

# Selection Footprint in the FimH Adhesin Shows Pathoadaptive Niche Differentiation in *Escherichia coli*

Evgeni V. Sokurenko,\*<sup>1</sup> Michael Feldgarden,†, Elena Trintchina,\* Scott J. Weissman,\* Serine Avagyan,\* Sujay Chattopadhyay,\* James R. Johnson,‡, and Daniel E. Dykhuizen†

\*Department of Microbiology, University of Washington, Seattle; †Department of Ecology and Evolution, SUNY at Stony Brook; ‡VA Medical Center and Department of Medicine, University of Minnesota Medical School, Minneapolis

Spread of biological species from primary into novel habitats leads to within-species adaptive niche differentiation and is commonly driven by acquisition of point mutations in individual genes that increase fitness in the alternative environment. However, finding footprints of adaptive niche differentiation in specific genes remains a challenge. Here we describe a novel method to analyze the footprint of pathogenicity-adaptive, or pathoadaptive, mutations in the *Escherichia coli* gene encoding FimH—the major, mannose-sensitive adhesin. Analysis of distribution of mutations across the nodes and branches of the FimH phylogenetic network shows (1) zonal separation of evolutionary primary structural variants of FimH and recently derived ones, (2) dramatic differences in the ratio of synonymous and nonsynonymous changes between nodes from different zones, (3) evidence for replacement hot-spots in the FimH protein, (4) differential zonal distribution of FimH variants from commensal and uropathogenic *E. coli*, and (5) pathoadaptive functional changes in FimH brought by the mutations. The selective footprint in *fimH* indicates that the pathoadaptive niche differentiation of *E. coli* is either in its initial stages or undergoing an evolutionary “source/sink” dynamic.

## Introduction

When a population spreads from its evolutionarily primary niche into a new habitat, some genes will not be optimally adapted to this new environment. Provided that the population can maintain itself for a period of time in the novel habitat, advantageous mutations in these genes will be selected leading to differentiation within the species. Niche differentiation is a fundamental biological process, representing the first step in formation of new species (Orr and Smith 1998). Niche differentiation can also lead to the emergence of pathogenic microbial clones from relatively benign lineages of the same species, because host compartments in which infection takes place are commonly separate from the principal habitat of the species (Levin and Bull 1994). Mutational gene changes that increase fitness of microorganisms as pathogens are called pathoadaptive mutations (Sokurenko, Hasty, and Dykhuizen 1999). It remains a challenge, however, to predict which genes might be subject to (patho)adaptive mutation and then selection in the course of niche differentiation (Page and Holmes 1998). Limited information about genes under selection in the new habitat also impedes our understanding of the DNA “footprints” that indicate a gene is undergoing (patho)adaptive selection. At its initial stages, in particular, the adaptive niche differentiation might involve numerous populations evolving independently of one another at relatively few genes, selecting only a few mutations. Studying sequence polymorphisms of genes expected to be involved in the niche differentiation might provide an insight on the footprints of selection under these conditions. Here we report a novel type of selection footprint within the gene encoding FimH, the major adhesive protein of *Escherichia coli*, mutations in which are pathoadaptive for uropatho-

genic *E. coli* clones and contribute to niche differentiation of the species.

The large intestine of healthy individuals provides the primary niche for *E. coli* in humans. *E. coli* strains, however, are also associated with a variety of diseases at extraintestinal sites (Johnson and Russo 2002), particularly the urinary tract, which may be considered the alternative niche of the species. Among both commensal and uropathogenic *E. coli* populations, the vast majority of strains are capable of expressing type 1 fimbriae—hair-like, adhesive appendages present in the hundreds on the bacterial cell surface (Brinton 1959). At the tip of each fimbria is the FimH protein, the 30-kDa lectin-like adhesin that determines mannose-sensitive binding of bacteria to target cells (Klemm and Christiansen 1987). For intestinal *E. coli*, FimH contributes to fecal/oral transmission by mediating transient colonization of the oropharyngeal epithelium (Bloch, Stocker, and Orndorff 1992). For uropathogenic *E. coli*, FimH is a critical determinant of tropism for the urinary tract epithelium (Hung et al. 2002). Thus, FimH adhesin is important in colonizing strikingly different niches of *E. coli*, primary and alternative alike, and provides a good model for studying role of a single gene adaptation in species niche expansion. Indeed, it was shown that different structural variants of FimH vary in the strength of their binding to uroepithelial cells (Sokurenko et al. 1995, 1997). The strength of uroepithelial cell binding depends on the adhesins’ ability to bind cell receptors that contain single terminal mannosyl, or monomannose, residues (Sokurenko et al. 1997). The high monomannose-binding capability of urotropic FimH variants in turn depends on the presence of point structural mutations in the *fimH* gene (Sokurenko et al. 1995). These replacement mutations are of a diverse nature and span the protein. FimH mutations were shown to provide significant advantage to bacteria in the colonization of the urinary bladder in a murine model (Sokurenko et al. 1998) and to correlate with extraintestinal virulence of *E. coli* (Hommais et al. 2003). Therefore, the *fimH* replacements belong to the class of pathoadaptive mutations, i.e., gene changes

Key words: bacterial pathogens, niche differentiation, selection footprint.

E-mail: evs@u.washington.edu.

*Mol. Biol. Evol.* 21(7):1373–1383. 2004

doi:10.1093/molbev/msh136

Advance Access publication March 24, 2004

that enhance microbial virulence (Sokurenko, Hasty, and Dykhuizen 1999).

We hypothesize that evolution of the FimH adhesin might reflect an ongoing adaptive niche expansion of *E. coli* coupled with the increased uropathogenicity. If so, it should leave some footprint in the phylogeny of *fimH* alleles that might clarify the stage and evolutionary dynamics of this expansion. We expect that, depending on the overall selection on the alleles and the time since the initial expansion into new habitat, different phylogenetic patterns in the selected gene would emerge. The first outcome is balanced polymorphism, in which differentially adapted alleles are maintained for extended periods of time by different selection pressures (Kreitman and Hudson 1991). This leads to an overall excess of polymorphism that could be detected by specific molecular evolutionary tests (e.g., it would give a significantly positive value to Tajima *D*'s statistics or Fu and Li *D*\* statistics tests) (Tajima 1989; Fu and Li 1993). The second is allelic replacement, wherein newly adapted alleles confer a higher overall fitness in the new and primary habitats and ultimately replace the primary alleles entirely (Kreitman and Hudson 1991). It will behave like a selective sweep purging allelic variation (this should give a significantly negative value to the Tajima *D*'s statistics or Fu and Li *D*\* statistics tests). The third possible outcome is an "evolutionary source-sink" process (Pulliam 1988). In this process, novel alleles that are adaptive in a secondary habitat (the "sink") continuously emerge from the primary pool of alleles (the "source") but their overall long-term fitness across all habitats is lower than the primary alleles. Thus, from an evolutionary perspective, this would lead to the relatively short persistence and rapid extinction of the newly adapted alleles.

To understand whether there is a specific selection footprint of the evolution of the FimH adhesin of *E. coli* we have analyzed DNA variation patterns in *fimH* alleles from *E. coli* isolates of commensal and extraintestinal pathogenic origin. As a control, we have analyzed in parallel variations in the gene encoding the molecular chaperone of type 1 fimbriae, FimC, that is not expressed on the surface of bacteria and, thus, is unlikely to be under selection that affects receptor-binding properties of the fimbriae.

