

Chaperone-Usher Fimbriae of *Escherichia coli*

Daniël J. Wurpel, Scott A. Beatson*, Makrina Totsika, Nicola K. Petty, Mark A. Schembri*

Australian Infectious Diseases Research Centre, School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Queensland, Australia

Abstract

Chaperone-usher (CU) fimbriae are adhesive surface organelles common to many Gram-negative bacteria. *Escherichia coli* genomes contain a large variety of characterised and putative CU fimbrial operons, however, the classification and annotation of individual loci remains problematic. Here we describe a classification model based on usher phylogeny and genomic locus position to categorise the CU fimbrial types of *E. coli*. Using the BLASTp algorithm, an iterative usher protein search was performed to identify CU fimbrial operons from 35 *E. coli* (and one *Escherichia fergusonii*) genomes representing different pathogenic and phylogenetic lineages, as well as 132 *Escherichia spp.* plasmids. A total of 458 CU fimbrial operons were identified, which represent 38 distinct fimbrial types based on genomic locus position and usher phylogeny. The majority of fimbrial operon types occupied a specific locus position on the *E. coli* chromosome; exceptions were associated with mobile genetic elements. A group of core-associated *E. coli* CU fimbriae were defined and include the Type 1, Yad, Yeh, Yfc, Mat, F9 and Ybg fimbriae. These genes were present as intact or disrupted operons at the same genetic locus in almost all genomes examined. Evaluation of the distribution and prevalence of CU fimbrial types among different pathogenic and phylogenetic groups provides an overview of group specific fimbrial profiles and insight into the ancestry and evolution of CU fimbriae in *E. coli*.

Citation: Wurpel DJ, Beatson SA, Totsika M, Petty NK, Schembri MA (2013) Chaperone-Usher Fimbriae of *Escherichia coli*. PLoS ONE 8(1): e52835. doi:10.1371/journal.pone.0052835

Editor: Cecile Wandersman, Pasteur Institute, France

Received: August 31, 2012; **Accepted:** November 22, 2012; **Published:** January 30, 2013

Copyright: © 2013 Wurpel et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by grants from the Australian National Health and Medical Research Council (631654 and APP1012076). MAS is the recipient of an Australian Research Council (ARC) Future Fellowship (FT100100662). SAB is the recipient of an ARC Australian Research Fellowship (DP0881247). MT is the recipient of an ARC Discovery Early Career Researcher Award (DE130101169). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: s.beatson@uq.edu.au (SAB); m.schembri@uq.edu.au (MAS)

Introduction

Fimbriae are long proteinaceous organelles that extend from the surface of many bacteria and mediate diverse functions, including adherence and biofilm formation. Fimbrial adhesins, which are often located at the tip of the organelle, typically recognize specific receptor targets in a lock-and-key fashion, thus enabling the bacterium to target a specific surface and display tissue tropism. Many different types of fimbriae have been described in Gram-positive and Gram-negative bacteria [1]. In Gram-negative bacteria, fimbriae are assembled via a range of different protein translocation systems, including the chaperone-usher (CU) pathway, the type IV secretion pathway and the extracellular nucleation precipitation pathway [2].

Among the fimbrial types produced by Gram-negative bacteria, the CU class of fimbriae is the most abundant. The genes encoding for CU fimbriae are found in most members of the *Enterobacteriaceae* (e.g. *Escherichia coli*, *Salmonella spp.*, *Klebsiella spp.*, *Proteus spp.*, *Enterobacter spp.*, *Citrobacter spp.*) as well as bacteria from other genera including *Pseudomonas*, *Haemophilus*, *Bordetella*, *Burkholderia* and *Acinetobacter* [3,4]. The CU pathway is a highly conserved bacterial secretion system for the assembly of fimbriae on the bacterial cell surface. Fimbrial biogenesis by the CU pathway requires a periplasmic chaperone and an outer membrane assembly platform termed the usher. The chaperone facilitates several essential steps in the pathway; it mediates the folding of fimbrial subunit proteins, prevents their polymerization in the periplasm and directs their passage to the usher. The usher in turn

acts as an assembly platform; it forms a binding scaffold for fimbrial subunit protein-chaperone complexes from the periplasm and facilitates the assembly of the fimbrial structural organelle [5,6,7,8,9,10,11].

The prototypical CU fimbriae are type 1 and P fimbriae from uropathogenic *Escherichia coli* (UPEC), which mediate binding to specific receptors in the bladder and upper urinary tract, respectively, via an adhesin located at the tip of the organelle. The biogenesis, regulation and function of type 1 and P fimbriae have been comprehensively studied [12,13,14,15,16,17]. Type 1 fimbriae are 0.2–2.0 μm long tubular structures predominantly comprised of a major structural subunit (FimA) and containing a tip fibrillum composed of several minor components including the FimH adhesin [16,18,19]. Type 1 fimbriae confer binding to α -D-mannosylated proteins such as uroplakins, which are abundant in the bladder [20]. The expression of type 1 fimbriae by UPEC enhances colonization and host response induction in the murine urinary tract infection (UTI) model, and promotes biofilm formation and host cell invasion [21,22,23]. Like type 1 fimbriae, P fimbriae are composed of a major structural protein (PapA), however they contain a larger tip fibrillum, which is comprised of major (PapE) and minor (PapF, PapK, PapG) components. P fimbriae are strongly associated with acute pyelonephritis; they contribute to the establishment of UTI by binding to the α -D-galactopyranosyl-(1–4)- β -D-galactopyranoside receptor epitope in the globoseries of glycolipids and activate innate immune responses in animal models and in human infection [24,25,26,27].

Table 1. *Escherichia* genomes analysed and CU fimbrial operons identified per strain.

