2 of the *E. coli* species.

The vast majority of both intestinal and extraintestinal *E. coli* express type 1 fimbriae – hair-like, adhesive appendages peritrichously distributed around the bacterial surface (Brinton, 1959). The fimbrial body is built from the major pilin subunit, FimA (15 kDa), while the adhesive subunit, mannose-binding protein FimH (30 kDa), is located at the fimbrial tip (Klemm and Christiansen, 1987). In both primary intestinal and secondary extraintestinal habitats, type 1 fimbriae play an important role. For intestinal *E. coli*, FimH is thought to mediate transient colonization of the oropharyngeal epithelium and, in this way, contribute to fecal/oral transmission between hosts (Bloch *et al.*, 1992). For extraintestinal *E. coli*, FimH is a critical determinant of tropism for the urinary tract (Hung *et al.*, 2002) and vaginal epithelium (Klumpp *et al.*, 2002). It was shown previously that naturally occurring amino acid replacements in FimH can modify its tropism towards uroepithelium and various components of basement membranes (Korhonen *et al.*, 1992; Sokurenko *et al.*, 1995; 1997; Pouttu *et al.*, 1999). These replacements increase the monomannose (1M)-binding capability of FimH by affecting shear-dependent conformational properties of the protein (Thomas *et al.*, 2002). The mutated FimH variants were shown to be associated with uropathogenic, but not intestinal, *E. coli* strains; to be pathoadaptive, providing advantage in the colonization of the urinary tract (Sokurenko *et al.*, 1998); and to correlate with overall extraintestinal virulence of *E. coli* (Hommais *et al.*, 2003). Thus, FimH mutations may be adaptive in secondary, but not primary, habitats of *E. coli*.

Although the structural SNPs enhancing 1M binding by FimH appear to be adaptive in nature, molecular evidence for positive selection in *fimH* is rather indirect (Sokurenko *et al.*, 2004). This is due to the fact that structural mutations in the alleles (i) occur singly or in small numbers; (ii) are of diverse nature and scattered across the gene length; and (iii) arise independently in diverse allelic backgrounds. Under such mutational dynamics, SNP variation among alleles randomly sampled at the species level is hard to distinguish from neutral evolution. For example, the most direct method of detecting positive selection for structural SNPs – the ratio

of rates of non-synonymous and synonymous mutation (Dn/Ds) – failed to uncover evidence of selection for amino acid replacements in the FimH adhesin, as Dn was significantly below Ds (Sokurenko *et al.*, 2004). Furthermore, neither the clonal dynamics nor the rate of molecular evolution of *fimH*, relative to other fimbrial and housekeeping genes, are known.

In this study, we have analysed molecular variation in *fimH*, *fimA* and eight housekeeping loci at the level of clonally related ExPEC strains, with a (nearly) identical set of eight housekeeping loci, that represent one of the most epidemiologically prominent groups among ExPEC isolates (Kunin *et al.*, 1993; Johnson *et al.*, 2001a). Such novel clonal analysis of microevolutionary events directly demonstrates that structural SNPs in *fimH* occur under strong positive selection for pathoadaptive functional change, and accumulate at a dramatically higher rate than the other genes studied. Similar results were obtained with another important adhesin of ExPEC strains – the P fimbrial subunit, PapG.

Results

ST95 includes K1-bearing strains with various O- and H-antigens

To identify a group of clonally related ExPEC isolates (Table 1), we selected a widely disseminated group of K1 capsule-bearing strains belonging to serotypes O1:K1:H7, O2:K1:H7 and O18:K1:H7, shown by multiple locus enzyme electrophoresis (MLEE) to be closely related to one another (Achtman and Pluschke, 1986), and which includes such archetypal ExPEC strains as Nu14 (cystitis) and RS218 (newborn meningitis – NBM). Strains from this group belong to phylogenetic group B2 (Johnson *et al.*, 2000), one of four major phylogenetic clusters identified through study of the *E. coli* reference collection (ECOR, Ochman and Selander, 1984), and one that is particularly rich in ExPEC strains (Boyd and Hartl, 1998a; Johnson and Russo, 2002). We also tested strains of unknown phylogeny that were suspected to belong to this group of strains based on O:K serotype and virulence factor carriage (data not shown). These included strains of vaginal origin belonging to serotypes O1:K1:H1 and O2:K1:NM, isolated in Japan; K1 reference strain U5/41 (ATCC 11775, O1:K1:H7); and model fecal isolate F-18 (OR:K1:H5). For comparison, we included representatives of two K1 serotypes known to be electrophoretically distinct from the O1/2/18:K1:H7 strains: strain JVEC10 belonging to serotype O2:K1:H6 from phylogenetic group B2, and strain JVEC9 belonging to serotype O1:K1:NM from phylogenetic group D, another distinct ExPEC-rich cluster. We also included archetypal, non-K1 strains of known phylogenetic origin: model pyelonephritis strains 536, CFT073 and J96 (all phylogenetic group B2), as well as model K-12 laboratory strain MG1655 (phylogenetic group A).

To establish phylogenetic relationships among study strains, we subjected the entire collection to multiple locus sequence typing (MLST) by sequencing of 450–500 bp internal fragments from eight housekeeping gene loci around the *E. coli* chromosome – *adk*, *fumC*, *gyrB*, *icd*, *mdh*, *metG*, *purA* and *recA*.

All but three strains of the well-known serotypes O1:K1:H7, O2:K1:H7 and O18:K1:H7 as well as strains of serotypes O1:K1:H1, O2:K1:NM, and model strains U5/41 (O1:K1:H7) and F-18 (OR:K1:H5) exhibited complete nucleotide sequence identity over all housekeeping loci sequenced. Based on these housekeeping gene haplotypes, this clonal group belongs to sequence type (ST) 95, as defined in the *E. coli* MLST database (<http://web.mpiib-berlin.mpg.de/mlst/dbs/Ecoli>). Strains IHE3040, IHE3047 and IHE3080 (serotype O18:K1:H7) differed from the ST95 haplotype by a synonymous SNP in *gyrB*, resulting in assignment to ST390. As ST95 is considered the founding ST to ST390, both STs are assigned to the ‘ST95 complex’.

Multiple locus sequence typing analysis of the remaining study collection strains showed them to be distinct from the ST95 complex strains. Using concatenated 3865 bp sequences for each strain, we generated a maximum-likelihood phylogram (Fig. 1). As expected, other B2 strains were distinct from – but rather closely clustered with – ST95, while strain MG1655 (group A) and the serotype O1:K1:NM strain (group D) located quite distant from ST95 and from each other.

Thus, *E. coli* strains belonging to serotypes O1:K1:H7, O2:K1:H7 and O18:K1:H7, along with serotypes O1:K1:H1, O2:K1:NM and OR:K1:H5, constitute a complex of very closely related strains – the ST95 complex, according to the MLST-based clonal grouping of the *E. coli* species. The conservation of housekeeping gene sequences within the ST95 complex indicates that these K1 strains are clonal, having diverged from a common ancestor within an evolutionarily recent time span. The observed variability in LPS and flagellar antigen composition in ST95 complex strains is likely generated by inter-strain horizontal transfer of large gene clusters, followed by localized gene exchanges with insertion-flanking sequences (Nelson and Selander, 1994; Milkman *et al.*, 2003), and reflects the contribution of host immune pressure to clonal genetic diversification (Reeves, 1992).

Allelic variation in *fimA* during clonal divergence

Located towards the 5' end of the *fim* cluster, the *fimA* gene encodes the major structural subunit of the type 1 fimbria. It is among the most genetically diverse loci in *E. coli* strains (Boyd and Hartl, 1998b) and exhibits extensive nucleotide polymorphism within regions predicted to encode surface exposed epitopes (Peek *et al.*, 2001). The entire locus (approximately 550 bp) was sequenced for all study isolates, and a *fimA* maximum-likelihood phylogram was constructed. Among the 44 strains of the ST95 complex, four distinct lineages of *fimA* alleles were identified (Fig. 2): lineage A was carried by the single O2:K1:NM strain; lineage B was carried by some O1:K1:H7 and all O2:K1:H7 strains; lineage C was found in all O18:K1:H7 and O1:K1:H1 strains, as well as in the OR:K1:H5 strain; and lineage D appeared in the remaining O1:K1:H7 strains. Within two lineages, two SNPs were identified: (i) in lineage B, a single O1:K1:H7 strain carried *fimA* allele B_1 , derived from primary allele B_0 by one synonymous SNP; and (ii) in lineage C, a single O18:K1:H7 strain carried *fimA* allele C_1 , derived from primary allele C_0 by one non-synonymous SNP.