## Materials and Methods

### Strain Collection

To avoid selection bias, isolates included in the present study were selected systematically from larger collections.

Twenty-eight fecal isolates from healthy adults were collected from three different groups of volunteers without signs or symptoms of *E. coli* infection. The groups included women receiving (or eligible to receive) care at the University of Minnesota Student Health Center; employees of the Minneapolis VA Medical Center and their household members; and female patients at a family practice clinic in St. Paul, Minnesota. Fecal samples were collected and processed to isolate *E. coli* as previously described (Johnson et al. 1998).

Fourteen cystitis isolates were recovered from the urine of women seen at the University of Minnesota Student Health Center with clinically diagnosed acute cystitis plus microscopic pyuria. Sixteen pyelonephritis isolates were recovered from the urine of women with uncomplicated pyelonephritis of mild-to-moderate severity during a multi-center treatment trial conducted in the mid-1990s (Talan et al. 2000; Johnson et al. 2002). Fourteen urosepsis isolates were blood isolates from patients with bacteremia of urinary tract origin, as previously published (Johnson and Stell 2000). In addition to the UTI strains, forty non-urinary extraintestinal clinical isolates were studied that were collected at the Minneapolis VA Medical Center. This included eighteen sepsis isolates from patients without UTI or pulmonary infection, fourteen strains isolated from the blood or sputum of patients with pulmonary infection, eight wound and five catheter tip isolates from non-bacteremic patients.

In addition to the clinical isolates, we used the sequences from the following archetypal strains: human intestinal isolate F-18 (Krogfelt et al. 1991); cystitis isolate NU14 (Schaeffer 2002); strain F3 from a patient with recurrent cystitis (Stapleton, Moseley, and Stamm 1991); cystitis isolate PY1013, representing a recently identified, fast-spreading trimethoprim-sulfamethoxazole-resistant clonal group of uropathogenic *E. coli* (Manges et al. 2001; Johnson et al. 2002); and model pyelonephritis strains CFT073, 536 and J96. In addition, *fimH* and *fimC* gene sequences were determined for 9 *E. coli* reference (ECOR) strains—ECOR 1, 2, 28, 38, 42, 52, 61, 64, and 72—representing the major phylogenetic branches of *E. coli*.

In summary, both *fimH* and *fimC* gene sequences were obtained from 115 isolates, and *fimH* from 18 additional isolates. The use of the additional *fimH* sequences provided additional power for the analysis of the distribution of FimH variants among strains of different origin, but it did not affect the comparative analysis of *fimH* and *fimC* genes.

### Sequence Analysis

Sequences for *fimH* genes from archetypal strains CFT073 and J96 were obtained from GenBank, and *fimH* sequences of strains F18 and NU14 were reported previously (Sokurenko et al. 1998). *fimH* sequences from the remaining strains and all *fimC* sequences were determined in this study by standard methods. The genes were sequenced by PCR amplification. The following primers were used for the *fimH* genes: FIMH3'-42:CGTGCAGGT-TTTTAGCTTCA; FIMH5'-49:TCAGGGAACCATTCAGGCA; FIMH5'-12:ACCTACAGCTGAACCCGAAG; FIMH3'-(-21):TTATTGATAAACAAGTCAC; FIMH5'-INT:GGTATTACCTCTCCGGCACA; FIMH3'-INT:GACGCGGTATTGGTGAAAAT. (The usual primers for PCR of *fimH* are the FIMH5'-49 and FIMH3'-42, with the FIMH5'-12 used sometimes and the FIMH3'-(-21) used a few times. The numbers represent the number of bases from the end of the primer to the beginning of gene sequence. The two primers marked INT are internal primers. These were used to sequence the entire *fimH* in

both directions.) The following primers were used for the *fimC* genes: FIMC5'-65:CAGGCCTGGTTCTCTTTA-ACC; FIMC3'-44:CCCGGCAGTCAATTCTTTT. (The two *fimC* primers were used on all strains and internal primers were not needed.). The method of analysis proposed in this paper, zonal analysis, emphasizes the differences in related sequences and consequently emphasizes sequencing errors. Thus, all sequences were done in both directions and all singletons were checked, often by resequencing.

ClustalW alignment of the gene and protein sequences was performed using MacVector 6.5.3. DNA polymorphism analysis was performed using DnaSP 3.53 software.

Construction of phylogenetic trees of FimH and FimC protein variants was based on maximum likelihood phylogenetic trees (unrooted phylograms) of the *fimH* and *fimC* genes, respectively, using the PAUP\* 4.0b software package. DNA trees were built using the General Time-Reversible model with estimated base frequencies site-specific by codon position distribution. Substitution rates were obtained from the sequence data. Molecular clock and topological constraints were not enforced. To conserve computing time, duplicate sequences were removed from the input sample. When a single maximum likelihood tree was obtained for each gene, branches containing only silent changes were collapsed, leaving branches that contained either replacement changes only or both replacement and silent mutations. In this way, structurally identical FimH variants that emerged independently are presented as separate nodes on the tree. The creation of the zones was done by hand for this paper. A programmed version will be developed.

Calculation of nonsynonymous and synonymous variations at single codon sites was done using ADAPT-SITE 1.2 software based on a previously described method (Suzuki and Gojobori 1999).

#### Determination of Monomannose-Binding Properties

The monomannose-binding capability of *E. coli* strains was determined essentially as described previously (Sokurenko et al. 1997). In brief, the expression of type 1 fimbriae was locked "on" by transforming strains with a pPKL9/91 plasmid encoding the positive regulator of type 1 fimbrial expression, FimB. The transformed strains were radiolabeled by growing bacteria overnight in Luria broth containing 3H-thymidine, and they were tested for mannose-sensitive binding to yeast mannan (the model monomannose-like substrate) in a microtiter-plate assay as described previously.

#### Statistical Analysis

The G-test (adjusted by William's correction) and, where necessary, two-tailed Fisher's Exact test were used to evaluate whether number distributions between the groups were different from chance. The Wilcoxon two-sample test was used to compare values of monomannose-binding between different isolate groups.

## Results

### DNA Polymorphism in *fimH* and *fimC* Genes

One of the traditional approaches used to detect the footprint of adaptive protein evolution is to determine rates of silent ( $K_s$ ) and replacement ( $K_a$ ) nucleotide substitutions, where  $K_a > K_s$  provides an indication of positive selection for protein modification (Nei and Gojobori 1986). However, though this method provides unambiguous proof for selection, it is highly conservative (Sharp 1997). We have analyzed DNA polymorphism patterns for *fimH* (900 bp) and *fimC* (720 bp), encoding the type 1 fimbrial adhesive subunit and molecular chaperone, respectively, from *E. coli* isolates of fecal and extraintestinal pathogenic origin.

Among 133 *fimH* genes sequenced, 63 distinct allelic variants were identified, with 96 unique mutations found at 89 polymorphic sites resulting in a total nucleotide diversity of  $1.64 \pm 0.07\%$  (table 1). All mutations were point substitutions, i.e., single nucleotide polymorphisms. The corresponding analysis of *fimC* genes from the same strains (115 strains) identified 40 distinct alleles, with 45 unique mutations (all point substitutions) found at 43 polymorphic sites and resulting in a total nucleotide diversity of  $1.11 \pm 0.03\%$ . Thus, the level of diversity in both genes (measured by Theta; see table 1) is within the range found in housekeeping genes (Pupo et al. 1997).