<i>E.coli</i>	Phylogroup	Total CU Operons	Intact CU Operons	Reference
UPEC				
CFT073	B2	12	11	Welch <i>et al.</i> 2002 [41]
536	B2	14	12	Hochhut <i>et al.</i> 2006 [70]
F11	B2	13	11	Rasko <i>et al.</i> 2008 [71]
UTI89	B2	12	10	Chen <i>et al.</i> 2006 [72]
EC958	B2	10	9	Totsika <i>et al.</i> 2011 [73]
UMN026	D	12	12	Touchon <i>et al.</i> 2009 [62]
IAI39	D	14	11	Touchon <i>et al.</i> 2009 [62]
ABU				
83972	B2	12	9	Zdziarski <i>et al.</i> 2010 [74]
NMEC				
S88	B2	10	8	Touchon <i>et al.</i> 2009 [62]
APEC				
APEC01	B2	10	9	Johnson <i>et al.</i> 2006 [75]
EAEC				
55989	B1	16	14	Touchon <i>et al.</i> 2009 [62]
042	D	12	11	Chaudhuri <i>et al.</i> 2010 [76]
EPEC				
O127:H6 E2348/69	B2	9	8	Iguchi <i>et al.</i> 2009 [77]
O55:H7 CB9615	E	14	12	Zhou <i>et al.</i> 2010 [78]
ETEC				
O78:H11 H10407	A	13	8	Crossman <i>et al.</i> 2010 [79]
E24377A	B1	15	12	Rasko <i>et al.</i> 2008 [71]
EHEC				
O26:H11 11368	B1	15	14	Ogura <i>et al.</i> 2009 [80]
O103:H2 12009	B1	14	10	Ogura <i>et al.</i> 2009 [80]
O111:H-11128	B1	14	12	Ogura <i>et al.</i> 2009 [80]
O157:H7 EDL933	E	14	12	Perna <i>et al.</i> 2001 [81]
O157:H7 Sakai	E	14	12	Hayashi <i>et al.</i> 2001 [82]
O157:H7 EC4115	E	14	11	Eppinger <i>et al.</i> 2011 [83]
O157:H7 TW14359	E	14	11	Kulasekara <i>et al.</i> 2009 [84]
Commensal				
ATCC 8739	A	11	9	Joint Genome Institute [85]
IAI1	B1	15	14	Touchon <i>et al.</i> 2009 [62]
ED1a	B2	10	5	Touchon <i>et al.</i> 2009 [62]
HS	A	13	9	Rasko <i>et al.</i> 2008 [71]
SE11	B1	17	16	Oshima <i>et al.</i> 2008 [86]
SE15	B2	9	8	Toh <i>et al.</i> 2010 [87]
Environmental				
SMS-3-5	D	12	10	Fricke <i>et al.</i> 2008 [88]
Laboratory				
BL21(DE3)	A	12	9	Jeong <i>et al.</i> 2009 [89]
B REL606	A	12	9	Jeong <i>et al.</i> 2009 [89]
K-12 MG1655	A	12	9	Blattner <i>et al.</i> 1997 [45]
K-12 DH10 β	A	11	8	Durfee <i>et al.</i> 2008 [90]

Table 1. Cont.

<i>E. coli</i>	Phylogroup	Total CU Operons	Intact CU Operons	Reference
K-12 BW2952	A	11	8	Ferenci <i>et al.</i> 2009 [91]
<i>E. fergusonii</i>		Total CU Operons	Intact CU Operons	Reference
Commensal				
EFER ATCC 35469	-	7	7	Touchon <i>et al.</i> 2009 [62]

UPEC: uropathogenic *E. coli*, ABU: asymptomatic bacteriuria *E. coli*, NMEC: neonatal meningitis *E. coli*, APEC: avian pathogenic *E. coli*, EAEC: enteroaggregative *E. coli*, EPEC: enteropathogenic *E. coli*, EHEC: enterohaemorrhagic *E. coli*.
doi:10.1371/journal.pone.0052835.t001

E. coli represents the most comprehensively studied organism with respect to CU fimbriae. In addition to type 1 and P fimbriae, many other CU fimbriae have been characterised and often the adherence properties of these fimbriae are associated with certain *E. coli* pathotypes. For example, P, F1C and S fimbriae are commonly associated with extra-intestinal *E. coli* (ExPEC; including UPEC and meningitis-associated *E. coli* [NMEC]) [26,28,29], aggregative adherence fimbriae (AAF) are associated with enteroaggregative *E. coli* (EAEC) [30], long polar fimbriae

(LPF) with enteropathogenic *E. coli* (EPEC) and enterohaemorrhagic *E. coli* (EHEC) [31], CS1-CFA/I are associated with human enterotoxigenic *E. coli* [32] and K88 (F4) and K99 (F5) fimbriae with porcine, bovine and ovine enterotoxigenic *E. coli* (ETEC) [33,34]. The significant increase in bacterial genome sequencing that has occurred over the last decade has also resulted in the identification of many CU fimbrial gene clusters that remain uncharacterised. This includes CU fimbriae from commensal *E.*

