



## Genes under positive selection in *Escherichia coli*

Lise Petersen, Jonathan P. Bollback, Matt Dimmic, et al.

*Genome Res.* 2007 17: 1336-1343 originally published online August 3, 2007  
Access the most recent version at doi:[10.1101/gr.6254707](https://doi.org/10.1101/gr.6254707)

---

**References** This article cites 79 articles, 38 of which can be accessed free at:  
<http://genome.cshlp.org/content/17/9/1336.full.html#ref-list-1>

### License

**Email Alerting Service** Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

---

An advertisement banner with a teal background. On the left, the text reads "CRISPR and RNAi Genetic Screening. Your new superpower." In the center, there is a white button with the text "LEARN MORE". On the right, there is a photograph of a woman wearing a red superhero mask and cape, and the Cellecta logo, which consists of a green molecular structure and the word "CELLECTA" in white capital letters.

CRISPR and RNAi Genetic Screening.  
Your new superpower.

LEARN MORE

CELLECTA

---

To subscribe to *Genome Research* go to:  
<https://genome.cshlp.org/subscriptions>

---

Copyright © 2007, Cold Spring Harbor Laboratory Press

# Genes under positive selection in *Escherichia coli*

Lise Petersen,<sup>1,5</sup> Jonathan P. Bollback,<sup>1,2</sup> Matt Dimmic,<sup>3</sup> Melissa Hubisz,<sup>4</sup> and Rasmus Nielsen<sup>2</sup>

<sup>1</sup>Bioinformatics Centre, University of Copenhagen, Copenhagen DK-2200, Denmark; <sup>2</sup>Center for Comparative Genomics, Institute of Biology, University of Copenhagen, Copenhagen DK-2100, Denmark; <sup>3</sup>Divergence, Inc., St. Louis, Missouri 63141, USA;

<sup>4</sup>Department of Human Genetics, University of Chicago, Chicago, Illinois 60637, USA

We used a comparative genomics approach to identify genes that are under positive selection in six strains of *Escherichia coli* and *Shigella flexneri*, including five strains that are human pathogens. We find that positive selection targets a wide range of different functions in the *E. coli* genome, including cell surface proteins such as beta barrel porins, presumably because of the involvement of these genes in evolutionary arms races with other bacteria, phages, and/or the host immune system. Structural mapping of positively selected sites on *trans*-membrane beta barrel porins reveals that the residues under positive selection occur almost exclusively in the extracellular region of the proteins that are enriched with sites known to be targets of phages, colicins, or the host immune system. More surprisingly, we also find a number of other categories of genes that show very strong evidence for positive selection, such as the enigmatic *rhs* elements and transposases. Based on structural evidence, we hypothesize that the selection acting on transposases is related to the genomic conflict between transposable elements and the host genome.

[Supplemental material is available online at [www.genome.org](http://www.genome.org).]

Comparative genomics data is ideal for identifying genes affected by natural selection. By comparing the genomes of different organisms, genomic elements or positions with reduced substitution rates due to negative selection, or increased substitution rates due to positive selection, can be identified. While negative selection is pervasive in functional genetic elements, there has been particular interest in identifying positive selection because it provides evidence for adaptive changes in function. Recently, several genomic scans have been published identifying positively selected genes by comparing genomic sequences from human, chimpanzees, and mouse (Clark et al. 2003; Nielsen et al. 2005), and different species of *Drosophila* (Richards et al. 2005).

The statistics used to identify positive selection in these studies is the nonsynonymous to synonymous substitution rate ratio,  $\omega$  ( $d_N/d_S$ ). An excess of nonsynonymous substitutions ( $d_N$ ) over synonymous substitutions ( $d_S$ ) is an unambiguous indicator of positive natural selection at the molecular level (Hughes and Nei 1988; Yang and Bielawski 2000). However, because most positions in functional genes are conserved, the average value of  $\omega$  tends to be  $>1$ , even in genes that have experienced large amounts of positive selection. To account for this fact, methods that allow variation in  $\omega$  among sites have been developed (Nielsen and Yang 1998). These methods can identify positive selection when only a small fraction of sites in the gene have been affected. In addition, they have been shown to be highly robust in simulation studies and to have considerable statistical power even when relatively few sequences are included in the study (Wong et al. 2004).

Genome-wide studies have identified a number of categories enriched with genes targeted by positive selection, in particular immune and defense-related genes, which are involved in dynamic host–pathogen interactions.

The objective of this study is to further characterize the ef-

fect of positive selection on the *Escherichia coli* genome. Recently, Chen et al. (2006) studied positive selection that is specific to uropathogenic *E. coli* strains. We are particularly interested in examining the degree to which positive selection can be attributed entirely to the effect of host–pathogen interactions, or whether there are other important forces leading to positive selection in *E. coli*. We use  $d_N/d_S$ -based methods to scan the genomes of *E. coli* (four strains) and *Shigella flexneri* (two strains) for evidence of positive selection. The genetic similarity between *E. coli* and *S. flexneri* is sufficient to allow inclusion in the same species, but for historical reasons this has not been the case (Rolland et al. 1998; Wei et al. 2003). The comparison of all six genomes provides an excellent basis for studying evolution in *E. coli*, because the level of divergence between the strains is high enough to detect the signature of selection, yet low enough that severe problems with alignment and saturation by multiple substitutions are avoided.

The statistical methods used in this study will detect positive selection when it has led to multiple nonsynonymous substitutions at the same codon site in one or more strains. The natural habitat of *E. coli* is the intestinal tract of warm-blooded animals, where it forms part of the normal flora. However, certain strains are also (opportunistic) pathogens of warm-blooded animals. We included strains belonging to a variety of ecological niches to optimize detection of genomic variation. The K12 is a laboratory strain, and to our knowledge this strain has not been implicated in human disease, even though most (or all) strains are pathogenic in one form or another, and all will have some interactions with the host immune system. The other strains, the UPEC strain *E. coli* CFT073, the two EHEC strains *E. coli* Sakai and EDL933, and the *S. flexneri* strains 2457T and 301, are pathogenic to man. Genes involved in interactions with the host immune and defense system may therefore be expected to show evidence for positive selection, due to selection to avoid immune recognition or selection promoting the ability to invade and colonize host cells. Other genes involved in direct dynamic interactions with the environment, such as attacks from other microorgan-

<sup>5</sup>Corresponding author.

E-mail [lpe@binf.ku.dk](mailto:lpe@binf.ku.dk); fax 45-3532-1281.

Article published online before print. Article and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.6254707>.

isms or to otherwise adapt to a changing environment, are also obvious candidates for positive selection. This universal selection can be expected to have occurred in pathogenic as well as commensal strains.