Among the *fimH* mutations, 28 were amino acid replacements and 68 were silent substitutions. Overall, the rate of replacement substitutions ( $K_a = 0.0042$ ) in *fimH* genes was significantly lower than the rate of silent mutations ( $K_s = 0.052$ ), with  $K_a/K_s = 0.081$  (table 1). In *fimC*,  $K_a$  and  $K_s$  values were both somewhat lower than in *fimH*, but the overall  $K_a/K_s$  ratio was similar. In *fimH*, but not *fimC*, the  $K_a/K_s$  ratio for isolates of UTI origin was higher than for isolates of fecal or non-UTI pathogenic origin, suggesting selection for diversity in that environment. However,  $K_a/K_s$  values for *fimH* within either source group were significantly less than 1, and they were similar to values found among *E. coli* housekeeping genes, suggesting purifying selection.

Other traditional (and also conservative) evolutionary tests for selection/neutrality in DNA sequences likewise showed no evidence for the expected diversifying selection in *fimH*. Values for Tajima's *D* statistics and Fu and Li's *D\** statistics for *fimH* and *fimC* did not significantly differ from zero ( $P > .1$ ) for all strains combined (table 1) or for strains of specific origin (not shown).

Thus, the traditional tests failed to provide any evidence for the action of selection, even though previous experimental evidence suggests considerable advantage for change in function of FimH in uropathogenic *E. coli* (Sokurenko et al. 1998). We believe that novel analytical approaches should be employed to search for the putative selection footprint in the FimH adhesin.

#### Distribution of FimH Variants Across the Phylogenetic Tree

We analyzed the distribution of structural variants of the FimH adhesin across the protein phylogenetic tree. In

**Table 1**  
**DNA Polymorphism in *fimH* and *fimC* Genes Among 133 Fecal and Extraintestinal *E. coli* Isolates**

	<i>fimH</i>	<i>fimC</i>
Total number of genes sequenced	133	115
Number of distinct alleles	63	40
Unique mutations (polymorphic sites)	96 (89)	45 (43)
Nucleotide diversity	1.64 ± 0.07%	1.11 ± 0.03%
Replacements	28	10
Silent substitutions	68	35
$K_a/K_s$ = ratio		
All isolates	0.0042/0.052 = 0.081	0.0026/0.038 = 0.069
Fecal isolates	0.0033/0.053 = 0.062	0.0026/0.041 = 0.065
UTI isolates	0.0047/0.053 = 0.089	0.0023/0.036 = 0.065
Non-UTI pathogenic isolates	0.0041/0.053 = 0.077	0.0028/0.038 = 0.076
Theta (for replacements)	0.008	0.003
Tajima D's statistics test	-0.38	-0.16
Fu and Li D* statistics test	-1.18	-0.62

this tree, presented as an unrooted phylogram (fig. 1), the nodes represent specific structural variants of the adhesin, with the connecting branches corresponding to amino acid changes. This tree was constructed from the DNA-based unrooted phylogram as described in *Materials and Methods*. Because of the method of construction of the protein phylogram, structurally identical FimH variants that evolved independently are represented as separate nodes.

A total of 45 distinct (resolved) nodes were identified on the tree, with 43 nodes represented by naturally-occurring FimH variants from our sample (two nodes were resolved but not represented by alleles in our collection). Four additional nodes were hypothetical (i.e., unresolved). Most of the naturally occurring FimH variants (30 of 43) were found in only one study isolate and comprised the single-isolate nodes (singletons). Remaining FimH variants were found in two or more strains and comprised multiple-isolate nodes. The largest (i.e., most populous) node on the tree was represented by a FimH variant found in 19 isolates that also represents a consensus structure for all FimH variants in the sample (the Consensus node).

Based on the occurrence of silent mutations within FimH nodes and structural difference between the FimH variants, all nodes on the protein tree fall into three distinct zones.

#### Primary Zone

This zone was formed by nodes within which silent nucleotide polymorphisms occurred; that is, each primary node variant was encoded by multiple distinct phylogenetically linked *fimH* alleles. (Singletons connecting two primary nodes would also be placed in the Primary zone.) The Primary zone occupied the center of the tree, encompassing 4 nodes (the Consensus node, along with nodes S91, N99, and V263), which were linked together via single-replacement branches and were encoded by 14, 5, 3, and 2 *fimH* alleles, respectively. When the *fimH* gene sequence from *Klebsiella pneumonia* was used as out-group, it rooted the *E. coli* sequences in the primary node S91 (not shown). However, because of the relatively high

divergence of *K. pneumonia* and *E. coli fimH* (about 20% heterogeneity), the most ancestral basal node in the FimH tree cannot be determined reliably.

#### Secondary Zone

This zone was formed by multiple-isolate nodes with no synonymous variation and singletons that were connected to a Primary zone node via a single amino acid replacement. That is, the nodes in this zone represented FimH alleles differing from a corresponding primary FimH variant by only a single amino acid. The Variant zone was formed by 23 distinct nodes immediately surrounding the Primary zone.

#### Extended Zone

This zone was comprised by nodes differing from a Primary zone node by two or more amino acid replacement changes. The Extended zone occupied the outermost area of the tree and consisted of 18 nodes.

The accumulation of synonymous nucleotide diversity within the nodes in the Primary zone indicates that these FimH alleles have a long history in the population and are likely to be under purifying selection against structural variation. At the same time, the lack of synonymous nucleotide diversity within the large multi-strain nodes of the Secondary zone indicates the recent emergence of these FimH variants. Some of these nodes contain isolates of different clonal origin. For example, the largest Secondary node, V48a, contains isolates of at least 7 different serotypes (not shown) and the phylogenetically diverse strains ECOR1, ECOR61, and ECOR64. Therefore, some FimH variants from the Secondary zone presumably have recently spread horizontally among large numbers of *E. coli* strains and, thus, might carry an adaptive value. In contrast, it is difficult to assess the selective value of terminal singleton nodes within the Secondary zone. These nodes may represent recently emerged adaptive variants but could likewise represent neutral or even slightly deleterious FimH variants that are slowly being removed by purifying selection. The Extended variants are structurally the most divergent forms of FimH, and the multiple