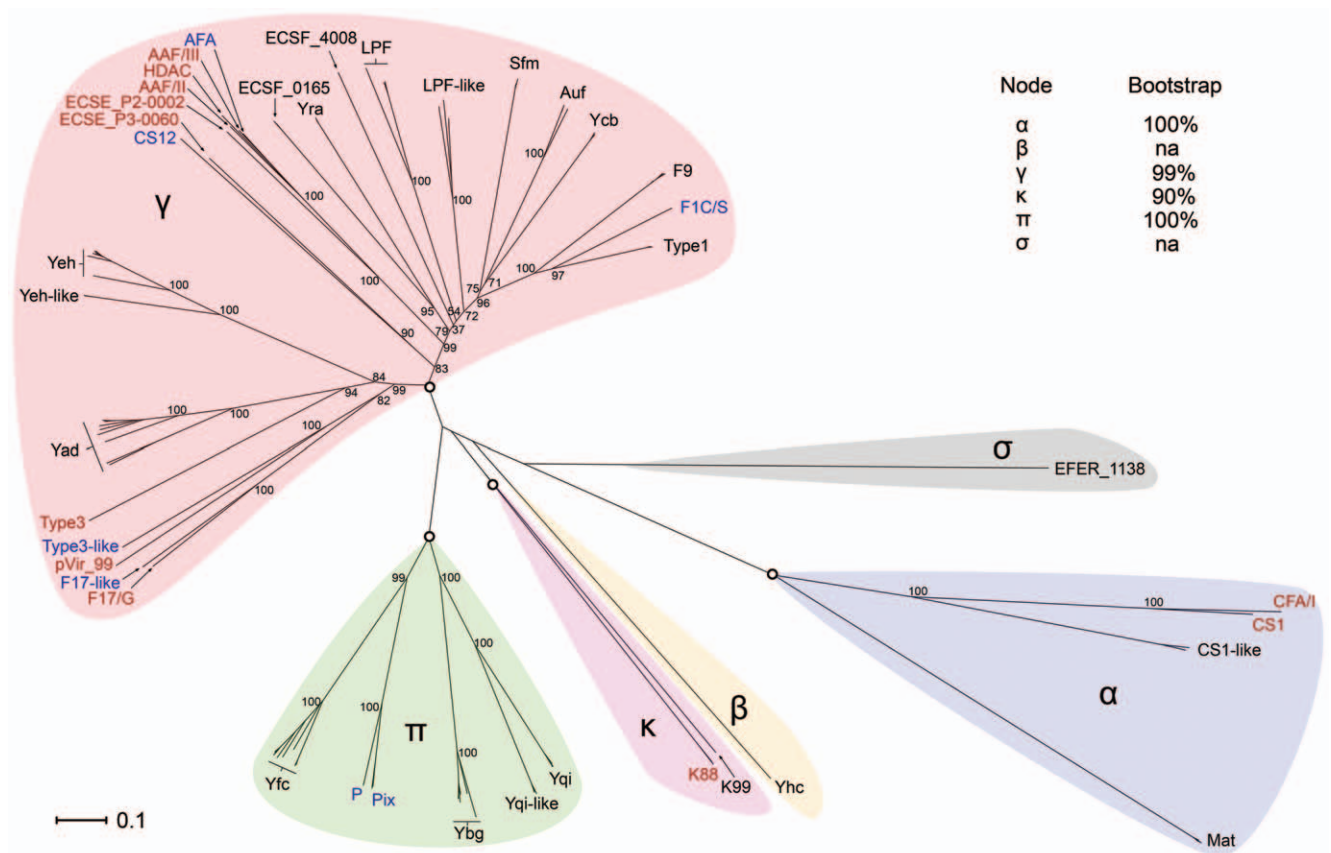


Figure 1. Unrooted phylogram of fimbrial usher proteins identified in *Escherichia coli*. A total of 1075 amino acid positions were used to infer the evolutionary relationship of 383 aligned usher proteins. These consist of 379 usher amino acid sequences belonging to intact fimbrial gene clusters and an additional four usher amino acid sequences of disrupted fimbrial gene clusters (Yhc and AAF/II), which lack intact representatives in the genome sequenced strains examined in this dataset. The corresponding 383 fimbrial gene clusters can be classified as 38 types based on the evolutionary phylogeny of usher amino acid sequence and genetic locus position. Fimbrial gene clusters were grouped according to the Nuccio clade system (α , β , γ , κ , π , σ , open circles represent cladistic nodes) [3], and highlighted in colour. The text of fimbrial types located on PAI's or plasmids is highlighted in blue and red, respectively. Bootstrap values (1000) are displayed as percentage on major nodes. The scale represents the number of amino acid substitutions per site.