## Results and Discussion

A total of 3757 genes from *E. coli* K12 were found to have an ortholog in at least two other strains with sufficient similarity. Among these, 16 genes were eliminated because of suspected alignment errors, frameshift mutations, or because of ambiguous assignment of orthology. The tests for positive selection used here are not robust to the effect of intragenic recombination among strains. A statistical test for recombination revealed 236 genes with some evidence for recombination among the strains, which were subsequently eliminated from the analysis (see Supplemental material for genes that show evidence of recombination events).

For all remaining genes ( $n = 3505$ ), a test for positive selection was applied (see Methods for details). Genes were assigned to functional categories according to the EcoCyc database, and the total evidence for positive selection in a category was determined using Fisher's combined  $P$ -value (Fisher 1948).

Eight gene categories showed evidence of positive selection (Table 1) after correcting for multiple tests (disregarding categories containing less than five genes). A total of 23 genes were identified to be under positive selection (Table 2). As in previous studies of eukaryotes (Clark et al. 2003; Nielsen et al. 2005), we find that the proteins that typically undergo positive selection are exposed on the cell surface. Many of them (belonging to categories BC-4.1.B beta barrel porins, BC-7.4 outer membrane, BC-1.6.11 glycoprotein, and BC-8.4 colicin related), are directly involved in interactions with the host, phages, or other bacteria, and our results may suggest they are involved in co-evolutionary arms races.

However, a number of genes, or categories of genes, show evidence for positive selection without any obvious connection to host–pathogen interactions or interactions with other microorganisms. For example, a number of genes involved in horizontal DNA transfer (in BC-8 and subcategories), including the IS30 transposases (b1404 and b4282), show evidence of positive selection. Interestingly, we also identify several proteins in categories BC-1.6.3.2 core region and BC-1.5.3.11 menaquinone, ubiquinone, of which four (RfaC, WecD, PqqL, and UbiF) have enzymatic functions, and one (PurR) is a transcriptional repressor (see below). This finding may indicate that a substantive amount of positive selection may be related to other adaptive processes than immune/defense evasion and interactions with phages and other

microorganisms, possibly metabolic adaptations in response to a changing environment.

We may have missed evidence for positive selection among the 236 genes excluded due to evidence for recombination. Genes with high recombination rates may possibly also be the genes that are more likely to experience positive selection. It is quite possible that recombination is an important contributor to genomic evolution in bacteria (e.g., Guttman and Dykhuizen 1994), and an investigation of the impact of recombination based on the complete *E. coli* genomes is clearly very interesting, but beyond the scope of the present study.

Our results differ from the results by Chen et al. (2006) in several important ways, primarily because Chen et al. (2006) focused on selection specific to the uropathogenic strain, and disregarded selection that was (simultaneously) acting in the other strains. In contrast, the methods used here identify selection that acts in one or more of the six strains used.

Only four genes were identified in both studies (*ompF*, *ompC*, *fhuA*, and *yegO/mdtC*). *MdtC* is involved in antibiotic resistance (Nagakubo et al. 2002), whereas the others code for beta barrel porins that are known to have multiple functions in uptake of colicins and other antibiotic compounds and phages, as well as interactions with the host (see below). It is likely that selection, even though it may be specific to the uropathogenic strains, is not exclusively attributable to host–pathogen interactions, as those strains propagate in the mammalian intestinal tract as well as other reservoirs in close contact with phages and other microorganisms. Furthermore, proteins that are under positive selection due to specific host–pathogen interactions may simultaneously be involved in interactions with other microorganisms/phages, and therefore be under positive selection in all strains.

Our approach resulted in the identification of 19 additional genes, including several that are believed to be important in human infection. In the following, we will discuss the results obtained for the major EcoCyc gene categories with significant evidence for positive selection as determined by Fisher's combined  $P$ -value.

### Beta barrel porins

Beta barrel porins are pore proteins in the outer membrane of Gram-negative bacteria, mitochondria, and chloroplasts. Their structure has been thoroughly described elsewhere (Schirmer 1998; Endriss and Braun 2004). They allow the passage of small, hydrophilic, or charged molecules (Nikaido 1994, 1996; Schirmer 1998); however, porins are also important in interactions with the host immune system (Fourel et al. 1993; Moeck et al. 1995; Andersen et al. 1999; Massari et al. 2003; Secundino et al. 2006) and in pathogenesis (Nikaido 1996; Snyder et al. 2004). Further-

**Table 1.** Gene categories that show evidence of positive selection ( $P \geq 0.01$ )

EcoCyc category	Description	Category size (no. of genes under positive selection)	Fisher $P$ -value
BC-4.1.B	Beta barrel porins	18 (5)	$1.47 \times 10^{-13}$
BC-8.3	Transposon related	46	$1.18 \times 10^{-10}$
BC-7.4	Outer membrane	59	$1.50 \times 10^{-10}$
BC-8	Extrachromosomal	128	$4.33 \times 10^{-10}$
BC-1.6.3.2	Core region	13	$7.73 \times 10^{-8}$
BC-1.5.3.11	Menaquinone, ubiquinone	25	$3.01 \times 10^{-6}$
BC-1.6.11	Glycoprotein	6	$3.86 \times 10^{-5}$
BC-8.4	Colicin related	10	$3.66 \times 10^{-3}$

**Table 2.** Genes that show evidence of positive selection

Gene name	Gene	Likelihood ratio	$d_N/d_S$	Function
<i>fhuA</i>	b0150	13.78	2.66	Outer membrane protein receptor for ferrichrome, colicin M, and phages T1, T5, and phi80
<i>eaeH</i>	b0297	17.11	999.00	Attaching and effacing protein, pathogenesis factor
<i>nmpC</i>	b0553	9.80	105.79	Outer membrane porin protein; locus of <i>qsr</i> prophage
<i>ubiF</i>	b0662	7.24	7.96	2-Octoprenyl-3-methyl-6-methoxy-1,4-benzoquinone hydroxylase
<i>ompF</i>	b0929	13.90	36.78	Outer membrane protein 1a (Ia;b;F)
<i>ompA</i>	b0957	61.93	209.08	Outer membrane protein 3a (II*;G;d)
<i>ycgV</i>	b1202	12.94	999.00	Putative adhesion and penetration protein
<i>yciD</i>	b1256	10.63	8.41	Putative outer membrane protein
<i>rzpR</i>	b1362	11.03	998.97	Putative Rac prophage endopeptidase
<i>tra8_2</i>	b1404	84.33	653.04	IS30 transposase
<i>yddK</i>	b1471	38.05	13.02	Putative glycoprotein
<i>pqqL</i>	b1494	95.25	21.52	Putative zinc protease
<i>nohA</i>	b1548	9.06	513.98	DNA packaging protein NU1 homolog from lambdoid prophage Qin
<i>purR</i>	b1658	20.55	5.68	Transcriptional repressor for <i>pur</i> regulon, <i>glyA</i> , <i>glnB</i> , <i>prsA</i> , <i>speA</i>
<i>yeeU</i>	b2004	11.59	11.14	Toxin of the YeeV–YeeU toxin–antitoxin pair
<i>yeeV</i>	b2005	34.01	31.09	Antitoxin of the YeeV–YeeU toxin–antitoxin pair
<i>mdtC</i>	b2076	6.55	13.15	Component of MdtABC-TolC Multidrug Efflux Transport System
<i>ompC</i>	b2215	39.28	62.96	Outer membrane protein 1b (Ib;c)
<i>insA_6</i>	b3444	11.93	1.57	IS1 protein InsA
<i>rfaC</i>	b3621	113.36	56.37	Heptosyl transferase I; lipopolysaccharide core biosynthesis
<i>wecD</i>	b3790	6.33	12.85	dTDP-fucosamine acetyltransferase
<i>lamB</i>	b4036	54.49	545.22	Phage lambda receptor protein; maltose high-affinity receptor
<i>tra8_3</i>	b4284	84.33	653.05	IS30 transposase