The variability depicted for ST95 complex strains in the *fimA* phylogram is incongruent with the nearly monophyletic appearance of this complex in the housekeeping gene tree. Specifically, the four *fimA* lineages within the ST95 complex are as distinct from one another at the nucleotide level as they are from *fimA* alleles in strains outside the complex. The average nucleotide distance between the primary *fimA* alleles A, B_0 , C_0 and D was $7.4 \pm 1.9\%$, while that between these four *fimA* sequences and those from the out-group strains was $8.3 \pm 3.2\%$; the corresponding average amino acid distances were 13.2 ± 3.8 aa between complex strains and 14.1 ± 5.5 aa between complex and non-complex strains. The lack of congruence between the housekeeping and *fimA* trees is further demonstrated by the observation that certain ST95 complex *fimA* alleles were closely related to those carried in distantly related out-group strain backgrounds: (i) *fimA* allele A was almost identical to the *fimA* of strains MG1655 and J96, which themselves belong to very distinct phylogenetic groups (A and B2 respectively); and (ii) *fimA* allele C_0 was clustered with the *fimA* of strain JVEC9 (phylogenetic group D). The incongruence of housekeeping and *fimA* trees indicates the evolutionarily recent occurrence of horizontal transfer of *fimA* alleles into the ST95 complex, at a frequency comparable to that of genes defining somatic and flagellar antigens (four *fimA* variant groups compared with three O- and three H-antigens), and is likely driven by immune pressure for the antigenic diversification of the pilin subunit of type 1 fimbriae.

Although two within-lineage *fimA* SNPs provided further genetic diversity in the complex, only one produced an amino acid replacement. Even though this amino acid change is in one of six variable regions predicted to be exposed as fimbrial shaft epitopes (Peek *et al.*, 2001) and could result in antigenic shift, the vast majority of the pilin gene diversity within the ST95 complex derives from the horizontal movement of genetically diverse alleles, rather than from gene mutation.

Allelic variation in *fimH* during clonal divergence

In type 1 fimbriae, the adhesive function is provided by a minor fimbrial subunit, FimH (300 aa nascent protein), encoded at the opposite end of the *fim* cluster from *fimA*. Present in 1/100–1/1000 the amount of the major structural subunit (Madison *et al.*, 1994), FimH is unlikely to be under strong selective pressure as an antigen (Lindberg *et al.*, 1987; Hanson and Brinton, 1988). Indeed, unlike *fimA*, the *fimH* gene is highly conserved (Sokurenko *et al.*, 1998). To compare within-clonal complex variability of FimH to that observed in FimA, we sequenced the type 1 fimbrial adhesin locus in all study strains.

Within the ST95 complex, we identified five *fimH* lineages (Fig. 3), named *a*, *b*, *b'*, *c* and *d* (to correspond with the names of the *fimA* lineages from the same strains). As with *fimA*, the primary *fimH* alleles within the ST95 complex were as distinct from one another as they were from alleles outside the complex, although the variability of *fimH* was much less than that of *fimA*, as expected. The five primary *fimH* alleles of the ST95 complex differed from one another by $1.9 \pm 0.4\%$ of nucleotide positions, and from those of the out-group strains by $2.0 \pm 0.8\%$. Correspondingly, the average number of amino acid differences was 3.2 ± 1.0 aa within the complex and 3.1 ± 1.7 aa between complex and non-complex strains.

The *fimH* allele *a*, carried by strain JVEC39 (serotype O2:K1:NM), was identical to *fimH* in the MG1655 and J96 strains. Because these three strains also carry almost identical *fimA* alleles, it appears that not only the individual *fimA* and *fimH* loci, but an entire *fim* cluster has entered into three phylogenetically distinct lineages (two lineages in group B2, one in group A). Perfect sequence identity in the mid-cluster loci *fimI* and *fimC* of these three strains further supports this conclusion (data not shown). Thus, *fimH* alleles move horizontally between *E. coli* strains, typically together with the antigenically variable *fimA* locus.

On the other hand, O1/2:K1:H7 strains that carried *fimA* allele *B* were split into two groups – one carrying *fimH* allele *b*, the other carrying a structurally distinct allele, *b'* (Table 1). This indicates that recombination can also occur within the *fim* cluster. The overall incongruence between the *fimA* and *fimH* trees (for instance, lineages *B*, *C* and *D* cluster together in the *fimA* tree, but lineage *c* is distant from *b/b'* and *d* in the *fimH* tree) also suggest that intra-cluster recombination is common.

Finally, three of five primary *fimH* alleles were subjected to further diversification via point mutation, with a single *fimH* SNP each identified in the *b'* and *d* lineages, and five SNPs in the *c* lineage; all seven mutations were non-synonymous. Thus, while *fimH* alleles move horizontally into the ST95 complex, they also frequently acquire point mutations following transfer. The physiologic significance of mutation acquisition in *fimH* is the main focus of the remainder of the study.

FimH is evolving under positive selection

We compared the number of point mutations in *fimH* acquired by the primary alleles within the ST95 complex strains with the number of analogous point mutations in housekeeping loci and in *fimA*, normalizing by locus sequence length. The number of SNPs in *fimH* and *fimA* were both greater than that for all positions in housekeeping loci – $P < 0.0001$ and $P = 0.04$

respectively (Fisher's exact test, two-tailed). However, the number of non-synonymous SNPs was significantly greater for *fimH*, but not *fimA*, relative to housekeeping loci ($P < 0.0001$ and $P = 0.12$ respectively). Thus, the *fimH* locus is targeted by non-synonymous point mutations at a higher frequency than are *fimA* or the housekeeping genes.

Next, we determined the ratio of the non-synonymous mutation rate (Dn) to the synonymous mutation rate (Ds). Dn/Ds ≤ 1 indicates purifying or neutral selection, favouring amino acid conservation, while Dn/Ds > 1 indicates positive selection, favouring amino acid substitutions (Hughes and Nei, 1989; Tanaka and Nei, 1989). Here, we treated all primary alleles as a single concatenated haplotype and, for the purpose of calculation, considered only the SNPs acquired within the clonal complex. Because the absence of synonymous *fimH* mutations in the ST95 complex strains precludes calculation of Dn/Ds, we inserted a hypothetical single synonymous mutation into the *fimH* allele pool. The presence of seven non-synonymous SNPs and one (hypothetical) synonymous SNP within the ST95-complex *fimH* lineages corresponds to Dn/Ds 3.32 ± 1.14 (SE), consistent with strong positive selection for structural evolution of FimH. Calculation of this ratio using the Dn/Ds module of the Phylogenetic Analysis by Maximum Likelihood (PAML) software package (<http://abacus.gene.ucl.ac.uk/software/paml.html>), based on modification (Goldman and Yang, 1994) of Nei and Gojobori (1986), yielded Dn/Ds 4.11, indicating even more strongly the operation of positive selection on FimH. In contrast, the one non-synonymous SNP and one synonymous SNP in the *fimA* lineages and one synonymous mutation in the MLST loci do not support the operation of positive selection for non-synonymous SNPs in either pilin or housekeeping loci during clonal diversification.

Finally, in clones exhibiting *fimH* mutations, we determined whether the 900 bp *fimH* locus is specifically subject to high rates of mutation within the 10 kb *fim* cluster region. We used a Cel-1 nuclease-based SNP screening technique to compare complete *fim* clusters from three O18:K1:H7 strains bearing successive SNPs within *fimH* lineage *c*: strain 89-1449, which carries the ancestral c_0 allele; strain RS218, which carries allele c_1 (c_0 with t247g polymorphism) and strain F3, which carries allele c_2 (c_1 with c284t polymorphism). Only one Cel-1-specific cut was detected in heteroduplexes created between the 89-1449/RS218, RS218/F3 and 89-1449/F3 *fim* cluster pairs (Fig. 4B), each corresponding to the position of known single (or, in the latter case, adjacent) *fimH* mutations. Therefore, the mutation screen demonstrated lack of additional *fim* cluster mutations outside the adhesin locus, indicating that within the 10 kb *fim* cluster, *fimH* is specifically targeted by SNPs.