doi:10.1371/journal.pone.0052835.g001

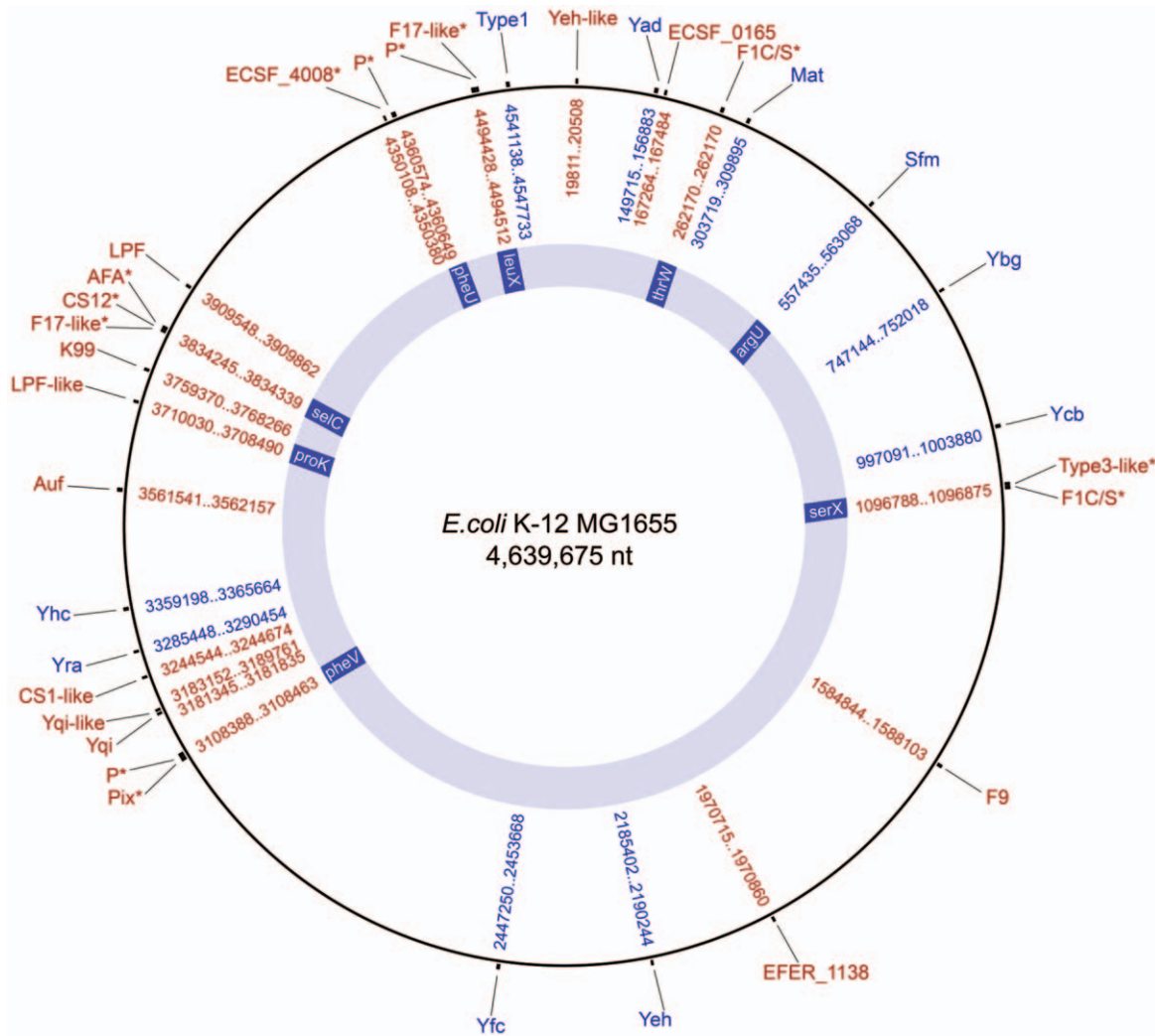


Figure 2. Locus positions of chromosome-borne CU fimbrial operons relative to the genome of *E. coli* MG1655. The *E. coli* K-12 MG1655 chromosome (outer black ring) was used as a reference map to visualise the locus position of 30 chromosome-borne CU fimbrial types. Types highlighted in blue are present in *E. coli* K-12 MG1655, types in red are absent in this strain. Fimbrial types associated with PAIs are indicated by an asterisk. A number of PAI associated fimbrial gene clusters occupy different locus positions relative to the MG1655 genome. tRNA sites that flank CU-containing PAIs are indicated on the inner blue ring. doi:10.1371/journal.pone.0052835.g002

coli strains, where the expression of many CU fimbriae is cryptic and repressed by the histone-like protein H-NS [35].

Early attempts to distinguish between different types of fimbriae from *E. coli* and other Gram-negative bacteria were based either on morphology, function or serology [36,37,38]. More recently, a phylogenetic clade system was established that defines CU fimbriae according to evolutionary descent [3]. In this scheme, CU fimbriae phylogeny is based on the sequence of the usher protein due to its ubiquitous association with all CU gene clusters and the fact that the usher-encoding gene is present in a single copy in all CU gene loci. Here we have employed the classification scheme developed by Nuccio *et al.* [3] to define the repertoire of CU fimbriae in *E. coli*. Thirty five *E. coli* (and one *E. fergusonii*) genomes representing commensal, diarrheagenic and ExPEC strains were searched for genes encoding putative fimbrial usher proteins. A total of 458 usher-encoding genes were identified and individually interrogated for the presence of an adjacent cognate chaperone-encoding gene as well as at least one fimbrial subunit-encoding gene. The CU fimbrial genes were analysed for their

distribution, genetic conservation and genetic location among *E. coli* pathotypes.

Methods

Identification of Chaperone-Usher Operons

The NCBI BLAST2.2.25+ program [39] was utilised to examine two datasets, one consisting of the whole genomes (chromosomes and plasmids) of 36 *Escherichia* strains (Table 1) and the second dataset containing 132 *Escherichia* plasmids (with no associated chromosome sequence available) (Table S1), for the presence of usher sequences. All amino acid sequences encoded by the genomes and plasmids listed in Table 1 and S1 were downloaded from UniProt [40] and used to build a local BLAST database. The 10 usher amino acid sequences annotated in *E. coli* CFT073 [41,42] were used as an initial BLASTp query dataset to probe the local BLAST database. BLASTp searches were performed using the BLOSUM62 series algorithm and an E-value cut-off score of 0.1. Newly identified proteins with a reported