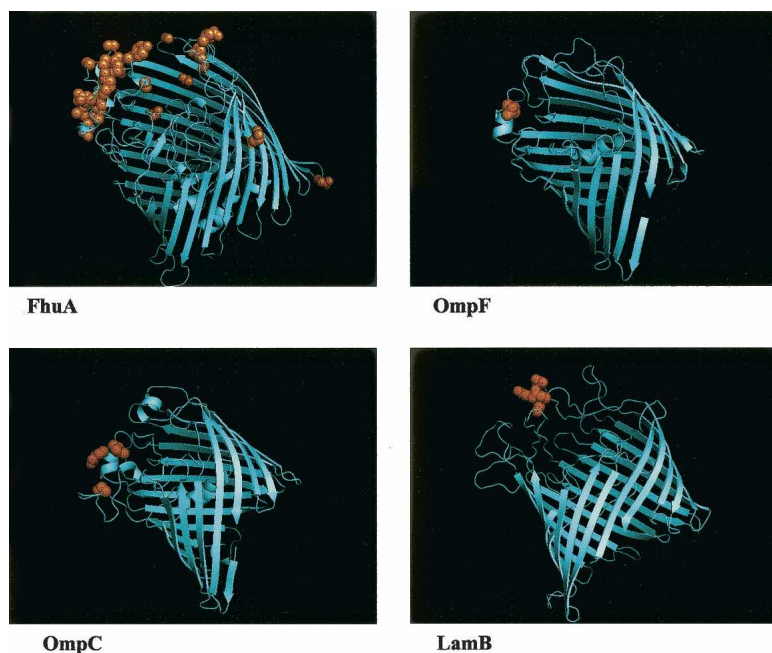
more, they serve as receptors for phages (Charbit 2003; Endriss and Braun 2004; Housden et al. 2005), antibiotics (Endriss and Braun 2004), and colicins (James et al. 1996; Lazdunski et al. 1998; Riley and Gordon 1999).

Five of the 18 beta barrel porin loci (*lamB*, *ompC*, *ompF*, *ompW*, and *fhuA*) show significant evidence of positive selection. The protein structures for four of these porins have been experimentally determined (Fig. 1).

LamB's primary role is as a high-affinity receptor for malt-

ose. However, in addition, it functions as the receptor for phage lambda. Several sites/regions located on different loops are known to be important for phage binding, including residues 382, 386, 387, and 389 on loop 9 (Clement et al. 1983; Charbit et al. 1988, 1994; Andersen et al. 1999). Andersen et al. (1999) found that loop 9, but apparently not the other loops, is important for recognition by monoclonal antibodies. All four sites predicted to be under positive selection in our analysis are located in this very region on loop 9 (residues 383, 384, 385, and 386) (Fig. 1). This strongly suggests that the positive selection acting on the *lamB* gene is caused by selection to avoid phage binding and/or binding of defense-related molecules from the host immune/defense system.

OmpC and OmpF are some of the most abundant proteins in *E. coli*, representing up to 2% of the total protein of the cell (Nikaido 1996). Expression studies show that the *ompC* and *ompF* genes are significantly up-regulated during urinary tract infections (Nikaido 1996; Snyder et al. 2004). OmpC has been reported to play a key role in virulence of *Salmonella* (Negm and Pistole 1999), and it also functions as the receptor for several phages. A major varied segment between residues 176 and 226 has been reported to affect phage infectivity (Tomassen et al. 1985; Mizuno et al. 1987; Hikita et al. 1989; Yu et al. 2000). We found evidence of positive selection in residues 163, 202, and 203, located on loops 4 and 5, suggesting that the positive selection acting on *ompC* is selection to avoid phage binding, but selection to



**Figure 1.** Three-dimensional structures of beta barrel porins FhuA, OmpF, OmpC, and LamB. Sites that show evidence of positive selection ( $P < 0.01$ ) are depicted as orange spheres.

enhance OmpC's function as a virulence factor may also be important. OmpF has putative antigenic epitopes located on several loops, including loop 5 (Klebba et al. 1990; Fourel et al. 1993; Williams et al. 2000), and it also functions as a receptor and as a translocation channel for a group of colicins (Nikaido 1996), and several residues on this loop are also involved in colicin import (Fourel et al. 1993; Jeanteur et al. 1994; de Zamaroczy and Buckingham 2002; Zakharov and Cramer 2004; Housden et al. 2005). We found evidence for positive selection at residue 203 on loop 5. In summary, our results suggest that selection to avoid the host immune system, and possibly also to avoid colicin uptake, is acting on *ompF*.

FhuA, in addition to its role as ferrichrome receptor, functions as the receptor for several phages (T1, T5,  $\Phi$ 80) as well as colicins, microcins, and antibiotics (Nikaido 1996; Endriss and Braun 2004). It has been reported that loops 4, 5, and 8 are involved in phage specificity, and loops 3, 4, 7, 8, and 11 are involved in the sensitivity to colicins and antibiotics (Bos et al. 1998; Endriss and Braun 2004). Furthermore, loops 4 and 5 are targets for monoclonal antibodies raised against FhuA (Moeck et al. 1995). We found that multiple residues on loops 4, 5, 6, 7, and 8 are affected by positive selection (Fig. 1; see Supplemental Table S.7). Our results strongly suggest that there is a massive selection pressure acting on *fhuA* to avoid recognition by the host immune system, phage binding, and/or bacteriocin binding.

In the four porins with known protein structure, the vast majority of sites predicted to be under positive selection are surface exposed (Fig. 1). Furthermore, the locations coincide with sites experimentally determined to be important for interactions with phages and/or interactions with the host immune/defense system, and interactions with bacteriocins. Positive selection in these genes seems to be associated with selection to avoid recognition by a host immune system and/or binding of phages and colicins.

### Outer membrane/glycoprotein

These two categories contain the beta barrel porins discussed above, together with a number of other loci, of which five (*ompA*, *mdtC*, *caeH*, *yddK*, and *ycgV*) show evidence of positive selection.