Thus, molecular evolutionary analysis indicates that within the ST95 complex, the *fimH* gene is a preferred target for point mutations (compared with housekeeping and other genes in the *fim* cluster), due specifically to positive selection for structural replacements in the adhesin. The high Dn/Ds ratio provides the strongest evidence for the selection, as this is a conservative test producing a biologically intuitive statistic. Interestingly, when Dn/Ds is calculated for *FimH* variation across the *E. coli* species, the ratio is much less than 1, as the signal from purifying selection overwhelms the signal from selection for non-synonymous mutations (Sokurenko *et al.*, 2004). However, the analysis of ST95 complex strains provides unambiguous evidence for strong positive selection operating on *fimH*.

Functional evidence for selection on FimH

We tested whether the *fimH* replacement mutations affected binding function of the FimH adhesin. Of the seven FimH amino acid substitutions found in the ST95 complex, six occurred in the mature peptide coding region of *fimH* – four as solitary mutations in primary alleles (b^1 , c_1 , c_4 and d_1) and two as secondary mutations (c_3 and c_5) – and one occurred in the leader peptide (c_4). We evaluated the impact of the mature peptide mutations on FimH-mediated binding by cloning and expressing primary and mutant alleles in an isogenic background.

Among the five primary ST95-complex FimH alleles, relatively strong monomannose (1M) binding was mediated by alleles *a* and *b* (Fig. 5, black bars), neither of which gave rise to further mutation in this collection; 1M binding levels encoded by the remaining primary alleles (*b'*₀, *c*₀ and *d*₀) were distinctly lower, with *c*₀ encoding one of two evolutionarily conserved FimH variants that are most common among fecal isolates (Sokurenko *et al.*, 2004). Each of the mature peptide mutations acquired in primary alleles during ST95 complex evolution significantly increased 1M binding relative to the respective parent allele, all comparisons by *z*-test $P < 0.0001$ (two-tailed), except *c*₄ versus *c*₀, $P < 0.05$. Double-mutant alleles exhibited stronger binding than their respective parent single-mutant alleles – $P < 0.03$ for *c*₃ versus *c*₁ and $P < 0.0001$ for *c*₄ versus *c*₅.

Because colonization of the vaginal introitus by fecal *E. coli* is relatively common and believed to be a key event leading to urinary tract infection (Schaeffer *et al.*, 1999), and because a considerable proportion of the ST95 complex strains in our collection were of vaginal origin, we then evaluated binding of the study alleles to primary human vaginal epithelial cells. FimH primary alleles with weak 1M-binding conferred poor binding to the epithelial cell monolayers, at levels comparable to that of the *fimH*-null control strain; however, alleles that had conferred stronger 1M binding with purified receptors likewise produced increased epithelial cell binding, as shown in the plot of vaginal epithelial cell binding values against 1M binding values (Fig. 6). The scatter plot indicates a linear relationship with correlation coefficient (*R*) value of 0.86, consistent with strong correspondence between the two phenotypes.

Thus, FimH point mutations acquired in the course of within-clone microevolution are functional in nature, increasing binding to 1M-containing receptors; such mutations were shown previously to be adaptive for uropathogenic *E. coli* (Sokurenko *et al.*, 1997). Here we have shown that increased 1M binding by FimH also correlates with increased adhesion to vaginal epithelial cells. This phenotype may provide selective advantage to *E. coli* strains colonizing extraintestinal niches such as the genitourinary tract, adding functional support to the hypothesis that in the ST95 complex, the *fimH* adhesin is adapting under positive selective pressure.

Allelic variation in P fimbrial determinants

To determine whether other bacterial fimbriae might be under selective pressures similar to those operating on the type 1 fimbria, we studied the within-clonal variability of pilin and adhesive subunits of the P fimbriae, which are more prevalent among ExPEC isolates than fecal isolates (Johnson, 1991). Similar in organization to the *fim* cluster, the *pap* operon contains 11 genes, including the pilin-encoding locus, *papA*, and the adhesin-encoding locus, *papG* (Kuehn *et al.*, 1992). Like type 1 fimbriae, P fimbriae are built primarily from a major subunit, PapA. At least a dozen antigenically distinct variants (F-types) of PapA have been identified by serotyping (Orskov and Orskov, 1990). PapG is the tip-associated adhesin subunit that confers binding to di-galactose carbohydrates (Leffler *et al.*, 1982), common components of cellular glycolipid receptors such as those found in kidney epithelium; three genetically distinct classes of the PapG adhesin have been distinguished based on structural homology and receptor specificity (Marklund *et al.*, 1992). Unlike the ubiquitous *fim* cluster, *pap* operons are introduced into a subset of *E. coli* clonal lineages, as components of pathogenicity-associated islands (Schmidt and Hensel, 2004).

Using universal primers for the major structural subunit locus and class-specific primers for the adhesin locus, we obtained *papA* and *papG* sequences for 35 of 44 strains in the ST95 complex. Sequence analysis for the 35 *pap*-positive ST95 complex strains revealed two phylogenetically distinct primary alleles at both *papA* (Fig. 7) and *papG* (Fig. 8) loci, corroborating previous observations that *pap*-positive O18:K1:H7 strains express the F10 major structural subunit and PapG class III adhesin, while O1:K1:H7 and O2:K1:H7 strains

typically encode an F11 structural subunit in combination with a class II PapG (Johnson and Stell, 2000; Johnson *et al.*, 2000; 2001a). Thus, as with type 1 fimbrial determinants, horizontal mobility of P-fimbrial clusters has contributed to genetic diversity of the ST95 complex; specifically, two distinct *pap* clusters appear to have entered the complex via independent events, likely within larger pathogenicity islands.

Both *papA* and *papG* primary alleles have further diversified by SNP acquisition. We found two non-synonymous SNPs within the F11 *papA* allele lineage. Both *papA* structural mutations were located in hypervariable regions – regions of 50% percent homology or less (Denich *et al.*, 1991) – indicating that they may have been acquired under positive selection, driven by immune escape. However, due to the low total of *papA* mutations detected, this hypothesis cannot be evaluated statistically. In *papG* alleles, we found seven SNPs – five non-synonymous and one synonymous in the class II allele group, and one non-synonymous in the class III group. This distribution furnished evidence of positive selection in *papG*, with Dn/Ds ratio = 2.1.

Thus, similar to the type 1 fimbrial cluster, P fimbrial clusters have contributed to genetic diversification of the ST95 complex both by horizontal mobility and by SNP acquisition, primarily structural mutations of the adhesin. Like FimH, the P fimbrial adhesin may be undergoing pathoadaptive functional modification of its receptor-binding properties (currently under investigation).

Discussion

This study, an analysis of the evolutionary dynamics of *E. coli* adhesins at the level of a recently diverged clonal complex, generated four novel observations. First, strains of diverse bacterial serotypes can carry identical housekeeping loci throughout the chromosomal backbone and, thus, form a clonal complex by MLST. This serologic diversity indicates evolutionarily rapid turnover of major surface antigens – such as LPS and flagella – via localized recombination. The diversification is likely to be driven by negative frequency-dependent selection, in which the fitness advantage of a phenotype increases as its frequency decreases. In this case, strains bearing rare or novel surface antigens will enjoy selective advantage relative to strains bearing antigens against which adaptive immune responses have been developed across the host population. Second, fimbrial operons enter into clonal complexes at a rate comparable to determinants of the classic serotype antigens. Such movement is likely to be driven by frequency-dependent selection also, as the fimbrial pilins constitute major surface antigens. Third, despite high levels of structural variability of major fimbrial subunits, single amino acid replacements are not strongly selected in the (antigenic) pilin loci during clonal diversification. Fourth, in contrast to pilin and housekeeping genes, non-synonymous mutations are acquired at a high rate in fimbrial adhesins, driven by adaptive functional modification. The strongest evidence for strong selection on these adhesin mutations was obtained for the type 1 fimbrial adhesin, FimH. These findings were made possible by the novel approach to molecular evolutionary analysis undertaken here – defining primary alleles in strains from a single clonal complex and tracing mutation acquisition within these allele lineages. This clonal-level analysis provided much better resolution than the species-level sampling traditionally used, wherein synonymous variation overwhelms the non-synonymous variation selected during adaptation.