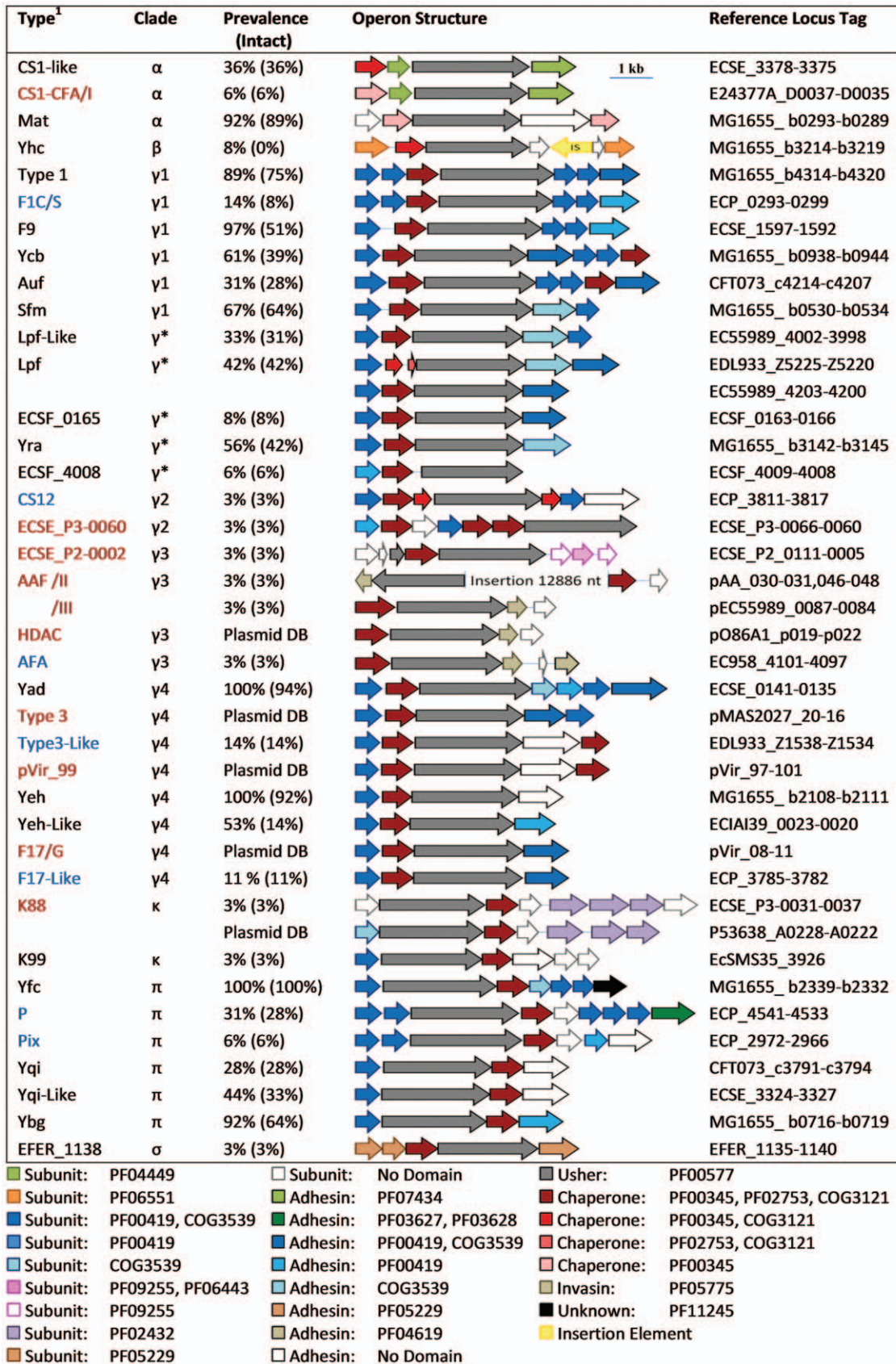


Figure 3. Prevalence and genetic organisation of CU fimbrial types identified in *Escherichia*. The genetic organisation of the different fimbrial types is depicted diagrammatically. Fimbriae are grouped according to the Nuccio clading scheme [3]. Fimbrial prevalence is represented as a percentage of all the strains in the genome dataset. Plasmid-borne fimbriae not part of a genome are highlighted as 'Plasmid DB'. Genes are colour

coded according to predicted function of the corresponding protein product, with associated Pfam and COG domains indicated. The scale represents DNA length in kilo base pair. Reference operon locus tags for individual fimbrial types are displayed on the right. ¹PAI and plasmid-borne operons are highlighted in blue and red, respectively.
doi:10.1371/journal.pone.0052835.g003

E-value of 0 were directly added to the usher database, hits with an E-value >0 were screened for the presence of an usher protein family domain (PF00577) and/or flanking chaperone (PF00345, PF02753 or COG3121) encoding genes before they were added to the usher query dataset. The NCBI Conserved Domain Database (CDD) was used to examine amino acid sequences for conserved domains [43]. After each BLASTp run, the updated usher query dataset was used to re-probe the genome and plasmid sequences until no new usher sequences were found.

Operon Structure Prediction and Analysis of Genetic Context

To determine the genetic organisation of an operon, flanking regions of usher nucleotide sequences were visualised in xBASE

[44]. Fimbrial encoding genes were identified using conserved protein domain searches [43] and sequence homology to annotated genes. Intergenic regions >200 bp were investigated for the presence of protein encoding sequences with conserved fimbrial domains or significant sequence identity to fimbrial subunits. To determine the locus position of chromosome-borne fimbrial operons, the genetic context of each operon was visualised in xBASE and aligned with the genome of *E. coli* K-12 MG1655 [45]. Plasmid-borne fimbrial operons were compared to the closest homologous annotated fimbrial sequences, and analysed for genetic organisation and subunit sequence similarity.