OmpA is one of the most abundant proteins in the outer membrane of *E. coli*, and is believed to function in shape stabilization of the cell and as a diffusion channel. In contrast to the beta barrel porins, the conformation of OmpA changes between open and closed channel forms, resulting in a significantly reduced (~50-fold) flux of solutes through OmpA (Sugawara and Nikaido 1994; Koebnik 1999). OmpA is a prime target of the host immune system during infection, and is involved in invasion of epithelial cells in newborn meningitis (Prasadarao et al. 1996, 1999). It also functions as a phage and colicin receptor (Smajs et al. 1997; Power et al. 2006), and several residues on loops 1–3 are involved in phage susceptibility (Smajs et al. 1997; Koebnik 1999; Power et al. 2006). The *ompA* gene shows strong evidence of positive selection in residues 25 (loop1), 67 and 68 (loop2), and 111 (loop3), and each one of them is exposed on the long loops (Fig. 2), consistent with a role in adhesion/invasion, as well as receptor function.

*MdtC* is part of the multidrug resistance cluster *mdtABCD*, where *mdtB/C* are encoding a heteromultimeric transmembrane complex located in the inner membrane. *MdtABC* is involved in resistance against bile salts and antibiotics (Nagakubo et al.



**Figure 2.** Three-dimensional structure of outer membrane protein A (OmpA). Sites that show evidence of positive selection ( $P < 0.01$ ) are depicted as orange spheres.

2002). Two sites show evidence of positive selection, and they are both surface exposed (data not shown).

Three additional proteins that are believed to be involved in host–pathogen interaction have been identified: (1) EaeH (putative adhesin/invasion, Uniprot Q8X6G3), (2) YddK (putative glycoprotein, likely to be involved in protein–protein interactions based on the presence of multiple leucine-rich repeats) (Kobe and Kajava 2001), and (3) YcgV (adhesin homolog), presumed to be involved in biofilm formation (Roux et al. 2005). Intracellular biofilm formation has been reported to be important in virulence of uropathogenic *E. coli* (Justice et al. 2004).

In summary, we identified five genes that are all highly relevant in human disease. Our results indicate that positive selection, unsurprisingly, plays an important role in pathogenesis. Avoidance of host immune defenses is clearly an important force driving positive selection in *E. coli*.

### Colicin related

Colicins are antibacterial compounds produced by *E. coli* strains and are active against closely related organisms (Pattus et al. 1990; Lazdunski et al. 1998; Housden et al. 2005). Colicins are plasmid encoded (Braun et al. 2002; Riley and Wertz 2002a), and are therefore not included in this study. Diverse mechanisms for colicin translocation across the membranes have been described elsewhere (Lazdunski et al. 1998; Journet et al. 2001; Braun et al. 2002; Cao and Klebba 2002; Zakharov and Cramer 2004). There are obvious reasons why colicin-related genes may be under positive selection, as they are involved in bacterial–bacterial competitive interactions (James et al. 1996; Riley 1998; Riley and Wertz 2002b).

In this category, we identified *fhuA* (discussed under “Beta barrel porins”) and *purR* (discussed under “Menaquinone, ubiquinone”). We will not pursue analysis of this category any further here, because of the redundancy in the annotation, and because the EcoCyc annotation for this category apparently is incomplete. For instance, OmpF is known to function as a colicin receptor (described above under “Beta barrel porins”) but is not included here.

### Core region

The core region forms part of the bacterial lipopolysaccharide (LPS). The inner core region is highly conserved in *E. coli*, whereas the outer core is more variable, and five distinct outer core types are known (Raetz and Whitfield 2002). The core region interacts with the host immune system, and has been used as a vaccine target (Dalwai and Chart 2003; Cross et al. 2004; Gibbs et al. 2004). Furthermore, colicin U can bind to the outer core

(Smajs et al. 1997). Interestingly, two genes that code for enzymes involved in core region formation/modification show evidence of positive selection, *rfaC* (heptosyl transferase) (Kadmas and Raetz 1998) and *wecD* (dTDP-fucosamine acetyltransferase, UniProt P27832). This important finding indicates that positive selection targets all levels in cellular processes and not only structural proteins that are directly involved in interactions.

### Menaquinone and ubiquinone

We identify three genes, *pqqL* (putative zinc peptidase), *ubiF* (2-octaprenyl-3-methyl-6-methoxy-1,4-benzoquinone hydroxylase), and *purR* (transcriptional repressor) (Rolfes and Zalkin 1988; Kawamukai 2002). PurR shows strong evidence of positive selection with a large region being affected (Fig. 3). The precise function of the region identified here is, to our knowledge, not well understood, even though the domain structure of PurR has been described (Sinha et al. 2003); however, our results emphasize the importance of this region. Similar to the Core region category discussed above, we found evidence of positive selection in two enzymes. In this case, however, the end product is not a molecule directly involved in interaction with the host or other microorganisms. Selection seems in this case to be related to gene regulation, and is likely to play a role in adaptation to spatial environmental heterogeneity.

### Extrachromosomal

Within the Extrachromosomal category, 11 of 128 genes showed significant evidence for positive selection. Two genes (*yeeV* and *yeeU*) constitute a toxin–antitoxin pair (Brown and Shaw 2003), while one (*purR*) encodes a known transcriptional repressor (discussed above). However, the remaining genes fall into two general categories: phage related and transposon related. Briefly, we discuss each of these in turn (see Supplemental Table S1).

### Transposon related

Three genes detected as being under positive selection are related to transposable elements, or insertion sequences, two copies of the IS30 transposase gene (*tra8\_2* and *tra8\_3*), and the IS1 gene *insA*. Transposable elements are selfish genomic elements that can mobilize to new positions in the genome, often resulting in a growth in copy number.

The IS30 element is around 1221 bp in length with two imperfect inverted terminal repeats (IR) and codes for a 383 amino acid residue transposase. During transposition, IS30 appears to target two distinct motifs in bacterial genomes: (1) regions known as GOHS (genomic oligonucleotide hot-spot) and

(2) sequence tracts that match the left and right inverted repeats. Nagy et al. (2004) showed that the deletion of a portion of the N-terminal region—the helix-turn-helix motif 1 (HTH1)—of the transposase gene results in a nearly complete elimination of intermolecular transposition to GOHS targets, but has no effect on transposition to IR sites. We found evidence of positive selection in 16 sites in the IS30 transposase gene, of which 14 occur in the HTH1 region, suggesting that selection on the sequences may be acting on the element’s ability to target GOHS sites during transposition.

Insertion of transposable elements may have a negative fitness effect, driving selection in the host genome to control or eliminate insertion. However, selection will act on transposable elements to favor insertion (only transposable elements that keep inserting, albeit at a low rate, will survive in the long run). It is interesting to hypothesize that selection on these genes is the result of an ongoing arms race between the host and the genomic parasite, creating an evolutionary dynamic very similar to host/pathogen interactions (for review, see Hartl et al. 1984; Ajioka and Hartl 1989). Our results show that positive selection drives the evolution in the HTH1 region and suggests the presence of a dynamic co-evolutionary interaction where selection acts on the transposase HTH1 regions to facilitate insertion in the bacterial genome.