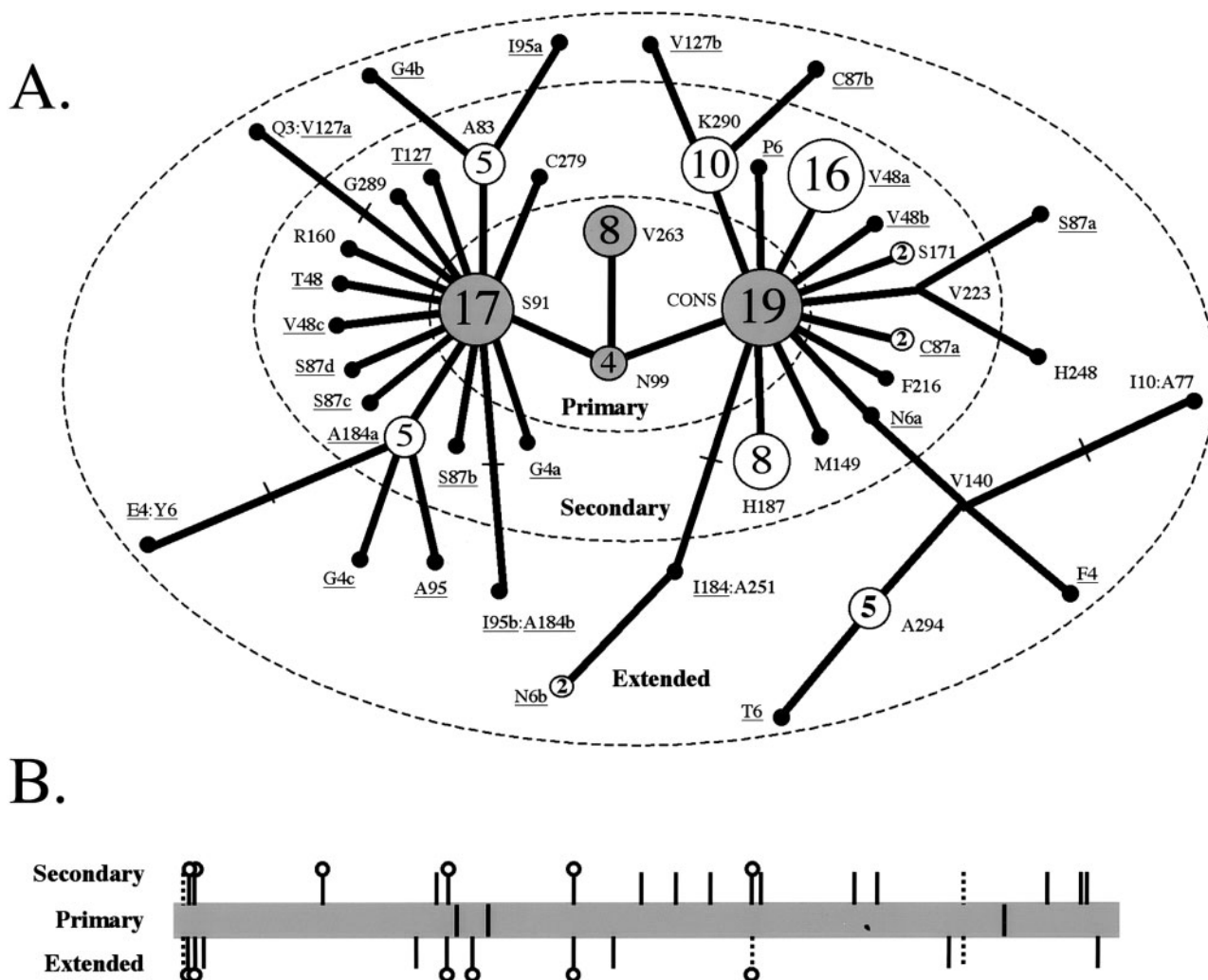


FIG. 1.—(A) Phylogenetic tree of FimH protein variants. CONS—node corresponding to the FimH of consensus structure. All nodes are marked according to the replacement mutation from the consensus structure or the immediate ancestral variant. Replacements of the same amino acid in the same position that were acquired independently are distinguished by lowercase letters (a, b, etc.). Small filled circles represent single-strain nodes. Circles containing numbers represent multiple-strain nodes, and indicate the total number of strains in the collection that carry the corresponding protein variant. Grey circles represent nodes with intranodal synonymous variation. Open circles represent nodes without any synonymous variation. Thin bars mark hypothetical (unresolved) nodes. Nodes formed by parallel or coincidental mutations in “hot-spot” positions are underlined. (B) Distribution of replacement polymorphisms in the FimH protein. The grey bar corresponds to the full length FimH protein. The lines within the bar represent the Primary zone variations. The lines above the bar represent changes found in the Secondary zone. The lines below the bar represent changes found in the Extended zone. Mutational “hot-spot” positions are marked by open circles. The dashed lines represent changes of uncertain location.

amino acid changes suggest that they are selectively advantageous changes.

#### Distribution of Silent and Replacement Changes Along the Branches Across the Phylogenetic Zones

To determine whether the structural diversification in FimH alleles from the Secondary and Extended zones has been adaptive for *E. coli*, we have analyzed the occurrence of silent and replacement substitutions along the replacement branches connecting the nodes from different zones. Importantly, this criterion differs from that used above to place the strains into distinct zones, as here we calculated silent changes between, and not within, the nodes. In other words, estimation of silent polymorphism along the

branches connecting the nodes is distinct from the silent polymorphism within the nodes.

Three replacement branches linking the four Primary nodes with one another had a total of 10 silent substitutions—six between the Consensus and N99 nodes, two between N99 and S91, and two between N99 and V263 (table 2). In contrast, 23 replacement branches connecting the Primary zone nodes and the corresponding Secondary nodes had only 18 silent mutations ( $P = .037$ ). In greater contrast, along the branches connecting the Secondary zone nodes with the corresponding Extended nodes, or the Extended nodes with one another, there were 17 replacements with no silent changes ( $P \ll .01$ ). Such prevalence of replacement over silent changes is significantly higher than expected from a completely neutral accumulation of changes ( $P = .045$ ).

**Table 2**  
**Zonal Distribution of Nodes and Mutational Changes Along the Connecting Branches**

Protein		Phylogenetic Tree Zones		
		Primary	Secondary	Extended
FimH	Number of nodes	4	23	18
	Branch replacements vs. silent mutations	3 vs. 10	23 vs. 18	17 vs. 0 <sup>a</sup>
FimC	Number of nodes	9	9	1
	Branch replacements vs. silent mutations	7 vs. 8	9 vs. 8	2 vs. 0

<sup>a</sup> Three branches connecting primary and extended nodes directly (i.e., without a defined intermediate node in the Secondary zone) were excluded from the analysis as ambiguous.

It is difficult to evaluate a significance of the relative predominance of replacement mutations along the branches connecting the Primary and Secondary zone nodes. This analysis could be biased against the silent replacements, because the Primary and Secondary zones were separated on the basis of intra-nodal silent variations that is present in the former but absent in the latter. At the same time, separation of the Secondary and Extended nodes (as well as of the Extended nodes from each other) does not preclude inclusion in the analysis of silent changes along the branches, which makes the branch analysis unambiguous. Thus, in general, FimH variants in the Extended zone (and possibly some in the Secondary zone) have emerged from the Primary zone variants under positive diversifying selection. Unfortunately, relatively short length of the individual branches does not permit the statistically reliable estimation of their lengths in terms of replacement and silent substitutions by the method described previously

(Zhang, Rosenberg, and Nei 1998), and the identification of particular nodes under selection cannot be done (not shown).

We believe that the analysis of distribution patterns of silent and replacement mutations along the protein tree of FimH indicates that (1) a subpopulation of *E. coli* encodes FimH variants that are primary to the species and under purifying selection against structural changes, and (2) another subpopulation of *E. coli* expresses structural variants of FimH that have evolved from the primary forms by diversifying selection.