The dynamics of mutation acquisition in the *E. coli* *fimH* adhesin gene reflects existence of alternative habitats of the species – one where there is either no selection or selection against IM-enhancing mutations, and another where such mutations are advantageous and under positive selection. As most fecal isolates express evolutionarily original, non-mutated FimH proteins, the mammalian large intestine (with fecal-oral transmission between individuals) likely serves as the primary habitat for *E. coli*. We propose that the alternative habitat is likely

a secondary niche, represented by genitourinary compartments such as the vaginal introitus, periurethral area and/or bladder, which are frequently colonized in transient fashion by ExPEC isolates. This scenario implies that structurally conserved, low 1M-binding FimH variants are adapted to the primary niche, while high 1M-binding mutants confer a selective advantage in colonization of secondary habitats, potentially promoting success in host-to-host transmission and expanding the mutant population within the *E. coli* species at large. Indeed, *E. coli* urinary tract isolates express high 1M-binding FimH variants more frequently than do fecal isolates of *E. coli* (Sokurenko *et al.*, 1997). However, it is also possible that the secondary niche to which these clones are adapting may be not an extraintestinal site in the human host, but rather the intestinal tract of non-human or even non-mammalian hosts as, for example, ST95 strains appear to be common among avian isolates (Marc and Dho-Moulin, 1996).

For ExPEC strains, the transfer (and subsequent selection) of *fim* operons could happen in either the primary niche or the secondary habitat. If the transfer, likely driven by strong diversifying selection on FimA, occurs in the primary niche, the potentially pathogenic extraintestinal clone will likely acquire a type 1 fimbrial operon that contains a conserved FimH variant, not adapted for strong 1M receptor binding. These strains then continuously but transiently enter the secondary niches, where selection for 1M-enhancing adhesin mutations is strong, thus giving rise to pathoadapted adhesin variants. In the ST95 complex, this phenomenon is seen most clearly in the type 1 fimbrial operon bearing primary *fimH* allele *c₀*, which encodes one of most common FimH variants among intestinal isolates (Hommais *et al.*, 2003), and which exhibited more structural diversification than any other primary allele.

If, on the other hand, *fim* operon transfer happens in the secondary habitat, then the transferred operon may encode a FimH variant already pathoadapted for strong 1M binding, in which case no further adaptation would occur. This may describe what is seen with primary alleles *a* and *b*, which encode relatively high 1M-binding FimH variants and were not observed to acquire additional mutations. We conclude that these variants were acquired in this adapted state, rather than having emerged within the ST95 complex, because the identical pathoadaptive haplotypes are found in other, unrelated clonal groups (data not shown).

Interestingly, pathoadapted FimH variants have tended not to accumulate silent mutations, consistent with the ‘source-sink’ evolutionary dynamic (Sokurenko *et al.*, 2004). In this scenario, a bacterial population occupying a stable primary niche (‘source’) repeatedly seeds sub-populations into unstable secondary niches (‘sinks’), enabling continuous emergence in the latter of novel adaptive variants that do not sustain themselves over the long-term. The instability in this case is due to (i) the transient nature of secondary habitat colonization and (ii) selection against the mutant alleles in the primary niche, due to detrimental effects of the 1M-enhancing mutations in that setting. Indeed, while increases in 1M binding due to FimH mutation are adaptive for urovirulence (and, possibly, vaginal/periurethral colonization), such adaptation presents a functional trade-off; it is accompanied by decreased resistance to inhibition by soluble mannosylated compounds, which may in turn decrease oropharyngeal colonization in the presence of salivary components. Thus, pathogenic clones may not circulate in the host population long enough to accumulate silent substitutions to accompany the structural SNPs in *fimH*. However, it is possible that adhesin genes are among a small number of loci affected by the source-sink dynamic, such that clones carrying pathoadapted FimH variants could again become stable in the primary niche, if the mutant *fimH* were to be replaced with an evolutionarily conserved, non-mutant allele (via horizontal transfer of an entire or partial *fim* operon), or by either reversion or compensatory mutation.

We see a similar pattern of rapid acquisition of non-synonymous mutations in the P fimbrial adhesin. P fimbriae are generally considered to be a specific virulence factor for ExPEC isolates (Johnson, 1991), providing di-galactose-specific adhesion to upper urinary tract epithelium

(Korhonen *et al.*, 1986). However, experimental evidence also supports a role for these organelles in colonization of the large intestine (Tullus *et al.*, 1992; Wold *et al.*, 1992; Goetz *et al.*, 1999). Although P fimbriae are more common among ExPEC than fecal isolates, almost one-half of resident (as opposed to transient) intestinal *E. coli* strains encode P fimbriae (Nowrouzian *et al.*, 2001). The molecular evolutionary data presented here indicate that, similar to FimH, the P fimbrial adhesin may be subject to 'source-sink' evolutionary dynamics, with differential selective pressures in the primary and secondary niches. Thus, it may be that promoting intestinal colonization is the primary function of the P fimbrial adhesin. Because PapG also plays an important role in extraintestinal infections, the adhesin may be subject to structural modification, likely pathoadaptive in nature, in extraintestinal settings. The functional impact of PapG mutations is currently under investigation.

The process of clonal diversification in *E. coli* represents an intermediate level of evolution between adaptation of an individual strain during colonization or infection of a single host, and adaptation at the species level over more vast periods of time. Accordingly, the clonal analysis presented here takes as its unit of analysis contemporary isolate collections that encompass broad geographic variety, in an effort to sample this clonal diversity. MLST provides an evolutionary framework in which to measure the frequency and direction of recombinations, insertion, deletions and SNPs. When combined with functional studies, we can follow the precise history of a phenotypic trait (such as fimbrial adhesin-mediated binding), allowing inference of cause. Besides providing insight into the evolutionary dynamics of fimbrial genes, the study of adhesin mutations can provide the means to resolve the phylogenetic relationships of bacterial strains within an MLST clonal complex. Depending on the nature of the loci studied, clonal analysis may reveal aspects of structure/function relationship at the molecular level (for example, by revealing mutational 'hot spots') or of host/pathogen interactions at the organismal level, by delineating the evolutionary progression of functionally distinctive haplotypes. Thus, clonal analysis represents a powerful tool in the study of how virulent lifestyles emerge within bacterial populations.

Experimental procedures

Strains and plasmids

The primary collection consisted of 44 K1-encapsulated strains bearing O1, O2 and O18 antigens, in addition to four archetypal, non-K1 strains, including: NBM isolates RS218 and 891449, provided by K.S. Kim (Johns Hopkins University, Baltimore) and L. Spanjaard (Amsterdam) respectively; cystitis isolate F3, provided by A. Stapleton (University of Washington, Seattle); NBM isolates IHE3034, IHE3035, IHE3040, IHE3047 and IHE3080, provided by J. Hacker (Institut für Molekulare Infektionsbiologie, Würzburg, Germany); UTI strain U5/41 (ATCC11775), obtained from H. Rosen (University of Washington, Seattle); fecal isolate F-18, obtained from P. Cohen (University of Rhode Island); and archetypal pyelonephritis isolates CFT073, J96 and 536, obtained, respectively, from H. Mobley, Ann Stapleton and G. Blum-Ohler (Universität Würzburg, Germany). NBM isolates H16, U64 and C5, cystitis isolate Nu14, and fecal isolates S39 and S98 have been described previously (Johnson *et al.*, 2001a). K1-encapsulated isolates bearing O1, O2 or O18 antigens were selected from among vaginal isolates obtained in Japan from healthy and diseased women, as described previously (Obata-Yasuoka *et al.*, 2002). ST95 strains including bacteraemia isolates H25, H35 and 2H25, cystitis isolate Cy179, pyelonephritis isolate Py160, and catheter tip isolate (not associated with bacteraemia) VA1242, were identified within a previously described collection of clinical isolates (Sokurenko *et al.*, 2004).