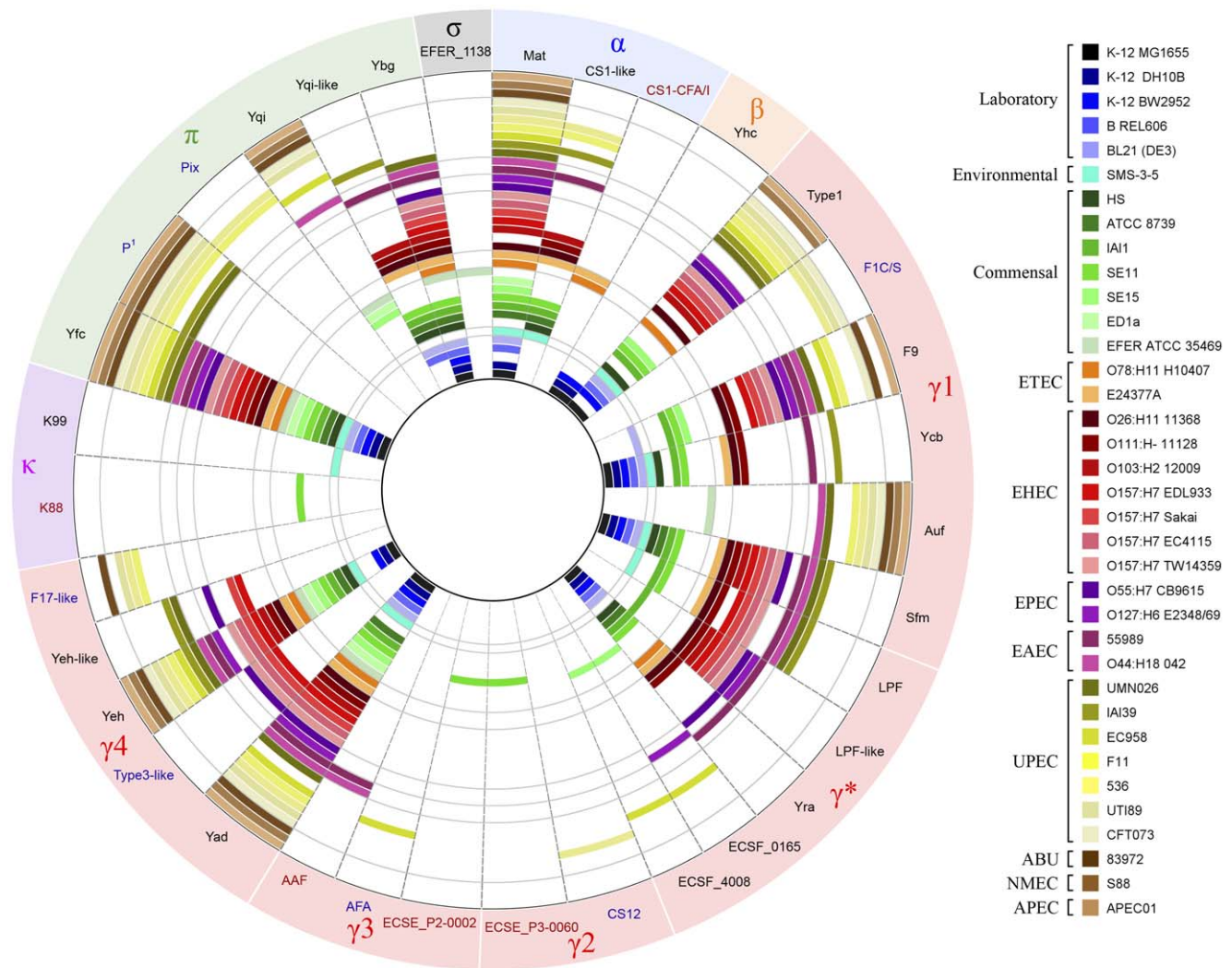


Figure 4. Distribution of CU fimbrial gene clusters among *E. coli* pathotypes. The inner ring represents the concatenated nucleotide sequences of the 38 fimbrial operons. Each segment is labelled in the outer ring according to the name and clade [3] of the corresponding fimbrial usher type with the intervening 36 rings displaying the presence of intact CU fimbrial gene clusters in each of the strains analysed. The legend on the right lists the colour of each strain that we included in our study, grouped according to pathogenicity class. Circular comparison was generated using BLAST ring image generator (BRIG) [69]. ¹CFT073 contains two copies of the P fimbriae operon.
doi:10.1371/journal.pone.0052835.g004