#### Distribution of Silent and Replacement Mutations Along the Protein Tree of FimC

We have constructed and analyzed in a similar manner an unrooted phylogram of FimC protein variants (fig. 2A). Overall, FimC variants were structurally less diverse than FimH variants. The 115 FimC variants formed only 18 nodes on the phylogenetic tree (compared to 45 nodes formed by 133 FimH variants,  $P < .01$ ). Furthermore, most (7 of 9) of the multi-isolate nodes exhibited intra-nodal allelic variation, forming a Primary zone similar to that observed with FimH. Most of other FimC nodes were terminal singletons forming a Secondary zone, i.e., derived from a corresponding Primary node by a single replacement. Only one node was located in the Extended zone (compared to 18 of 45 nodes on the FimH protein tree,  $P < .01$ ). This same pattern was seen in eight *E. coli* housekeeping genes (as exemplified by the Mdh tree in fig. 2B) that are considered to evolve in a neutral fashion. Thus, variation within the chaperone, FimC, forms the pattern expected for genes that evolve in a neutral fashion rather than the pattern seen for FimH.

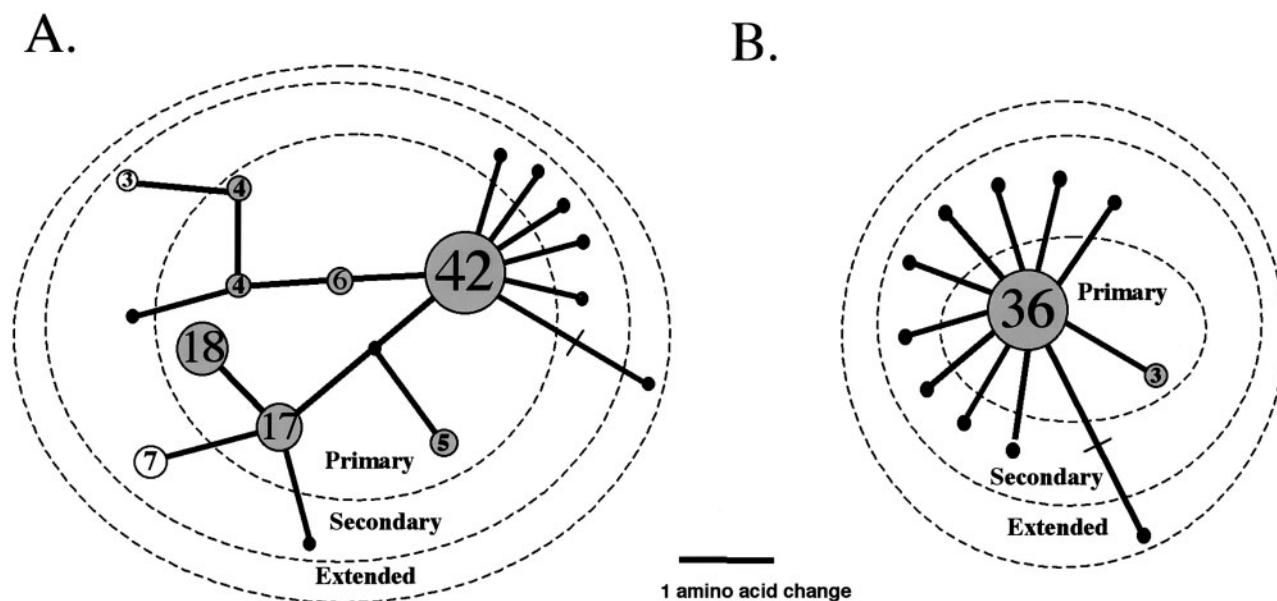


FIG. 2.—(A) Phylogenetic tree of FimC protein variants. Nodes and zones are defined as in figure 1, but specific mutations are not specified to conserve space. (B) Phylogenetic tree of malate dehydrogenase protein variants. Nodes and zones are defined as in figure 1. Partial sequence (452 bp) of *mdh* genes were obtained from a subset of ECOR strain isolates and provided by Mark Achtman, Max-Planck Institut für Infektionsbiologie, Berlin.

Also in contrast to FimH, the distribution of silent and replacement mutations in FimC along the branches connecting the Primary zone nodes (8 and 7 mutations, respectively) was very similar to that for branches connecting the Primary and Secondary zone nodes (8 silent and 9 replacements; table 2). It will be interesting to see if the approximate equal numbers of silent and replacement mutations along branches is expected if the gene evolves via the accumulation of primarily neutral mutations.

### Replacement Hot-Spot Positions in FimH

One characteristic of adaptive mutations is a preference for specific positions in the protein (Hughes and Nei 1988). In FimH, multiple replacements were found at seven amino acid positions (fig. 1A, underlined nodes). These replacement “hot-spot” positions included Val4 (changes to Phe, Glu, and three times to Gly), Thr6 (to Pro, Tyr, and twice to Asp), Ala48 (to Thr and three times to Val), Gly87 (twice to Cys and three times to Ser), Thr95 (to Ala and twice to Ile), Ala127 (to Thr and twice to Val), and Val184 (to Ile and twice to Ala). Though the  $K_a/K_s$  ratio in all codons encoding the hot-spot positions was above 1 (not shown), the overall low number of substitutions in FimH does not allow statistically reliable estimates of the prevalence of replacement over silent mutations at the level of individual codons (Suzuki and Gojobori 1999).

None of the nodes within the Primary zone were formed by replacements in hot-spots. However, among the replacements leading to formation of the Secondary and Extended zone nodes, 15 (60%,  $P = .04$ ) and 14 mutations (78%,  $P < .01$ ), respectively, occurred in the hot-spot positions (two mutations of ambiguous position, V127a and I184, were split between the zones). In contrast to FimH, none of the amino acid positions in the FimC protein had characteristics of mutational hot-spots ( $P < .001$ ). Thus, the existence of hot-spots provides additional evidence for the adaptive evolution of the Extended FimH variants and at least some Secondary zone FimH variants.

### Zonal Distribution of *E. coli* Strains of Different Origin

Uropathogenic isolates were less likely to be in the Primary zone and more likely to be in the Extended zone compared to FimH variants of fecal origin (table 3). This difference was especially prominent for FimH variants from isolates causing the most severe, invasive forms of UTI (pyelonephritis and urosepsis). Furthermore, FimH alleles of uropathogenic origin were more likely to be represented in the Secondary zone by multiple-isolate and hot-spot nodes (22 of 24 isolates) than FimH alleles from fecal strains (8 of 13 isolates,  $P = .035$ ). Such nodes, which included FimH alleles from model uropathogenic strains NU14 (node A83), CFT073 (node A184a), PY1013, and J96 (both node V48a), are more likely to consist of FimH variants carrying adaptive replacements than remaining singleton nodes, which could be comprised of rarely-occurring neutral or slightly deleterious FimH variants. Overall, we found FimH variants with replacements in

**Table 3**  
Zonal Distribution of Strains by Site of Origin

Protein	Isolate origin (n)	Phylogenetic Tree Zones		
		Primary	Secondary	Extended
FimH	All isolates (133)	48	63	22
	Fecal (29)	14	13	2
	All UTI (50)	13 ( $P = .048$ ) <sup>a</sup>	24	13 ( $P = .029$ )
	Pyelonephritis and urosepsis (33)	4 ( $P \ll .01$ )	20	9 ( $P = .033$ )
Non-UTI				
	pathogenic (45)	19	21	5
FimC	All isolates (115)	96 ( $P \ll .01$ )	18	1 ( $P \ll .01$ )

<sup>a</sup> The  $P$  values are relative to the isolates of fecal origin within the same zone (only significant values are indicated).

a hot-spot position (from either Secondary or Extended zones) significantly more often in uropathogenic isolates (24 of 50 isolates) than in fecal isolates (6 of 29 isolates,  $P = .014$ ). FimH variants from extraintestinal isolates of non-UTI origin did not differ in zonal distribution from the FimH variants from fecal isolates.