Phylotyping

Assignment of all study strains to four major phylogenetic groups of *E. coli* was performed using the rapid polymerase chain reaction (PCR)-based technique of Clermont (Clermont *et al.*, 2000).

Nucleotide sequencing

Sequences for the type 1 fimbrial major structural (*fimA*) and adhesin (*fimH*) subunits were obtained for all study strains. Eight housekeeping locus fragments were sequenced for all 50 strains in the study collection, including the seven currently in use (*adh*, *fumC*, *gyrB*, *icd*, *mdh*, *purA* and *recA*) by the publicly available *E. coli* MLST database, as well as *metG*. Primer sequences and relevant methodologic details for MLST are available at the publicly available website (http://web.mpiib-berlin.mpg.de/mlst/dbs/Ecoli/documents/primersColi_html). The primers used for amplification of *metG* were cacatccaggctgatgtctg (forward) and gcagatcaaa gaagaagtgt (reverse).

The primers used for amplification of fimbrial gene loci were: *fimA* F acgtttctgtggctcgacgcatt, R acgtccctgaacctggg taggtta; *fimH* F ggggggtgcactcagggaaccattcaggca, R gggg catgcttattgataaacaagaagtcac; *papA* F atggcagtggtgtcttttggtg, R cgtcccaccatagctgtcttc; *papG* F agtgccgtttcgaatgg, R tcgaggataattcaccag.

All sequences were obtained in both directions, and electropherograms for all singleton mutations were visually inspected for consistency between strands, and any ambiguous nucleotides were resolved by resequencing.

Evolutionary analysis

The eight housekeeping locus sequence fragments were concatenated into a single sequence for each strain prior to subsequent analysis. Nucleotide sequences were aligned using ClustalX 1.83 (Thompson *et al.*, 1997), fimbrial major structural and adhesin subunit protein sequences were predicted, and unrooted maximum-likelihood phylograms were generated using PAUP* 4.0b software package. Phylogram construction allowed identification of distinct allele lineages for all type 1 and P fimbrial loci.

A distance measure was calculated for type 1 fimbrial loci *fimA* and *fimH* both within the ST95 complex and between complex and non-complex (out-group) strains, using ProSeq v2.91 software (Filatov, 2002). To highlight the phylogenetic variability among the 'branches', rather than among the 'twigs', only the primary alleles from each ST95 complex allele lineage (four for *fimA*, five for *fimH*) were utilized in this analysis. For within-complex calculations, six pairwise comparisons were made for *fimA* (4×4 matrix), and 10 pairwise comparisons made for *fimH* (5×5 matrix). For complex/non-complex calculations, 20 pairwise comparisons were made for *fimA* (4×5 matrix), and 15 pairwise comparisons made for *fimH* (5×3 matrix, excluding the horizontally mobile J96/ K12 and JVEC9 alleles from the out-group). The average values were calculated for each matrix from the individual pairwise comparison values.

Rates of non-synonymous (Dn) and synonymous (Ds) mutation were calculated using DnaSP 3.99 software (Rozas *et al.*, 2003) for the major structural and adhesin subunits of both type 1 and P fimbriae, in order to determine the Dn/Ds ratio for each locus. In contrast to the average nucleotide distance calculations, all unique haplotypes observed within the ST95 complex were considered for each locus. To enable calculation of within-lineage mutation rates for each allele (corresponding to rates of mutation following entry of primary alleles into the clonal complex), each locus was considered as a concatenation of its individual lineages. Thus, each unique haplotype was concatenated alongside the primary alleles from the remaining lineages, and the resulting concatenated sequences were compared pairwise to determine mutation rates. In

fimA, for example, there were three unique concatenated haplotypes ($A + B_0 + C_0 + D$, $A + B_1 + C_0 + D$, and $A + B_0 + C_1 + D$), with each sequence totalling 546 + 546 + 552 + 552, or 2196, bp. Average Dn and Ds values for each locus were calculated from the individual pairwise comparisons, and Dn/Ds ratios were calculated from these average values. For each fimbrial type, the Dn/Ds ratios of the major structural and adhesin subunits were compared by Z-test to determine statistical significance.

CEL-1 analysis of nucleotide polymorphisms

Entire *fim* clusters were cloned from the strains of interest by recombinational cloning, adopted from Raymond *et al.* (2002). CEL-1 endonuclease restriction analysis was performed as described previously (Sokurenko *et al.*, 2001). The 10 kb *fim* clusters cloned from *E. coli* O18:K1:H7 strains 89–1449, RS218 and F3 were compared in pairwise manner.

Construction of isogenic recombinant strain

All recombinant strains were created as described elsewhere (Sokurenko *et al.*, 1994). Briefly, *fimH* genes were cloned from wild-type strains by PCR, subcloned into pACYC184-based plasmid pGB17 (Cm^r), and introduced into *fimH*-null *E. coli* strain AAEC191A (pPKL114) for substrate binding and tissue adhesion assays.

Substrate binding assays

Binding properties of FimH variants to yeast mannan (model 1M-like substrate) and bovine RNase B (model 3M-like substrate) was done as described previously (Sokurenko *et al.*, 1997).

Vaginal epithelial cell adhesion assays

Confluent monolayers of primary human vaginal epithelial cells were established from discarded tissues, under protocol approved by University of Washington Institutional Review Board, and maintained in Keratinocyte-SFM media with supplements (Invitrogen, New York) without antibiotics. Cells were seeded onto Laboratory-Tek Chamber Slides (Nalge Nunc Int., Naperville, IL) and grown in 5% CO₂ at 37°C for 24 h to confluence. After washing with Hanks' balanced salt solution (HBSS), cells were incubated at 37°C for 1 h in Keratinocyte-SFM media. Bacteria were adjusted in HBSS to OD₆₀₀ 0.6, and then diluted twofold with or without methyl-D-mannopyranoside inhibitor. Bacteria were added to cells, incubated at 37°C for 1 h, and then removed. Wells were washed six times with 0.5 ml of HBSS, then fixed and stained with Diff-Quik B4132-1 Stain Set (Dade Behring, Newark, DE). After wells were washed with PBS, slides were prepared for microscopic examination. Multiple fields were examined on each cover slip, with the number of bound bacteria on 20 epithelial cells being counted per well. Experiments were performed in triplicate wells.

Acknowledgements

The authors would like to thank Gerald van Belle for valuable statistical consultation, Elena Trinchina and Michael Feldgarden for assistance with cloning and sequencing and Steve Moseley for critical reading and discussion of the manuscript. Research supported by grants from the National Institutes of Health. J.R.J. has received grant support and/or honoraria from Bayer, Wyeth-Ayerst, Ortho-McNeil, Merck and Rochester Medical Group.

References

- Achtman M, Pluschke G. Clonal analysis of descent and virulence among selected *Escherichia coli*. *Annu Rev Microbiol* 1986;40:185–210. [PubMed: 3535645]
- Berger H, Hacker J, Juarez A, Hughes C, Goebel W. Cloning of the chromosomal determinants encoding hemolysin production and mannose-resistant hemagglutination in *Escherichia coli*. *J Bacteriol* 1982;152:1241–1247. [PubMed: 6754701]