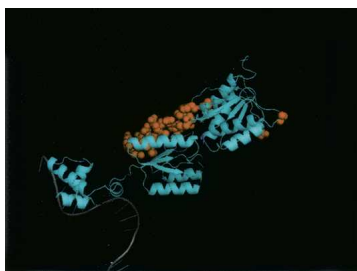
### Phage related

The other genes in this category that showed significant levels of positive selection encode proteins that function as phage receptors or are important in the life cycle of phages: (1) *lamB* and *fluA* (beta barrel porins, see above), (2) *nmpC* (outer membrane porin), located within the locus of a  $\lambda$ qsr prophage that confers sensitivity to some colicins and phages (Pugsley and Schnaitman 1978; Hindahl et al. 1984; Highton et al. 1985), and (3) *nohA* (DNA packaging protein from the prophage  $\lambda$ qin).

The beta barrel porins both serve multiple purposes, and selection to avoid phage uptake may or may not be a significant cause for the positive selection that we observe. The other two genes discussed here appear to be primarily phage related, and in summary, the results suggest an important role for positive selection in phage interactions.

### Other proteins

A number of *rhs* genes (Hill et al. 1994) show remarkably strong evidence for positive selection. The Rhs proteins have no known function, but properties predicted from their primary sequence suggest that they may be cell surface, ligand-binding proteins (Hill et al. 1994). They are not expressed under routine conditions in the laboratory; however, Hill et al. (1995) suggest that they respond to environmental signals and play a role in the success of the cell in natural settings. Hill et al. (1995) hypothesized that Rhs function relates in a key way to the periodic selection event(s) that influence *E. coli* population structure. However, as their function is still largely unknown, the exact causes for positive selection are hard to determine in this group of genes. The fact that *rhs* genes is the group of genes showing the very strongest evidence for positive selection in the *E. coli* genomes, suggests that these genes should be awarded more attention. They may possibly be involved in adaptations associated with changes in the environment, e.g., host shifts.



**Figure 3.** Three-dimensional structure of PurR. Sites that show evidence of positive selection ( $P < 0.01$ ) are depicted as orange spheres. DNA strand is shown as gray ball-and-stick.

## Conclusion

As expected, genes encoding cell surface proteins are the primary targets of positive selection because of their role in interactions with the host immune/defense system, phages, and bacteriocins. In genes with known structure and extensive functional annotation we find an extremely high correspondence between residues under positive selection and residues known to be involved in these types of interactions. These results validate the use of the statistical methodology and suggest that our results, and results of similar future studies, may identify which genes and which protein residues are associated with interactions among genes with less extensive functional and structural annotation.

Surprisingly, we also found many transposon and phage-related genes under positive selection. We hypothesize that these genes are involved in genomic conflicts that are driving positive selection. Another group of genes that show surprisingly strong evidence for positive selection are the *rhs* genes. These genes are likely to play a very important, yet unknown, role in the bacterium's interaction with the environment, possibly in interactions with the host immune/defense system.

## Methods

### Construction of files with aligned orthologous DNA sequences

All protein-coding gene sequences in *E. coli* K12 were blasted (*E*-value cut-off  $10^{-4}$ ) against the five *E. coli* and *S. flexneri* genomes (see details in Table 3).

If the higher scoring match for a given K12 gene in each of the five genomes corresponded to an annotated gene (an ortholog), the protein sequence was retrieved. Orthologous protein sequences were aligned with ClustalW, and subsequently, protein sequences were replaced by the corresponding DNA sequences, preserving the gaps obtained during protein alignment. The resulting DNA multiple alignments were scored using ClustalW, and sequences exhibiting <80% similarity with K12 were removed from the alignments. Also, sequences that had 1-, 2-, or 4-bp indels causing frameshifts were removed. In alignments with start- or stop-codon differences, the sequence overhangs were removed before analysis. Alignments containing the K12 gene and two or more orthologs satisfying the above criteria were analyzed as described below.

In genes where positive selection was identified we did reciprocal blast searches, and in a few cases where the original query sequence was not identified as 'best hit' (probably due to the presence of low-complexity regions that are disregarded in blast scoring), we compared the query and the original blast hits (in the five other genomes) using multiple alignments and literature/database searches to confirm orthology.

It should be noted that while we have taken efforts to avoid comparing paralogous genes, we cannot exclude that in a few cases we have made such comparisons. This will not affect our conclusions, as the bioinformatical methods used to detect posi-

tive selection are equally applicable to paralogous and orthologous genes. However, it is likely that our results regarding transposases may reflect the comparison of different paralogous copies.

### Grouping of genes

Genes were grouped according to the classification used in the EcoCyc database (Keseler et al. 2005).

To counteract erroneous conclusions due to incomplete categories, we disregard categories containing less than five genes.

### Tests for positive selection

To test for positive selection, we used the method of Nielsen and Yang (1998), as implemented in PAML (Yang 1997). This method tests for the presence of positively selected sites by allowing the  $d_N/d_S$  ratio ( $\omega$ ) to vary among sites. A model (M1a), which only allows sites in which  $\omega = 1$  or  $\omega = 0$  is compared with a model (M2a) that additionally allows a proportion ( $p$ ) of sites evolving with value  $\omega_s$ . If a likelihood ratio test rejects M1 in favor of M2, and if the estimate of  $\omega_s$  is larger than 1, this is interpreted as evidence for positive selection. The *P*-value for genes for which  $\omega_s < 1$  is set equal to 1. The genomes under study are closely related and, furthermore, genes showing evidence of recombination have been excluded; therefore, the same tree was used in the  $d_N/d_S$  test for each gene. This test was shown (Wong et al. 2004) to be conservative under a wide range of simulation conditions. However, it was also shown to be less powerful than an alternative test, based on comparing models denoted by M7 and M8 (Yang and Nielsen 2000). The original test was used here because the use of models M7 and M8 may in some cases involve very difficult optimization problems. For single data sets, this problem can usually be easily dealt with, but it is harder to automate methods for checking stability and accuracy of the optimization in genomic scans of thousands of genes. However, all analyses were repeated using the M7 versus M8 test and highly similar results were obtained (see Supplemental material). The optimization algorithm was run three times for each gene with different initial conditions to ensure that the global maximum had been found. To identify functional groups with excess positive selection, we calculate a combined *P*-value for each functional group using Fisher's combined score (Fisher 1948).