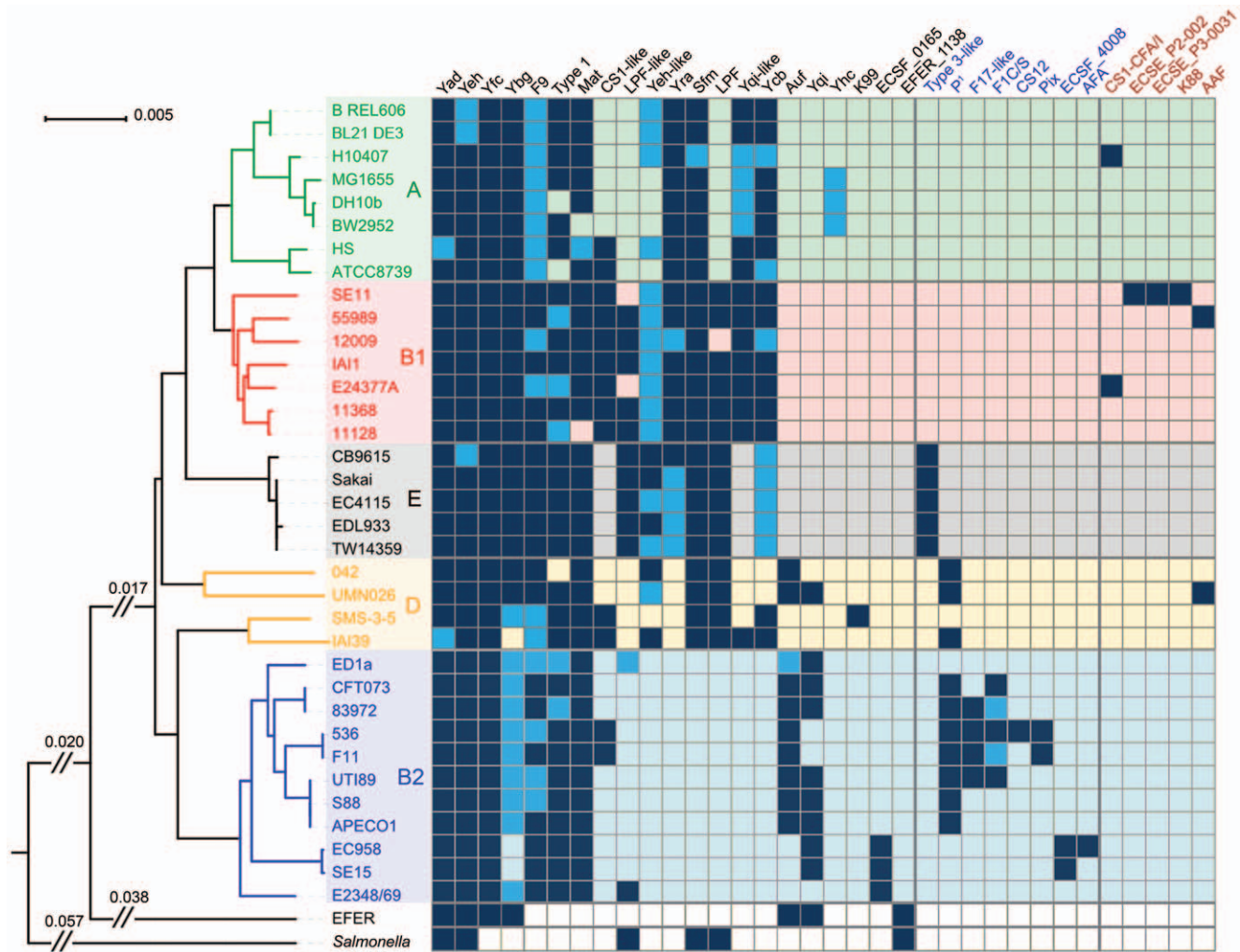


Figure 5. Distribution of *E. coli* fimbrial gene clusters in an evolutionary context. Left: The phylogeny of the *Escherichia* strains is displayed as inferred using the Neighbour Joining method on the concatenated nucleotide sequence of 7 housekeeping genes (~9 kb). *E. coli* strains are colour-coded according to phylogroup (A, B1, B2, D and E). To determine CU fimbrial gene cluster ancestry, the *Salmonella* pan-genome was investigated for the presence of *Escherichia* fimbrial types. The scale indicates the number of substitutions per nucleotide. Right: The names of fimbrial types are displayed along the top of a fimbrial gene cluster matrix, with the names of PAI or plasmid-born CU fimbrial gene clusters highlighted in blue and red, respectively. Dark blue and light blue cells represent intact and disrupted CU fimbrial gene clusters, respectively. The heterogeneous distribution of CU fimbrial types identified in our dataset suggests substantial acquisition and loss of CU fimbrial gene clusters during the evolution of the *Escherichia* genus. Depending on their distribution, CU fimbrial types can be classified as core-associated, clade-specific, or sporadic. ¹CFT073 possesses two copies of the P fimbriae operon. doi:10.1371/journal.pone.0052835.g005

Multiple Sequence Alignment and Phylogenetics

Full-length usher amino acid sequences from intact fimbrial operons (as well as the Yhc and AAF/II ushers) were used to infer evolutionary relationships. Sequences were aligned in ClustalX2.1 [46] using BLOSUM30 for pair-wise alignment with a gap opening penalty of 10 and gap extension penalty of 0.1, and the BLOSUM series matrix for multiple alignment with a gap opening penalty of 10 and a gap extension penalty of 0.2 (default parameters). Phylogenetic analyses were performed with the MEGA5 software package [47]. Protein distance matrices were predicted using the Poisson correction model with default settings. The Neighbour-Joining method was used to generate a phylogenetic tree, which was displayed as an unrooted phylogram using iTOL [48]. To estimate the confidence in the tree topology, a bootstrap test of 1000 replicates was performed. Alignment and phylogenetic tree construction was repeated with usher sequences

of previously published usher phylograms [3] to verify tree validity (data not shown).

The evolutionary relationship of the 35 *E. coli* and one *E. fergusonii* strains included in our analysis was predicted by Multi-Locus Sequence Typing (MLST) of the concatenated nucleotide sequences of 7 housekeeping genes (*adh*, *fumC*, *gyrB*, *icd*, *mdh*, *purA*, *recA*) as previously described [49]. MLST data of *Salmonella enterica* serovar Typhimurium LT2 [50] was incorporated as representative for the *Salmonella* outgroup. Sequences were aligned in ClustalX2.1 using the ClustalW(1.6) DNA weight matrix under default settings. The Neighbour-Joining method of MEGA5 was used to infer the evolutionary history, with distances computed by the Jukes-Cantor method. The resulting phylogenetic tree was visualised in iTOL [48] as a rooted phylogram.

Results and Discussion

Identification of CU fimbrial operons in *Escherichia*

A bioinformatic approach was used to identify CU fimbrial gene clusters in *Escherichia*. Fimbrial operons were identified using an iterative usher BLASTp search against a selection of 36 *Escherichia* complete genomes and 132 *Escherichia* spp. plasmids (Table 1 and S1). These genomes represent strains from ExPEC, diarrheagenic and commensal *E. coli*, as well as *Escherichia fergusonii*. Fimbrial operons were defined as polycistronic gene clusters containing at least an usher and a chaperone encoding sequence, and flanked by one or more genes encoding fimbrial subunits. Usher genes of disrupted operons may be subject to increased change, potentially distorting our interpretation of the phylogenetic relationships amongst usher proteins [51]. To prevent potential bias, CU operons that contained transposon insertion elements or truncated structural genes were considered disrupted and excluded from the evolutionary phylogeny analysis.

A total of 458 CU fimbrial gene clusters were identified from the combined whole genome and plasmid-only datasets. In the whole genome dataset, 449 operons containing usher and chaperone encoding sequences were identified (average 12 ± 2.14 operons per strain; maximum 17, minimum 7) (Table 1 and Table S2). Analysis of the genetic organisation of these CU fimbrial gene clusters revealed that 370 operons were intact (average 10 ± 2.28 intact operons per strain; maximum 16, minimum 5). The vast majority of fimbrial gene clusters in the whole genome dataset were chromosomally located (442/449), while 7 CU fimbrial gene clusters were located on plasmids. In the plasmid-only dataset, another nine CU fimbrial gene clusters were identified, all of which appeared to be intact (Table S2). No orphan usher encoding genes were discovered.

Classification of *Escherichia* fimbriae

To display the evolutionary relationship of the CU fimbriae usher amino acid sequences, an unrooted phylogram was constructed (Figure 1). This analysis included all the 379 usher amino acid sequences from the intact operons described above, as well as four usher sequences from disrupted operons that lacked intact representatives in the dataset (i.e. Yhc, and AAF/II). The CU clading scheme described previously by Nuccio et al. divides Gram-negative CU fimbriae into six clades (α , β , γ , κ , π , σ) and five sub-clades (γ_1 , γ_2 , γ_3 , γ_4 and γ^*), based on the evolutionary phylogeny of usher protein sequences [3]. A phylogenetic tree of *Escherichia* usher sequences based on the Nuccio scheme demonstrated that the *Escherichia* genus contains representatives of all six clades, which were labelled accordingly (Figure 1). The γ clade was the largest and encompassed 24 CU fimbrial types across five sub-clades, with the best-characterised fimbriae represented by type 1 fimbriae. The π clade contained 6 CU fimbriae, including the well-characterised P fimbriae from UPEC. The remaining four clades (α , β , κ , σ) comprised relatively few CU fimbrial types. The α clade was the most distantly related, and this is consistent with the classification of CS1-CFA/I fimbriae as members of an alternate CU pathway [52].