For FimC, the vast majority of *E. coli* isolates are in the Primary zone (96 of 115 isolates, or 83%; table 3), and there is no correlation of zonal distribution and origins (not shown).

Taken together, the strain distribution analysis shows that uropathogenic isolates are significantly more likely than intestinal commensal isolates to express FimH variants that have evolved under diversifying selection.

### Monomannose-Binding Capability of *E. coli* Strains

*E. coli* expressing Consensus node FimH variants mediated the lowest monomannose binding of any nodal group— $1.33 \times 10 + 6$  cfu/well. In contrast, *E. coli* expressing one of the most structurally divergent FimH variants, Extended zone allele I10:A77, exhibited the highest monomannose-binding— $8.94 \times 10 + 6$  cfu/well. Average monomannose binding among the Primary zone FimH variants was  $1.72 \pm 0.27 \times 10 + 6$  cfu/well. Most strains expressing FimH variants from the Secondary and Extended zones had a higher monomannose-binding capability than strains expressing the primary variants, with an average of  $3.03 \pm 0.42 \times 10 + 6$  cfu/well for the Secondary zone and  $3.68 \pm 0.77 \times 10 + 6$  cfu/well for the Extended zone. The average binding of *E. coli* expressing FimH represented by multiple-isolate and hot-spot nodes was significantly higher ( $3.32 \pm 0.44 \times 10 + 6$  cfu/well,  $P = .028$ ) than that of *E. coli* bearing Primary zone FimH variants. As was reported previously (Schembri, Sokurenko, and Klemm 2000), the monomannose-enhancing substitutions were distributed in different regions of both lectin and pilin domains of the FimH protein and, in general, relatively far from the binding site (not shown). It was proposed recently that these mutations affect conformational properties of FimH rather than the receptor-interacting residues themselves (Thomas et al. 2002).

Therefore, the molecular evolution of FimH leading to the formation of Secondary and Extended zones on the

protein tree has been accompanied by overall increased monomannose-binding capabilities in *E. coli*. It was shown previously that monomannose-binding correlates with the increased level of bacterial adhesion to uroepithelial cells and urinary bladder colonization in murine model of UTI (Sokurenko et al. 1995, 1997; Hung et al. 2002).

## Discussion

Receptor-specific bacterial adhesion is generally necessary for the successful colonization of any niche, and different habitats are likely to differ significantly in the composition and/or structure of surface receptors. Thus, one might expect adhesin genes to be favored targets for adaptive evolution during the expansion of bacterial clones into novel niches. In this study we have shown that the *E. coli* FimH adhesin is undergoing adaptive evolution, and this evolution contributes to niche differentiation of *E. coli* clones and increases uropathogenicity.

The strongest evidence for the action of positive selection upon the FimH adhesin is provided by novel, zonal analysis of the FimH phylogenetic tree, which is based on the separation of evolutionarily ancient nodes containing silent variation (the Primary zone) from nodes representing subsequent evolution (the Secondary and Extended zones). This separation then permits estimation of the combined number of replacement and silent substitutions along the branches that connect the different zones. Collective analysis of all connecting branches (rather than of individual branches or individual codons) makes zonal analysis very sensitive in detecting selective footprints when adaptive mutations (1) are few in number, (2) are scattered across the protein structure, and (3) arise independently in different allelic backgrounds. Under such mutational dynamics, DNA variation patterns among randomly sampled alleles are not particularly distinct from those expected to occur with selectively neutral evolution. Thus, it is not surprising that the determination of a total  $K_a/K_s$  ratio and the test of Tajima or other traditional approaches (Tajima 1989; Nei and Gojobori 1986) failed to uncover the influence of positive selection for structural mutations in the FimH adhesin. The Tajima's test assumes that recurrent mutations do not occur (i.e., an infinite sites model). Fu and Li's test also assumes an infinite site model, and it compares the number of singletons that represent changes on the tips of a phylogeny to the number of polymorphic sites that are not singletons and that represent changes on internal branches. Thus, the high rate of recurrent hot-spot mutation observed in *fimH* violates an important assumption of these tests. The Fu and Li test will underestimate the "tipiness" of *fimH* and consequently miss a significant signature of selection. In contrast to these traditional tests, zonal analysis accentuates this signature of selection.

Evidence from zonal analysis of the Secondary zone is not straightforward. Many of the singletons in this zone are likely to be selectively neutral or slightly detrimental variants that circulate in the *E. coli* populations at low frequency, as suggested by their appearance in the analysis of FimC, Mdh (fig. 2), and seven other MLST loci (data

not shown). However, the hypothesis that many FimH variants from the Secondary zone were selected by adaptive evolution is supported by the finding of multiple strain nodes in the Secondary zone that is not commonly seen in the FimC and housekeeping proteins. Also, most FimH variants from the Secondary zone and, especially, Extended zone were formed by mutations in hot-spot positions. Furthermore, FimH alleles of uropathogenic strains are more often found on the Extended zone nodes as well as multiple-strain and hot-spot nodes in the Secondary zone than FimH alleles of the commensal strains. Finally, the structural evolution of FimH augments the ability of the adhesin to bind monomannose, which represents the main mechanism of bacterial adhesion to uroepithelial cells (Sokurenko et al. 1995, 1997; Hung et al. 2002). Interestingly, a correlation was demonstrated recently between the extraintestinal pathogenicity of *E. coli* and two monomannose-enhancing mutations in FimH—Val48, the most common hot-spot mutation in the Secondary zone, and V140, which forms the largest clade within the Extended zone (Hommais et al. 2003). Thus, we believe that, in general, evolution of the FimH adhesin variants from the Extended and Secondary zones reflects adaptive niche differentiation of *E. coli*.

We would like to note, however, that the adaptive effect of FimH variations might not be limited to increased monomannose-binding. For example, positions Val4 and Thr6 are hot-spot mutations in the leader sequence (the first 21 amino acids) of the nascent protein. These mutations may affect fimbrial length and number rather than receptor specificity (currently under investigation). Furthermore, it is necessary to note that, although extended FimH variants are associated with uropathogenic strains and are pathoadaptive in nature (i.e., enhance *E. coli* urovirulence), it is unclear at this point whether FimH evolution is driven by the ability to cause disease itself or the uropathogenicity is merely a by-product of a different, possibly non-pathogenic, type of adaptive niche differentiation of *E. coli*.