### Test for recombination

Recombination can inflate the rate of false positives in studies aimed at detecting positive selection based on  $d_N/d_S$  ratios, because the common methods used for detecting positive selection assume that all sites share the same underlying phylogeny (Anisimova et al. 2003; Shrinier et al. 2003). Therefore, we eliminated any genes in the analysis that showed evidence for recombination among strains, using a statistical test implemented in HyPhy (Pond et al. 2005). In brief, the method compares a likelihood model assuming a single recombination breakpoint with two different topologies on each side of the breakpoint, with a model that assumes no recombination. If the model with recombination is supported by the Akaike Information Criterion (Burnham and Anderson 2002), a likelihood ratio test based on the parametric bootstrapping (e.g., Goldman 1993) is used to determine whether the model with no recombination can be rejected in favor of a model with recombination.

### Mapping of PS sites to three-dimensional structures of proteins

In each alignment the sequence with the fewest gaps around the identified PS-sites were selected. The sequence was blasted

**Table 3. Genome sequences**

Strain	RefSeq accession no.
<i>Escherichia coli</i> K12	NC_000913
<i>Escherichia coli</i> CFT073	NC_004431
<i>Escherichia coli</i> O157:H7 Sakai	NC_002695
<i>Escherichia coli</i> O157:H7 EDL933	NC_002655
<i>Shigella flexneri</i> 2457T	NC_004741
<i>Shigella flexneri</i> 301	NC_004337

against the PDB structure database, and the hit with the longest matching region was retrieved from the database. The PS-sites was mapped onto the structure using the BLAST alignment, and a .pml PyMOL script file was generated for visualization.

## Acknowledgments

This study was partly funded by grants from the Danish FSS, FNU, and Danmarks Grundforskningsfond to R.N., and by grants from the Danish FSS and FNU to J.B.P. We thank Sergei Kosakovsky-Pond for help with the application of the test for recombination.

## References

- Ajioka, J. and Hartl, D.L. 1989. Population dynamics of transposable elements. In *Mobile DNA* (eds. D. Berg and M.M. How), pp. 939–958. American Society for Microbiology, Washington, DC.
- Andersen, C., Bachmeyer, C., Tauber, H., Benz, R., Wang, J., Michel, V., Newton, S.M., Hofnung, M., and Charbit, A. 1999. In vivo and in vitro studies of major surface loop deletion mutants of the *Escherichia coli* K-12 maltoporin: Contribution to maltose and maltooligosaccharide transport and binding. *Mol. Microbiol.* **32**: 851–867.
- Anisimova, M., Nielsen, R., and Yang, Z. 2003. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. *Genetics* **164**: 1229–1236.
- Bos, C., Lorenzen, D., and Braun, V. 1998. Specific in vivo labeling of cell surface-exposed protein loops: Reactive cysteines in the predicted gating loop mark a ferrichrome binding site and a ligand-induced conformational change of the *Escherichia coli* FhuA protein. *J. Bacteriol.* **180**: 605–613.
- Braun, V., Patzer, S.L., and Hantke, K. 2002. Ton-dependent colicins and microcins: Modular design and evolution. *Biochimie* **84**: 365–380.
- Brown, J.M. and Shaw, K.J. 2003. A novel family of *Escherichia coli* toxin-antitoxin gene pairs. *J. Bacteriol.* **185**: 6600–6608.
- Burnham, K.P. and Anderson, D.R. 2002. *Model selection and multimodel inference: A practical information-theoretic approach*. Springer-Verlag, New York.
- Cao, Z. and Klebba, P.E. 2002. Mechanisms of colicin binding and transport through outer membrane porins. *Biochimie* **84**: 399–412.
- Charbit, A. 2003. Maltodextrin transport through lamb. *Front. Biosci.* **8**: s265–s274.
- Charbit, A., Gehring, K., Nikaïdo, H., Ferenci, T., and Hofnung, M. 1988. Maltose transport and starch binding in phage-resistant point mutants of maltoporin. Functional and topological implications. *J. Mol. Biol.* **201**: 487–496.
- Charbit, A., Werts, C., Michel, V., Klebba, P.E., Quillardet, P., and Hofnung, M. 1994. A role for residue 151 of Lamb in bacteriophage lambda adsorption: Possible steric effect of amino acid substitutions. *J. Bacteriol.* **176**: 3204–3209.
- Chen, S.L., Hung, C.S., Xu, J., Reigstad, C.S., Magrini, V., Sabo, A., Blasiar, D., Bieri, T., Meyer, R.R., Ozersky, P., et al. 2006. Identification of genes subject to positive selection in uropathogenic strains of *Escherichia coli*: A comparative genomics approach. *Proc. Natl. Acad. Sci.* **103**: 5977–5982.
- Clark, A.G., Glanowski, S., Nielsen, R., Thomas, P.D., Kejariwal, A., Todd, M.A., Tanenbaum, D.M., Civello, D., Lu, F., Murphy, B., et al. 2003. Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. *Science* **302**: 1960–1963.
- Clement, J.M., Lepouce, E., Marchal, C., and Hofnung, M. 1983. Genetic study of a membrane protein: DNA sequence alterations due to 17 lamb point mutations affecting adsorption of phage lambda. *EMBO J.* **2**: 77–80.
- Cross, A.S., Opal, S., Cook, P., Drabick, J., and Bhattacharjee, A. 2004. Development of an anti-core lipopolysaccharide vaccine for the prevention and treatment of sepsis. *Vaccine* **22**: 812–817.
- Dalwai, F. and Chart, H. 2003. Human antibody responses to R3-core epitopes on the lipopolysaccharide of *Escherichia coli* O157. *Lett. Appl. Microbiol.* **37**: 429–432.