- Blattner FR, Plunkett G 3rd, Bloch CA, Perna NT, Burland V, Riley M, et al. The complete genome sequence of *Escherichia coli* K-12. *Science* 1997;277:1453–1474. [PubMed: 9278503]
- Bloch CA, Stocker BA, Orndorff PE. A key role for type 1 pili in enterobacterial communicability. *Mol Microbiol* 1992;6:697–701. [PubMed: 1349417]
- Boyd EF, Hartl DL. Chromosomal regions specific to pathogenic isolates of *Escherichia coli* have a phylogenetically clustered distribution. *J Bacteriol* 1998a;180:1159–1165. [PubMed: 9495754]
- Boyd EF, Hartl DL. Diversifying selection governs sequence polymorphism in the major adhesin proteins *fimA*, *papA*, and *sfaA* of *Escherichia coli*. *J Mol Evol* 1998b;47:258–267. [PubMed: 9732452]
- Brinton C. Nonflagellar appendages of bacteria. *Nature* 1959;183:782–786. [PubMed: 13644195]
- Clermont O, Bonacorsi S, Bingen E. Rapid and simple determination of the *Escherichia coli* phylogenetic group. *Appl Environ Microbiol* 2000;66:4555–4558. [PubMed: 11010916]
- Denich K, Blyn LB, Craiu A, Braaten BA, Hardy J, Low DA, O'Hanley PD. DNA sequences of three *papA* genes from uropathogenic *Escherichia coli* strains: evidence of structural and serological conservation. *Infect Immun* 1991;59:3849–3858. [PubMed: 1682251]
- Filatov DA. ProSeq: a software for preparation and evolutionary analysis of DNA sequence data sets. *Mol Ecol Notes* 2002;2:621–624.
- Goetz GS, Mahmood A, Hultgren SJ, Engle MJ, Dodson K, Alpers DH. Binding of pili from uropathogenic *Escherichia coli* to membranes secreted by human colonocytes and enterocytes. *Infect Immun* 1999;67:6161–6163. [PubMed: 10531281]
- Goldman N, Yang Z. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol* 1994;11:725–736. [PubMed: 7968486]
- Hanson MS, Brinton CC Jr. Identification and characterization of *E. coli* type-1 pilus tip adhesion protein. *Nature* 1988;332:265–268. [PubMed: 2894612]
- Hommais F, Gouriou S, Amarin C, Bui H, Rahimy MC, Picard B, Denamur E. The FimH A27V mutation is pathoadaptive for urovirulence in *Escherichia coli* B2 phylogenetic group isolates. *Infect Immun* 2003;71:3619–3622. [PubMed: 12761149]
- Hughes AL, Nei M. Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proc Natl Acad Sci USA* 1989;86:958–962. [PubMed: 2492668]
- Hull RA, Gill RE, Hsu P, Minshew BH, Falkow S. Construction and expression of recombinant plasmids encoding type 1 or d-mannose-resistant pili from a urinary tract infection *Escherichia coli* isolate. *Infect Immun* 1981;33:933–938. [PubMed: 6116675]
- Hung CS, Bouckaert J, Hung D, Pinkner J, Widberg C, DeFusco A, et al. Structural basis of tropism of *Escherichia coli* to the bladder during urinary tract infection. *Mol Microbiol* 2002;44:903–915. [PubMed: 12010488]
- Johnson JR. Virulence factors in *Escherichia coli* urinary tract infection. *Clin Microbiol Rev* 1991;4:80–128. [PubMed: 1672263]
- Johnson JR, Stell AL. Extended virulence genotypes of *Escherichia coli* strains from patients with urosepsis in relation to phylogeny and host compromise. *J Infect Dis* 2000;181:261–272. [PubMed: 10608775]
- Johnson JR, Russo TA. Extraintestinal pathogenic *Escherichia coli*: 'the other bad *E. coli*'. *J Lab Clin Med* 2002;139:155–162. [PubMed: 11944026]
- Johnson JR, Stell AL, Scheutz F, O'Bryan TT, Russo TA, Carlino UB, et al. Analysis of the F antigen-specific *papA* alleles of extraintestinal pathogenic *Escherichia coli* using a novel multiplex PCR-based assay. *Infect Immun* 2000;68:1587–1599. [PubMed: 10678978]
- Johnson JR, Delavari P, O'Bryan TT. *Escherichia coli* O18:K1:H7 isolates from patients with acute cystitis and neonatal meningitis exhibit common phylogenetic origins and virulence factor profiles. *J Infect Dis* 2001a;183:425–434. [PubMed: 11133374]
- Johnson JR, Weissman SJ, Stell AL, Trintchina E, Dykhuizen DE, Sokurenko EV. Clonal and pathotypic analysis of archetypal *Escherichia coli* cystitis isolate NU14. *J Infect Dis* 2001b;184:1556–1565. [PubMed: 11740731]
- Klemm P, Christiansen G. Three *fim* genes required for the regulation of length and mediation of adhesion of *Escherichia coli* type 1 fimbriae. *Mol Gen Genet* 1987;208:439–445. [PubMed: 2890081]

- Klumpp DJ, Forrestal SG, Karr JE, Mudge CS, Anderson BE, Schaeffer AJ. Epithelial differentiation promotes the adherence of type 1-piliated *Escherichia coli* to human vaginal cells. *J Infect Dis* 2002;186:1631–1638. [PubMed: 12447740]
- Korhonen TK, Virkola R, Holthofer H. Localization of binding sites for purified *Escherichia coli* P fimbriae in the human kidney. *Infect Immun* 1986;54:328–332. [PubMed: 2876959]
- Korhonen TK, Virkola R, Lahteenmaki K, Bjorkman Y, Kukkonen M, Raunio T, et al. Penetration of fimbriate enteric bacteria through basement membranes: a hypothesis. *FEMS Microbiol Lett* 1992;79:307–312. [PubMed: 1362172]
- Krogfelt KA, McCormick BA, Burghoff RL, Laux DC, Cohen PS. Expression of *Escherichia coli* F-18 type 1 fimbriae in the streptomycin-treated mouse large intestine. *Infect Immun* 1991;59:1567–1568. [PubMed: 1672304]
- Kuehn MJ, Heuser J, Normark S, Hultgren SJ. P pili in uropathogenic *E. coli* are composite fibres with distinct fibrillar adhesive tips. *Nature* 1992;356:252–255. [PubMed: 1348107]
- Kunin CM, Hua TH, Krishnan C, Van Arsdale White L, Hacker J. Isolation of a nicotinamide-requiring clone of *Escherichia coli* O18:K1:H7 from women with acute cystitis: resemblance to strains found in neonatal meningitis. *Clin Infect Dis* 1993;16:412–416. [PubMed: 8095816]
- Leffler H, Lomberg H, Gotschlich E, Hagberg L, Jodal U, Korhonen T, et al. Chemical and clinical studies on the interaction of *Escherichia coli* with host glycolipid receptors in urinary tract infection. *Scand J Infect Dis Suppl* 1982;33:46–51. [PubMed: 6127802]
- Lindberg F, Lund B, Johansson L, Normark S. Localization of the receptor-binding protein adhesin at the tip of the bacterial pilus. *Nature* 1987;328:84–87. [PubMed: 2885755]
- Madison B, Ofek I, Clegg S, Abraham SN. Type 1 fimbrial shafts of *Escherichia coli* and *Klebsiella pneumoniae* influence sugar-binding specificities of their FimH adhesins. *Infect Immun* 1994;62:843–848. [PubMed: 7906676]
- Marc D, Dho-Moulin M. Analysis of the *fim* cluster of an avian O2 strain of *Escherichia coli*: serogroup-specific sites within *fimA* and nucleotide sequence of *fimI*. *J Med Microbiol* 1996;44:444–452. [PubMed: 8636962]
- Marklund BI, Tennent JM, Garcia E, Hamers A, Baga M, Lindberg F, et al. Horizontal gene transfer of the *Escherichia coli* *pap* and *prs* pili operons as a mechanism for the development of tissue-specific adhesive properties. *Mol Microbiol* 1992;6:2225–2242. [PubMed: 1357526]
- Milkman R, Jaeger E, McBride RD. Molecular evolution of the *Escherichia coli* chromosome. VI. Two regions of high effective recombination. *Genetics* 2003;163:475–483. [PubMed: 12618387]
- Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 1986;3:418–426. [PubMed: 3444411]
- Nelson K, Selander RK. Intergeneric transfer and recombination of the 6-phosphogluconate dehydrogenase gene (*gnd*) in enteric bacteria. *Proc Natl Acad Sci USA* 1994;91:10227–10231. [PubMed: 7937867]
- Nowrouzian F, Adlerberth I, Wold AE. P fimbriae, capsule and aerobactin characterize colonic resident *Escherichia coli*. *Epidemiol Infect* 2001;126:11–18. [PubMed: 11293669]
- Obata-Yasuoka M, Ba-Thein W, Tsukamoto T, Yoshikawa H, Hayashi H. Vaginal *Escherichia coli* share common virulence factor profiles, serotypes and phylogeny with other extraintestinal *E. coli*. *Microbiology* 2002;148:2745–2752. [PubMed: 12213921]
- Ochman H, Selander RK. Standard reference strains of *Escherichia coli* from natural populations. *J Bacteriol* 1984;157:690–693. [PubMed: 6363394]
- Orskov I, Orskov F. Serologic classification of fimbriae. *Curr Top Microbiol Immunol* 1990;151:71–90. [PubMed: 1973373]
- Peek AS, Souza V, Eguiarte LE, Gaut BS. The interaction of protein structure, selection, and recombination on the evolution of the type-1 fimbrial major subunit (*fimA*) from *Escherichia coli*. *J Mol Evol* 2001;52:193–204. [PubMed: 11244580]
- Pouttu R, Puustinen T, Virkola R, Hacker J, Klemm P, Korhonen TK. Amino acid residue Ala-62 in the FimH fimbrial adhesin is critical for the adhesiveness of meningitis-associated *Escherichia coli* to collagens. *Mol Microbiol* 1999;31:1747–1757. [PubMed: 10209747]