Zonal analysis of *fimH* evolution also provides an insight into the evolutionary dynamics of adaptive niche expansion of *E. coli*. The absence of silent mutations (generally considered to be selectively neutral and to accumulate in random fashion and at constant rate for a given gene) along the FimH tree branches connecting Secondary and Extended nodes suggests that diversifying FimH evolution has occurred too recently for silent mutations to accumulate. By similar reasoning, the formation of large multi-strain, multi-clone nodes within the Secondary zone is also evolutionarily recent. Thus, it appears that adaptive evolution of FimH in the course of *E. coli* niche differentiation has occurred recently. The short-term nature of diversifying FimH evolution argues against the occurrence of a "balanced polymorphism" dynamic in this case, because balancing selection would be expected to accumulate many silent mutations along the branches for the differentially adapted alleles (Kreitman and Hudson 1991). Nor do the evolutionary dynamics observed in this zonal analysis fit the "population replacement" mode of niche expansion characterized by selective sweeps in the course of adaptive evolution (Kreitman and Hudson 1991). In



fact, in our sample, the extended FimH variants co-exist with a pool of primary variants.

We believe that the FimH variation footprint identified here fits instead with yet a third pattern of niche expansion—"source/sink" habitat dynamics. "Source/sink" habitats have been proposed as an ecological model with two fundamental requirements: (1) a stable primary niche (or "source" habitat), to which the species is well adapted and in which it maintains populations over a long period of time; and (2) the existence of alternative niches (or "sink" habitats) into which the organisms can spread, but in which, for one reason or another, they do not maintain a stable population (Pulliam 1988). Thus, genes involved in this "source/sink" adaptation will be constantly adapting to the alternative niches, but the adapted forms will also constantly become extinct. As a result, primary alleles well adapted to the main niche will contain silent variation, whereas those adapted to the alternative niches will not. This is in accordance with our data, where commensal *E. coli* from the principal, intestinal niche commonly expresses evolutionarily-stable, primary forms of FimH, while extraintestinal uropathogenic *E. coli* primarily express recently evolved, and obviously unstable in long-term, FimH variants. One needs to consider, however, an alternative explanation for the recent origin of adapted FimH variants. It is possible that novel habitats providing selective conditions for FimH evolution have become available to the *E. coli* clones relatively recently and that the endpoint dynamics of the niche differentiation has not yet emerged.

It was previously shown that the increased mannose-binding of FimH is accompanied by its increased susceptibility to inhibition by soluble mannosylated compounds, including salivary glycoproteins and intestinal mucin (Sokurenko et al. 1995, 1998). Thus, FimH mutations advantageous in the urinary tract (or other alternative niches) are likely to be selected against in the course of oral transmission and/or intestinal colonization of *E. coli* and to result in the evolutionary instability of uropathogenic *E. coli* clones. This supports the hypothesis that the adaptation to the UTI is a source-sink dynamic that has continued for considerable time.

With whole genomes becoming available for multiple strains within a species, can genes important in virulence be identified from the sequence data? It is easy to determine virulence factors by present-absent tests. Consequently, the acquisition of novel genetic material through horizontal gene transfer (e.g., pathogenicity islands) has been the major focus for studies of the evolution of bacterial virulence and niche expansion (Ochman and Moran 2001). It remains to be seen which acquisition of novel genetic material, through horizontal transfer or change of the genetic material already present, is more important in the evolution of virulence. We propose that zonal analysis will be a useful method for determining which genes are adapting to the pathogenic niche and thus be important in answering the above question.

Zonal analysis is, obviously, a promising method to determine selection acting on mutational changes in a gene. We do not know yet whether it will work as well for detecting adaptive recombination in bacterial genes. How-

ever, other methods are now available for finding the footprints of selection on recombination from a phylogenetic perspective. The relative rates of recombination can be determined by partitioning the species into clones and studying their diversification (Guttman and Dykhuizen 1994). Both MLEE (Multi Locus Enzyme Electrophoresis) and MLST (Multi Locus Sequence Typing) have shown most bacterial species contain high levels of diversity and that the diversity is in strong linkage disequilibrium (Feil and Spratt 2001). This disequilibrium is generated by the outgrowth of clones. In MLST eight gene fragments from eight well spaced cellular or "housekeeping" loci are chosen for DNA sequencing (Maiden et al. 1998). These loci are presumably not undergoing any selection, and any genetic changes, either by mutation or recombination, are selectively neutral as the clones diversify. Since recombination tends to introduce only short fragments, the resulting recombinant strains will be the same as the ancestor at most genes. Thus, complexes of clones can be defined. BURST algorithm has been developed recently to assign strains to clonal complexes and to determine the ancestral genotype of each clone and the variant alleles (Feil et al. 2001). Using this information, an estimation of the rates of mutation and recombination can be generated (Feil and Spratt 2001; Feil et al. 2003). Genes that show unusually high rates of recombination within a clone are candidate pathoadaptive loci. One might expect this system to flag loci under balancing selection, particularly those driven by the immune system. High rates of recombination have been seen for the TbpB (transferrin binding protein B) in *Neisseria meningitidis* (Linz et al. 2000) and the OspC gene in *Borrelia burgdorferi* (Dykhuizen and Baranton 2001), both important antigens, and for the MutS in *Escherichia coli* (Denamur et al. 2000), which promotes rapid adaptation by elevating the mutation rate.

Multiple genomes of a single species are becoming available. When enough complete genomes have been done, we should be able to use zonal analysis and analysis of clonal complexes to discover presumptive pathoadaptive loci. How many genomes will be required is still unknown. The zonal analysis was done with about 120 strains and the analysis of clonal complexes usually uses about 500–600 strains. For these to be practical methods for whole genome analysis, we expect that the analyses will have to be robust for tens of strains. Future work should be done to see how robust these methods are.

Discoveries of genes like *fimH* that are important in pathoadaptation will provide a more balanced understanding of the molecular basis of microbial pathogenesis and, in general, of niche differentiation events. Zonal analysis of gene sequences described here could prove to be a valuable approach for the identification of loci under selection.

## Acknowledgments

We thank Steve Moseley, Colin Manoil, and Kelly Hughes for valuable discussions and suggestions on improving the quality of manuscript. We also would like to thank Li Hao from Pennsylvania State University,

University Park, for the help with testing *fimH* sequences using the ADAPTSITE 1.2 software and Mark Achtman for providing MLST sequences to perform zonal analysis with housekeeping genes. The research was supported by grants from the National Institutes of Health and the National Science Foundation.