- de Zamaroczy, M. and Buckingham, R.H. 2002. Importation of nuclease colicins into *E. coli* cells: Endoproteolytic cleavage and its prevention by the immunity protein. *Biochimie* **84**: 423–432.
- Endriss, F. and Braun, V. 2004. Loop deletions indicate regions important for FhuA transport and receptor functions in *Escherichia coli*. *J. Bacteriol.* **186**: 4818–4823.
- Fisher, R.A. 1948. Combining independent tests of significance. *Am. Stat.* **2**: 30.
- Fourel, D., Mizushima, S., Bernadac, A., and Pages, J.M. 1993. Specific regions of *Escherichia coli* OmpF protein involved in antigenic and colicin receptor sites and in stable trimerization. *J. Bacteriol.* **175**: 2754–2757.
- Gibbs, R.J., Stewart, J., and Poxton, I.R. 2004. The distribution of, and antibody response to, the core lipopolysaccharide region of *Escherichia coli* isolated from the faeces of healthy humans and cattle. *J. Med. Microbiol.* **53**: 959–964.
- Goldman, N. 1993. Statistical tests of models of DNA substitution. *J. Mol. Evol.* **36**: 182–198.
- Guttman, D.S. and Dykhuizen, D.E. 1994. Clonal divergence in *Escherichia coli* as a result of recombination, not mutation. *Science* **266**: 1380–1383.
- Hartl, D., Dykhuizen, D., and Berg, D.E. 1984. Accessory DNAs in bacteria: Playground for coevolution. In *The origins and development of adaptation* (eds. D. Evered and G.M. Collins), CIBA Foundation Symposium No. 102, pp. 233–245. Pitman, London.
- Highton, P.J., Chang, Y., Marcotte Jr., W.R., and Schnaitman, C.A. 1985. Evidence that the outer membrane protein gene nmpC of *Escherichia coli* K-12 lies within the defective *qsr'* prophage. *J. Bacteriol.* **162**: 256–262.
- Hikita, C., Satake, Y., Yamada, H., Mizuno, T., and Mizushima, S. 1989. Structural and functional characterization of the OmpF and OmpC porins of the *Escherichia coli* outer membrane: Studies involving chimeric proteins. *Res. Microbiol.* **140**: 177–190.
- Hill, C.W., Sandt, C.H., and Vlazny, D.A. 1994. Rhs elements of *Escherichia coli*: A family of genetic composites each encoding a large mosaic protein. *Mol. Microbiol.* **12**: 865–871.
- Hill, C.W., Feulner, G., Brody, M.S., Zhao, S., Sadosky, A.B., and Sandt, C.H. 1995. Correlation of Rhs elements with *Escherichia coli* population structure. *Genetics* **141**: 15–24.
- Hindahl, M.S., Crockford, G.W., and Hancock, R.E. 1984. Outer membrane protein NmpC of *Escherichia coli*: Pore-forming properties in black lipid bilayers. *J. Bacteriol.* **159**: 1053–1055.
- Housden, N.G., Loftus, S.R., Moore, G.R., James, R., and Kleanthous, C. 2005. Cell entry mechanism of enzymatic bacterial colicins: Porin recruitment and the thermodynamics of receptor binding. *Proc. Natl. Acad. Sci.* **102**: 13849–13854.
- Hughes, A.L. and Nei, M. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* **335**: 167–170.
- James, R., Kleanthous, C., and Moore, G.R. 1996. The biology of E colicins: Paradigms and paradoxes. *Microbiol.* **142**: 1569–1580.
- Jeanteur, D., Schirmer, T., Fourel, D., Simonet, V., Rummel, G., Widmer, C., Rosenbusch, J.P., Pattus, F., and Pages, J.M. 1994. Structural and functional alterations of a colicin-resistant mutant of OmpF porin from *Escherichia coli*. *Proc. Natl. Acad. Sci.* **91**: 10675–10679.
- Journet, L., Bouveret, E., Rigal, A., Lloubes, R., Lazdunski, C., and Benedetti, H. 2001. Import of colicins across the outer membrane of *Escherichia coli* involves multiple protein interactions in the periplasm. *Mol. Microbiol.* **42**: 331–344.
- Justice, S.S., Hung, C., Theriot, J.A., Fletcher, D.A., Anderson, G.G., Footer, M.J., and Hultgren, S.J. 2004. Differentiation and developmental pathways of uropathogenic *Escherichia coli* in urinary tract pathogenesis. *Proc. Natl. Acad. Sci.* **101**: 1333–1338.
- Kadrmaz, J.L. and Raetz, C.R. 1998. Enzymatic synthesis of lipopolysaccharide in *Escherichia coli*. Purification and properties of heptosyltransferase I. *J. Biol. Chem.* **273**: 2799–2807.
- Kawamukai, M. 2002. Biosynthesis, bioproduction and novel roles of ubiquinone. *J. Biosci. Bioeng.* **94**: 511–517.
- Keseler, I.M., Collado-Vides, J., Gama-Castro, S., Ingraham, J., Paley, S., Paulsen, I.T., Peralta-Gil, M., and Karp, P.D. 2005. EcoCyc: A comprehensive database resource for *Escherichia coli*. *Nucleic Acids Res.* **33**: D334–D337.
- Klebba, P.E., Benson, S.A., Bala, S., Abdullah, T., Reid, J., Singh, S.P., and Nikaïdo, H. 1990. Determinants of OmpF porin antigenicity and structure. *J. Biol. Chem.* **265**: 6800–6810.
- Kobe, B. and Kajava, A.V. 2001. The leucine-rich repeat as a protein recognition motif. *Curr. Opin. Struct. Biol.* **11**: 725–732.
- Koebnik, R. 1999. Structural and functional roles of the surface-exposed loops of the  $\beta$ -barrel membrane protein OmpA from *Escherichia coli*. *J. Bacteriol.* **181**: 3688–3694.
- Lazdunski, C.J., Bouveret, E., Rigal, A., Journet, L., Lloubes, R., and Benedetti, H. 1998. Colicin import into *Escherichia coli* cells. *J. Bacteriol.* **180**: 4993–5002.
- Massari, P., Ram, S., Macleod, H., and Wetzler, L.M. 2003. The role of porins in neisserial pathogenesis and immunity. *Trends Microbiol.* **11**: 87–93.
- Mizuno, T., Kasai, H., and Mizushima, S. 1987. Construction of a series