- Raymond CK, Sims EH, Kas A, Spencer DH, Kutyavin TV, Ivey RG, et al. Genetic variation at the O-antigen biosynthetic locus in *Pseudomonas aeruginosa*. *J Bacteriol* 2002;184:3614–3622. [PubMed: 12057956]
- Reeves PR. Variation in O-antigens, niche-specific selection and bacterial populations. *FEMS Microbiol Lett* 1992;79:509–516. [PubMed: 1282486]
- Rozas J, Sanchez-Del Barrio JC, Messeguer X, Rozas R. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 2003;19:2496–2497. [PubMed: 14668244]
- Schaeffer, A.J., Rajan, N., Wright, E.T., Duncan, J.L., and Anderson, B.E. (1999) Role of vaginal colonization in urinary tract infections (UTIs). In *Advances in Bladder Research* Hayward, B.A. (ed.). New York: Kluwer Academic/Plenum, pp. 339–349.
- Schmidt H, Hensel M. Pathogenicity islands in bacterial pathogenesis. *Clin Microbiol Rev* 2004;17:14–56. [PubMed: 14726454]
- Sokurenko EV, Courtney HS, Ohman DE, Klemm P, Hasty DL. FimH family of type 1 fimbrial adhesins: functional heterogeneity due to minor sequence variations among *fimH* genes. *J Bacteriol* 1994;176:748–755. [PubMed: 7905476]
- Sokurenko EV, Courtney HS, Maslow J, Siitonen A, Hasty DL. Quantitative differences in adhesiveness of type 1 fimbriated *Escherichia coli* due to structural differences in *fimH* genes. *J Bacteriol* 1995;177:3680–3686. [PubMed: 7601831]
- Sokurenko EV, Chesnokova V, Doyle RJ, Hasty DL. Diversity of the *Escherichia coli* type 1 fimbrial lectin. Differential binding to mannosides and uroepithelial cells. *J Biol Chem* 1997;272:17880–17886. [PubMed: 9211945]
- Sokurenko EV, Chesnokova V, Dykhuizen DE, Ofek I, Wu XR, Krogfelt KA, et al. Pathogenic adaptation of *Escherichia coli* by natural variation of the FimH adhesin. *Proc Natl Acad Sci USA* 1998;95:8922–8926. [PubMed: 9671780]
- Sokurenko EV, Hasty DL, Dykhuizen DE. Pathoadaptive mutations: gene loss and variation in bacterial pathogens. *Trends Microbiol* 1999;7:191–195. [PubMed: 10354593]
- Sokurenko EV, Tchesnokova V, Yeung AT, Oleykowski CA, Trintchina E, Hughes KT, et al. Detection of simple mutations and polymorphisms in large genomic regions. *Nucleic Acids Res* 2001;29:E111. [PubMed: 11713331]
- Sokurenko EV, Feldgarden M, Trintchina E, Weissman SJ, Avagyan S, Chattopadhyay S, et al. Selection footprint in the FimH adhesin shows pathoadaptive niche differentiation in *Escherichia coli*. *Mol Biol Evol* 2004;21:1373–1383. [PubMed: 15044596]
- Stapleton A, Moseley S, Stamm WE. Urovirulence determinants in *Escherichia coli* isolates causing first-episode and recurrent cystitis in women. *J Infect Dis* 1991;163:773–779. [PubMed: 1672702]
- Staudinger BJ, Oberdoerster MA, Lewis PJ, Rosen H. mRNA expression profiles for *Escherichia coli* ingested by normal and phagocyte oxidase-deficient human neutrophils. *J Clin Invest* 2002;110:1151–1163. [PubMed: 12393851]
- Tanaka T, Nei M. Positive darwinian selection observed at the variable-region genes of immunoglobulins. *Mol Biol Evol* 1989;6:447–459. [PubMed: 2796726]
- Thomas WE, Trintchina E, Forero M, Vogel V, Sokurenko EV. Bacterial adhesion to target cells enhanced by shear force. *Cell* 2002;109:913–923. [PubMed: 12110187]
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 1997;25:4876–4882. [PubMed: 9396791]
- Tullus K, Kuhn I, Orskov I, Orskov F, Mollby R. The importance of P and type 1 fimbriae for the persistence of *Escherichia coli* in the human gut. *Epidemiol Infect* 1992;108:415–421. [PubMed: 1350997]
- Welch RA, Burland V, Plunkett G 3rd, Redford P, Roesch P, Rasko D, et al. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc Natl Acad Sci USA* 2002;99:17020–17024. [PubMed: 12471157]
- Wold AE, Caugant DA, Lidin-Janson G, de Man P, Svanborg C. Resident colonic *Escherichia coli* strains frequently display uropathogenic characteristics. *J Infect Dis* 1992;165:46–52. [PubMed: 1727897]

Fig. 1.

Unrooted phylogram of MLST loci of *E. coli* K1 and non-K1 archetypal strains. Based on concatenated 500 bp fragments of MLST housekeeping loci *adh*, *fumC*, *gyrB*, *icd*, *mdh*, *metG*, *purA* and *recA*. Branch lengths reflect absolute nucleotide differences between concatenated sequences. Circled numbers indicate number of strains at specified node. Letters in parentheses indicate major phylogenetic group assignment. Grey box indicates serotypes belonging to ST95 and ST390. CFT073, 536 and J96 are archetypal extraintestinal pathogenic *E. coli* strains. MG1655 and JVEC9 represent out-group strains.

Fig. 2.

Unrooted phylogram of *fimA* sequences from K1 and non-K1 archetypal strains. Labelling conventions same as for Fig. 1, with the following additions. Non-ST95 strains with sequences identical to complex strains are placed next to, but outside of, grey boxes, without branches. Letters in italics (with and without subscripts) indicate alleles as discussed in text.

Fig. 3.
Unrooted phylogram of *fimH* alleles from K1 and non-K1 archetypal strains. Figure conventions same as for Figs 1 and 2.

Fig. 4. CEL-1 digests of homo- and heteroduplexes of *fim* clusters cloned from three *E. coli* O18:K1:H7 strains with distinct *fimH* alleles. A. *fim* clusters prior to CEL-1 digestion. Lanes 1–3: homoduplexes of F3, RS218 and 89–1449; lanes 4–6: heteroduplexes of 89–1449/F3, RS218/F3 and 89–1449/RS218. B. *fim* clusters after CEL-I digestion (lane numbering same as for A). White arrowheads indicate fragments generated by CEL-1 cleavage at heteroduplex SNP sites, with similar pattern observed for all three pairs.

Fig. 5. Monomannose binding properties of distinct *fimH* alleles expressed in isogenic background. Monomannose binding without inhibitor shown in black bars, with inhibitor in grey. Alleles identified both by letter/subscript terminology used in text and by defining mutations (relative to primary allele in each lineage). SNP-containing allele lineages *b*₀, *c*₀ and *d*₀ demarcated by vertical lines. Error bars indicate standard error of the mean.

Fig. 6. Primary vaginal epithelial cell binding properties of variant *fimH* alleles expressed in isogenic background, plotted against monomannose binding strength. R^2 indicates slope of trend line. Error bars indicate standard error of the mean associated with vaginal cell binding data.

Fig. 7.

Unrooted phylogram of *papA* alleles. PapA pilin variants, serologic F-types F10 and F11. Other F-types indicated in parentheses. Circled numbers indicate number of strains at specified node. Branch lengths reflect absolute nucleotide differences between sequences. Grey boxes indicate serotypes belonging to ST95. CFT073, 536 and J96 are archetypal extraintestinal pathogenic *E. coli* strains. Note that two CFT073 alleles are included.