## Literature Cited

- Bloch, C., B. Stocker, and P. Orndorff. 1992. A key role for type 1 pili in enterobacterial communicability. *Mol. Microbiol.* **6**:697–701.
- Brinton, C. C. Jr. 1959. *Nature* (London) **183**:782–786.
- Denamur, E., G. Lecointre, P. Darfu et al. (12 co-authors). 2000. Evolutionary implications of the frequent horizontal transfer of mismatch repair genes. *Cell* **103**:711–721.
- Dykhuizen, D. E., and G. Baranton. 2001. The implications of a low rate of horizontal transfer in *Borrelia*. *Trends Microbiol.* **9**:344–350.
- Feil, E. J., J. E. Cooper, H. Grundmann et al. (12 co-authors). 2003. How clonal is *Staphylococcus aureus*? *J. Bact.* **185**:3307–3316.
- Feil, E. J., E. C. Holmes, D. E. Bessen et al. (12 co-authors). 2001. Recombination within natural populations of pathogenic bacteria: short-term empirical estimates and long-term phylogenetic consequences. *Proc. Natl. Acad. Sci. USA* **98**:182–187.
- Feil, E. J., and B. G. Spratt. 2001. Recombination and the population structures of bacterial pathogens. *Annu. Rev. Microbiol.* **55**:561–590.
- Fu, Y. X., and W. H. Li. 1993. Statistical tests of neutrality of mutations. *Genetics* **133**:693–709.
- Guttman, D. S., and D. E. Dykhuizen. 1994. Clonal divergence in *Escherichia coli* as a result of recombination. *Science* **266**:1380–1383.
- Hommais, F., S. Gouriou, C. Amorin, H. Bui, M. C. Rahimy, B. Picard, and E. Denamur. 2003. The FimH A27V mutation is pathoadaptive for urovirulence in *Escherichia coli* B2 phylogenetic group isolates. *Infect. Immun.* **71**:3619–3622.
- Hughes, A. L., and M. Nei. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* **89**:167–170.
- Hung, C. S., J. Bouckaert, D. Hung et al. (11 co-authors). 2002. Structural basis of tropism of *Escherichia coli* to the bladder during urinary tract infection. *Mol. Microbiol.* **44**:903–915.
- Johnson, J. R., J. J. Brown, U. B. Carlino, and T. A. Russo. 1998. Colonization with and acquisition of uropathogenic *Escherichia coli* as revealed by polymerase chain reaction-based detection. *J. Infect. Dis.* **177**:1120–1124.
- Johnson, J. R., A. R. Manges, T. T. O'Bryan, and L. R. Riley. 2002. A disseminated multidrug-resistant clonal group of uropathogenic *Escherichia coli* in pyelonephritis. *Lancet* **359**:2249–2251.
- Johnson, J. R., and T. A. Russo. 2002. Extraintestinal pathogenic *Escherichia coli*: “the other bad *E. coli*.” *J. Lab. Clin. Med.* **139**:155–62.
- Johnson, J. R., and A. L. Stell. 2000. Extended virulence genotypes of *Escherichia coli* strains from patients with urosepsis in relation to phylogeny and host compromise. *J. Infect. Dis.* **181**:261–272.
- Klemm, P., and G. Christiansen. 1987. Three *fim* genes required for the regulation of length and mediation of adhesion of *Escherichia coli* type 1 fimbriae. *Mol. Gen. Genet.* **208**:439–445.
- Kreitman, M., and R. R. Hudson. 1991. Inferring the evolutionary histories of the Adh and Adh-dup loci in *Drosophila melanogaster* from patterns of polymorphism and divergence. *Genetics* **127**:565–582.
- Krogfelt, K. A., B. A. McCormick, R. L. Burghoff, D. C. Laux, and P. S. Cohen. 1991. Expression of *Escherichia coli* F-18 type 1 fimbriae in the streptomycin-treated mouse large intestine. *Infect. Immun.* **59**:1567–1568.
- Levin, B. R., and J. J. Bull. 1994. Short-sighted evolution and the virulence of pathogenic microorganisms. *Trends Microbiol.* **2**:76–81.
- Lin, Z., M. Schenker, P. Zhu, and M. Achtman. 2000. Frequent interspecific genetic exchange between commensal neisseria and *Nisseria meningitidis*. *Mol. Microbiol.* **36**:1049–1058.
- Maiden, M. C. J., J. A. Bygraves, E. Feil et al. (13 co-authors). 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. USA* **95**:3140–3145.
- Manges, A. R., J. R. Johnson, B. Foxman, T. T. O'Bryan, K. E. Fullerton, and L. W. Riley. 2001. Widespread distribution of urinary tract infections caused by a multidrug-resistant *Escherichia coli* clonal group. *N. Engl. J. Med.* **345**:1007–1013.
- Nei, M., and T. Gojobori. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol.* **3**:418–426.
- Ochman, H., and N. A. Moran. 2001. Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. *Science* **292**:1096–1099.
- Orr, M. R., and T. B. Smith. 1998. Ecology and speciation. *Trends Ecol. Evol.* **13**:502–506.
- Page, R. D. M., and E. C. Holmes. 1998. Pp. 270–279 in *Molecular evolution. A phylogenetic approach*. Blackwell Science Ltd., Oxford.
- Pulliam, H. R. 1988. Sources, sinks, and population regulation. *Am. Nat.* **132**:652–661.
- Pupo, G. M., D. K. Karaolis, R. Lan, and P. R. Reeves. 1997. Evolutionary relationships among pathogenic and nonpathogenic *Escherichia coli* strains inferred from multilocus enzyme electrophoresis and *mdh* sequence studies. *Infect. Immun.* **65**:2685–2692.
- Schaeffer, A. J. 2002. Clonal and pathotypic analysis of archetypal *Escherichia coli* cystitis isolate NU14. *J. Urol.* **168**:1651–1652.
- Schembri, M. A., E. V. Sokurenko, and P. Klemm. 2000. Functional flexibility of the FimH adhesin: insights from a random mutant library. *Infect. Immun.* **68**:2638–2646.
- Sharp, P. M. 1997. In search of molecular Darwinism. *Nature* **385**:111–112.
- Sokurenko, E. V., V. Chesnokova, R. J. Doyle, and D. L. Hasty. 1997. Diversity of the *Escherichia coli* type 1 fimbrial lectin. Differential binding to mannosides and uroepithelial cells. *J. Biol. Chem.* **272**:17880–17886.
- Sokurenko, E. V., V. Chesnokova, D. E. Dykhuzien, I. Ofek, X.-R. Wu, K. A. Krogfelt, C. Struve, M. A. Schembri, and D. L. Hasty. 1998. Pathogenic adaptation of *Escherichia coli* by natural variation of the FimH adhesin. *Proc. Natl. Acad. Sci. USA* **95**:8922–8926.
- Sokurenko, E. V., H. S. Courtney, J. Maslow, A. Siitonen, and D. L. Hasty. 1995. Quantitative differences in adhesiveness of type 1 fimbriated *Escherichia coli* due to structural differences in *fimH* genes. *J. Bacteriol.* **177**:3680–3686.
- Sokurenko, E. V., D. L. Hasty, and D. E. Dykhuzien. 1999. Pathoadaptive mutations: gene loss and variation in bacterial pathogens. *Trends Microbiol.* **5**:191–195.

- Stapleton, A., S. Moseley, and W. E. Stamm. 1991. Urovirulence determinants in *Escherichia coli* isolates causing first-episode and recurrent cystitis in women. *J. Infect. Dis.* **163**:773–779.
- Suzuki, Y., and T. Gojobori. 1999. A method for detecting positive selection at single amino acid sites. *Mol. Biol. Evol.* **16**:1315–1328.
- Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA Polymorphisms *Genetics* **123**:585–595.
- Talan, D. A., W. E. Stamm, T. M. Hooton, G. J. Moran, T. Burke, A. Irvani, J. Reuning-Scherer, L. Faulkner, and D. Church. 2000. Comparison of ciprofloxacin (7 days) and trimethoprim-sulfamethoxazole (14 days) for acute uncomplicated pyelonephritis pyelonephritis in women: a randomized trial. *JAMA* **283**:1583–1590.
- Thomas, W. E., E. Trintchina, M. Forero, V. Vogel, and E. V. Sokurenko. 2002. Bacterial adhesion to target cells enhanced by shear force. *Cell* **109**:913–23.
- Zhang, J., H. F. Rosenberg, and M. Nei. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc. Natl. Acad. Sci. USA* **95**:3708–3713.

Antony Dean, Associate Editor

Accepted March 5, 2004