- of ompC-ompF chimeric genes by in vivo homologous recombination in *Escherichia coli* and characterization of their translational products. *Mol. Gen. Genet.* **207**: 217–223.
- Moeck, G.S., Ratcliffe, M.J., and Coulton, J.W. 1995. Topological analysis of the *Escherichia coli* ferrichrome-iron receptor by using monoclonal antibodies. *J. Bacteriol.* **177**: 6118–6125.
- Nagakubo, S., Nishino, K., Hirata, T., and Yamaguchi, A. 2002. The putative response regulator BaeR stimulates multidrug resistance of *Escherichia coli* via a novel multidrug exporter system, MdtABC. *J. Bacteriol.* **184**: 4161–4167.
- Nagy, Z., Szabo, M., Chandler, M., and Olasz, F. 2004. Analysis of the N-terminal DNA binding domain of the IS30 transposase. *Mol. Microbiol.* **54**: 478–488.
- Negm, R.S. and Pistole, T.G. 1999. The porin OmpC of *Salmonella typhimurium* mediates adherence to macrophages. *Can. J. Microbiol.* **45**: 658–669.
- Nielsen, R. and Yang, Z. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* **148**: 929–936.
- Nielsen, R., Bustamante, C., Clark, A.G., Glanowski, S., Sackton, T.B., Hubisz, M.J., Fedel-Alon, A., Tanenbaum, D.M., Civello, D., White, T.J., et al. 2005. A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.* **3**: e170. doi: 10.1371/journal.pbio.0030170.
- Nikaido, H. 1994. Porins and specific diffusion channels in bacterial outer membranes. *J. Biol. Chem.* **269**: 3905–3908.
- Nikaido, H. 1996. Outer membrane. In *Escherichia coli and Salmonella cellular and molecular biology* (eds. F.C. Neidhardt et al.), pp. 29–47. ASM Press, Washington, DC.
- Pattus, F., Massotte, D., Wilmsen, H.U., Lakey, J., Tsernoglou, D., Tucker, A., and Parker, M.W. 1990. Colicins: Prokaryotic killer-pores. *Experientia* **46**: 180–192.
- Pond, S.L., Frost, S.D., and Muse, S.V. 2005. HyPhy: Hypothesis testing using phylogenies. *Bioinformatics* **21**: 676–679.
- Power, E., M.L., Ferrari, B.C., Littlefield-Wyer, J., Gordon, D.M., Slade, M.B., and Veal, D.A. 2006. A naturally occurring novel *Escherichia coli* outer membrane protein A allele reduces sensitivity to bacteriophage. *Appl. Environ. Microbiol.* **72**: 7930–7932.
- Prasadarao, N.V., Wass, C.A., Weiser, J.N., Stins, M.F., Huang, S.H., and Kim, K.S. 1996. Outer membrane protein A of *Escherichia coli* contributes to invasion of brain microvascular endothelial cells. *Infect. Immun.* **64**: 146–153.
- Prasadarao, N.V., Wass, C.A., Stins, M.F., Shimada, H., and Kim, K.S. 1999. Outer membrane protein A-promoted actin condensation of brain microvascular endothelial cells is required for *Escherichia coli* invasion. *Infect. Immun.* **67**: 5775–5783.
- Pugsley, A.P. and Schnaitman, C.A. 1978. Identification of three genes controlling production of new outer membrane pore proteins in *Escherichia coli* K-12. *J. Bacteriol.* **135**: 1118–1129.
- Raetz, C.R. and Whitfield, C. 2002. Lipopolysaccharide endotoxins. *Annu. Rev. Biochem.* **71**: 635–700.
- Richards, S., Liu, Y., Bettencourt, B.R., Hradecky, P., Letovsky, S., Nielsen, R., Thornton, K., Hubisz, M.J., Chen, R., Meisel, R.P., et al. 2005. Comparative genome sequencing of *Drosophila pseudoobscura*: Chromosomal, gene, and cis-element evolution. *Genome Res.* **15**: 1–18.
- Riley, M.A. 1998. Molecular mechanisms of bacteriocin evolution. *Annu. Rev. Genet.* **32**: 255–278.
- Riley, M.A. and Gordon, D.M. 1999. The ecological role of bacteriocins in bacterial competition. *Trends Microbiol.* **7**: 129–133.
- Riley, M.A. and Wertz, J.E. 2002a. Bacteriocin diversity: Ecological and evolutionary perspectives. *Biochimie* **84**: 357–364.
- Riley, M.A. and Wertz, J.E. 2002b. Bacteriocins: Evolution, ecology, and application. *Annu. Rev. Microbiol.* **56**: 117–137.
- Rolfes, R.J. and Zalkin, H. 1988. Regulation of *Escherichia coli* purF. Mutations that define the promoter, operator, and purine repressor gene. *J. Biol. Chem.* **263**: 19649–19652.
- Rolland, K., Lambert-Zechovsky, N., Picard, B., and Denamur, E. 1998. *Shigella* and enteroinvasive *Escherichia coli* strains are derived from distinct ancestral strains of *E. coli*. *Microbiol.* **144**: 2667–2672.
- Roux, A., Beloin, C., and Ghigo, J.M. 2005. Combined inactivation and expression strategy to study gene function under physiological conditions: Application to identification of new *Escherichia coli* adhesins. *J. Bacteriol.* **187**: 1001–1013.
- Schirmer, T. 1998. General and specific porins from bacterial outer membranes. *J. Struct. Biol.* **121**: 101–109.
- Secundino, I., Lopez-Macias, C., Cervantes-Barragan, L., Gil-Cruz, C., Rios-Sarabia, N., Pastelin-Palacios, R., Villasis-Keever, M.A., Becker, I., Puente, J.L., Calva, E., et al. 2006. Salmonella porins induce a sustained, lifelong specific bactericidal antibody memory response. *Immunology* **117**: 59–70.
- Shriner, D., Nickle, D.C., Jensen, M.A., and Mullins, J.I. 2003. Potential impact of recombination on site-wise approaches for detecting positive natural selection. *Genet. Res.* **81**: 115–121.
- Sinha, S.C., Krahn, J., Shin, B.S., Tomchick, D.R., Zalkin, H., and Smith, J.L. 2003. The purine repressor of *Bacillus subtilis*: A novel combination of domains adapted for transcription regulation. *J. Bacteriol.* **185**: 4087–4098.
- Smajs, D., Pils, H., and Braun, V. 1997. Colicin U, a novel colicin produced by *Shigella boydii*. *J. Bacteriol.* **179**: 4919–4928.
- Snyder, J.A., Haugen, B.J., Buckles, E.L., Lockatell, C.V., Johnson, D.E., Donnerberg, M.S., Welch, R.A., and Mobley, H.L. 2004. Transcriptome of uropathogenic *Escherichia coli* during urinary tract infection. *Infect. Immun.* **72**: 6373–6381.
- Sugawara, E. and Nikaido, H. 1994. OmpA protein of *Escherichia coli* outer membrane occurs in open and closed channel forms. *J. Biol. Chem.* **269**: 17981–17987.
- Tommassen, J., van der Ley, P., van Zeijl, M., and Agterberg, M. 1985. Localization of functional domains in *E. coli* K-12 outer membrane porins. *EMBO J.* **4**: 1583–1587.
- Wei, J., Goldberg, M.B., Burland, V., Venkatesan, M.M., Deng, W., Fournier, G., Mayhew, G.F., Plunkett 3rd, G., Rose, D.J., Darling, A., et al. 2003. Complete genome sequence and comparative genomics of *Shigella flexneri* serotype 2a strain 2457T. *Infect. Immun.* **71**: 2775–2786.
- Williams, K.M., Bigley 3rd, E.C., and Raybourne, R.B. 2000. Identification of murine B-cell and T-cell epitopes of *Escherichia coli* outer membrane protein F with synthetic polypeptides. *Infect. Immun.* **68**: 2535–2545.
- Wong, W.S., Yang, Z., Goldman, N., and Nielsen, R. 2004. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics* **168**: 1041–1051.
- Yang, Z. 1997. PAML: A program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**: 555–556.
- Yang, Z. and Bielawski, J.P. 2000. Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* **15**: 496–503.
- Yang, Z. and Nielsen, R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* **17**: 32–43.
- Yu, S.L., Ko, K.L., Chen, C.S., Chang, Y.C., and Syu, W.J. 2000. Characterization of the distal tail fiber locus and determination of the receptor for phage AR1, which specifically infects *Escherichia coli* O157:H7. *J. Bacteriol.* **182**: 5962–5968.
- Zakharov, S.D. and Cramer, W.A. 2004. On the mechanism and pathway of colicin import across the *E. Coli* outer membrane. *Front. Biosci.* **9**: 1311–1317.

Received January 3, 2007; accepted in revised form June 14, 2007.