Fig. 8. Unrooted phylogram of *papG* alleles. PapG adhesin variants, classes II and III. Figure conventions otherwise same as for Fig. 7.

Table 1

Escherichia coli ST95 complex and non-complex archetypal strains.

Strain	Serotype	ST	Country	Site	<i>fim</i>			<i>pap</i>		Reference
					A	H		A	G	
JVEC37	O1:K1:H1	95	JP	V	C ₀	c ₄		F11	II	Obata-Yasuoka <i>et al.</i> (2002)
JVEC49	O1:K1:H1	95	JP	V	C ₀	c ₄		F11	II	Obata-Yasuoka <i>et al.</i> (2002)
JVEC59	O1:K1:H1	95	JP	V	C ₀	c ₄		F11	II	Obata-Yasuoka <i>et al.</i> (2002)
JVEC71	O1:K1:H1	95	JP	V	C ₀	c ₄		F11	II	Obata-Yasuoka <i>et al.</i> (2002)
JVEC84	O1:K1:H1	95	JP	V	C ₀	c ₅		—	—	Obata-Yasuoka <i>et al.</i> (2002)
JVEC48	O1:K1:H7	95	JP	V	B ₀	b		F11	II	Obata-Yasuoka <i>et al.</i> (2002)
JVEC72	O1:K1:H7	95	JP	V	B ₀	b' ₀		F11	II	Obata-Yasuoka <i>et al.</i> (2002)
JVEC74	O1:K1:H7	95	JP	V	B ₀	b' ₀		F11	II	Obata-Yasuoka <i>et al.</i> (2002)
JVEC86	O1:K1:H7	95	JP	V	B ₀	b' ₀		F11	II	Obata-Yasuoka <i>et al.</i> (2002)
JVEC87	O1:K1:H7	95	JP	V	B ₀	b' ₀		—	—	Obata-Yasuoka <i>et al.</i> (2002)
Cy179	O1:K1:H7	95	US	Cy	B ₁	b' ₁		F11	II	Sokurenko <i>et al.</i> (2004)
U5/41	O1:K1:H7	95	DK	Cy	D	d ₀		F11	II	Staudinger <i>et al.</i> (2002)
Py90	O1:K1:H7	95	US	Py	D	d ₁		F11	II	Sokurenko <i>et al.</i> (2004)
Py160	O1:K1:H7	95	US	Py	B ₀	b		F11	II	Sokurenko <i>et al.</i> (2004)
JVEC9	O1:K1:NM	59	JP	V	C-like	b' ₀		n/a	n/a	Obata-Yasuoka <i>et al.</i> (2002)
JVEC10	O2:K1:H6	141	JP	V	JVEC10	JVEC10		n/a	n/a	Obata-Yasuoka <i>et al.</i> (2002)
JVEC30	O2:K1:H7	95	JP	V	B ₀	b' ₀		F11	II	Obata-Yasuoka <i>et al.</i> (2002)
JVEC33	O2:K1:H7	95	JP	V	B ₀	b' ₀		F11	II	Obata-Yasuoka <i>et al.</i> (2002)
JVEC52	O2:K1:H7	95	JP	V	B ₀	b' ₀		F11	II	Obata-Yasuoka <i>et al.</i> (2002)
JVEC57	O2:K1:H7	95	JP	V	B ₀	b' ₀		F11	II	Obata-Yasuoka <i>et al.</i> (2002)
JVEC66	O2:K1:H7	95	JP	V	B ₀	b' ₀		F11	II	Obata-Yasuoka <i>et al.</i> (2002)
JVEC90	O2:K1:H7	95	JP	V	B ₀	b' ₀		F11	II	Obata-Yasuoka <i>et al.</i> (2002)
H25	O2:K1:H7	95	US	U	B ₀	b' ₀		F11	II	Obata-Yasuoka <i>et al.</i> (2002)
H35	O2:K1:H7	95	US	U	B ₀	b' ₀		F11	II	Johnson <i>et al.</i> (2000)
JVEC39	O2:K1:NM	95	JP	V	A	a		F11	II	Obata-Yasuoka <i>et al.</i> (2002)
JVEC13	O18:K1:H7	95	JP	V	C ₀	c ₁		F10	III	Obata-Yasuoka <i>et al.</i> (2002)
JVEC47	O18:K1:H7	95	JP	V	C ₀	c ₁		F10	III	Obata-Yasuoka <i>et al.</i> (2002)
JVEC62	O18:K1:H7	95	JP	V	C ₀	c ₁		F10	III	Obata-Yasuoka <i>et al.</i> (2002)
JVEC73	O18:K1:H7	95	JP	V	C ₀	c ₁		F10	III	Obata-Yasuoka <i>et al.</i> (2002)
IHE3034	O18:K1:H7	95	FI	NBM	C ₀	c ₁		—	—	Pouttu <i>et al.</i> (1999)
IHE3035	O18:K1:H7	95	FI	NBM	C ₀	c ₁		—	—	Pouttu <i>et al.</i> (1999)
IHE3040	O18:K1:H7	390	FI	NBM	C ₀	c ₀		—	—	Pouttu <i>et al.</i> (1999)
IHE3047	O18:K1:H7	390	FI	NBM	C ₀	c ₀		—	—	Pouttu <i>et al.</i> (1999)
IHE3080	O18:K1:H7	390	FI	NBM	C ₀	c ₀		—	—	Pouttu <i>et al.</i> (1999)
89-1449	O18:K1:H7	95	NL	NBM	C ₀	c ₀		—	—	Johnson <i>et al.</i> (2001a)
2H25	O18:K1:H7	95	US	U	C ₁	c ₃		F10	III	Johnson <i>et al.</i> (2000)
H16	O18:K1:H7	95	US	NBM	C ₀	c ₀		—	—	Johnson <i>et al.</i> (2001a)
C5	O18:K1:H7	95	US	NBM	C ₀	c ₁		F10	III	Johnson <i>et al.</i> (2001a)
F3	O18:K1:H7	95	US	NBM	C ₀	c ₂		F10	III	Stapleton <i>et al.</i> (1991)
Nu14	R:K1:H7	95	US	Cy	C ₀	c ₁		F10	III	Johnson <i>et al.</i> (2001b)
RS218	O18:K1:H7	95	US	NBM	C ₀	c ₁		F10	III	Johnson <i>et al.</i> (2001a)
S39	O18:K1:H7	95	US	F	C ₀	c ₁		F10	III	Johnson <i>et al.</i> (2001a)
S98	O18:K1:H7	95	US	F	C ₀	c ₁		F10	III	Johnson <i>et al.</i> (2001a)
U64	O18:K1:H7	95	US	NBM	C ₀	c ₁		F10	III	Johnson <i>et al.</i> (2001a)
VA1242	O18:K1:H7	95	US	Ca	C ₀	c ₁		F10	III	Sokurenko <i>et al.</i> (2004)

Strain	Serotype	ST	Country	Site	<i>fim</i>		<i>pap</i>		Reference
					A	H	A	G	
F-18	R:K1:H5	95	US	F	C ₀	c ₀	F11	II	Krogfelt <i>et al.</i> (1991)
196	O4:H5	12	US	Py	A-like	a	F13	III	Hull <i>et al.</i> (1981)
CFT073	O6:K2:H1	73	US	Py	CFT073	CFT073	F7-1/2	II	Welch <i>et al.</i> (2002)
536	O6:K15:H31	127	DE	Py	536	536	F536	III	Berger <i>et al.</i> (1982)
MG1655	n/a	98	US	F	A-like	a	—	—	Blattner <i>et al.</i> (1997)

NM, non-motile; R, rough LPS; n/a, not applicable or not available; ST, sequence type (per <http://web.mpiib-berlin.mpg.de/mlst/dbs/Ecoli/>); JP, Japan; US, United States; DK, Denmark; FI, Finland; NL, the Netherlands; DE, Germany; V, vagina; Cy, cystitis; Py, pyelonephritis; U, urosepsis; NBM, neonatal bacterial meningitis; F, feces; Ca, catheter tip. See text and Figs 2 and 3 for assignment of *fim* alleles.