The majority of CU fimbrial operons showed a strong relationship between chromosomal location and usher phylogeny. Accordingly, we superimposed the operon locus of each chromosomal CU type on the *E. coli* MG1655 reference genome (Figure 2). Based on usher phylogeny and locus position, the 458 CU operons identified in *Escherichia* can be classified as 38 fimbrial types. CU fimbrial genes that could not be mapped in this manner were either located on plasmids (i.e. CS1-CFA/I, ECSE_P2-002, ECSE_P3-0031, K88, AAF) or within pathogenicity-associated

islands (PAIs) (i.e. P, F1C, S, Pix and F17-like fimbriae), which are known to exist at various insertion sites on the *E. coli* chromosome backbone. CU operons associated with these mobile elements were typed according to usher phylogeny and conservation of their genetic organisation.

Genetic organisation of CU fimbrial gene clusters

The genetic organisation of the CU fimbrial gene clusters was predicted by reviewing the literature and inspecting individual genes for conserved fimbrial protein domains (Figure 3). In most instances, the genetic structure of operons belonging to the same fimbrial type was conserved. The exceptions were Lpf and K88 fimbriae, where additional subunit genes have been acquired or lost in certain strains. For example, in EHEC O157:H7 and EPEC O55:H7 strains the *lpf* operon contains an additional gene encoding a putative fimbrial subunit protein (COG3539 domain) at its 3'-end. The amino acid sequence of this fimbrial subunit protein shares strong identity (169/367 or 46% identical residues) and similarity (226/367 or 62% similar residues) with the amino acid sequence of the adjacent (conserved) subunit-encoding gene.

Distribution of CU fimbriae among *E. coli* pathotypes

In total, 38 distinct CU fimbrial operons were identified. The distribution of each intact CU fimbrial operon was assessed with respect to *E. coli* pathogenicity class (Figure 4). Five fimbrial types were common to most pathotypes: type 1, Yad, Yeh, Yfc and Mat (Ecp) fimbriae. Type 1 fimbriae, as discussed above, represent the most well characterised CU fimbriae and mediate binding to α -D-mannosylated receptors. The *yad*, *yeh* and *yfc* CU fimbrial genes encode for functional but cryptic surface organelles, and thus their precise role in colonisation remains to be determined [35]. Recently, Yad fimbriae were shown to be associated with adherence to UM-UC-3 bladder epithelial cells and biofilm formation [53], although their expression in wild-type strains remains to be demonstrated. Mat (meningitis associated and temperature regulated) fimbriae were first identified in neonatal meningitis *E. coli* (NMEC) [54] and have subsequently also been named ECP (*E. coli* common pilus) due to their apparent ubiquitous association with most *E. coli* strains [55]. Mat (ECP) fimbriae mediate biofilm formation and adherence to cultured epithelial cells [55,56].

Some of the CU fimbrial genes displayed a clear pathotype association. For example P, F1C/S, F17-like and Pix fimbriae genes were only found in ExPEC strains. P and F1C/S fimbriae are associated with colonisation of the urinary tract. F1C fimbriae bind to galactosylceramide targets present on epithelial cells in the kidneys, ureters and bladder as well as to globotriaosylceramide present only in the kidneys [28,57]. S fimbriae recognize α -sialyl-2,3-galactose receptors present on the surface of host glycoproteins [58]. Pix fimbriae, although functionally characterised from an *E. coli* strain isolated from the urinary tract, do not bind to receptor targets recognized by other UPEC fimbriae [59]. The function of F17-like fimbriae has not been characterised.

Other examples of pathotype association were also apparent. CS1-CFA/I fimbriae, which contribute to intestinal colonisation [60], were strongly associated with ETEC. The Lpf and Lpf-like fimbrial types were predominantly associated with diarrheagenic *E. coli* strains, although there were some exceptions for Lpf fimbriae as they were also detected in the UPEC strains UMN026 and IAI39. Similarly, Ybg fimbriae were predominantly found in commensal and diarrheagenic *E. coli* strains (except for UPEC strain UMN026) and type 3-like fimbriae were only found in EHEC and EPEC strains. In this dataset, AAF fimbriae were only present in EAEC. AFA fimbriae, which contribute to the virulence

of EIEC and UPEC [61], were only present in UPEC strain EC958 in this dataset. The Yhc, ECSE_P2-0002, ECSE_P3-0060, CS12, ECSF-4008, ECSF-0165, K88, K99 and EFER_1138 fimbriae were highly under-represented in the strains selected for our analysis.

Distribution of fimbriae among *Escherichia* lineages

To examine the evolutionary history of the 35 *E. coli* and one *E. fergusonii* strains in our dataset, we constructed a phylogenetic tree based on multi-locus sequence typing (MLST) of the concatenated nucleotide sequences of seven housekeeping genes [49]. Integration of the *Escherichia* phylogeny with the distribution of fimbrial gene clusters enabled us to evaluate the evolutionary history of CU fimbriae in the genus (Figure 5). As the majority of chromosome-borne fimbrial types occupy a single locus position, the most parsimonious evolutionary scenario suggests that the corresponding fimbrial gene clusters were acquired by a common ancestor through horizontal gene transfer or homologous recombination, and subsequently lost or disseminated vertically in its descendants. Exceptions are CU fimbriae located on PAIs. These elements are inherently prone to recombination events and can be found in a number of integration “hot-spots” (typically tRNA sites) relative to the *E. coli* chromosome backbone (Figure 2) [62]. Parsimony inference of the heterogeneous fimbrial presence (complete/partial) or absence pattern reveals extensive gain or loss of CU fimbrial gene clusters during the evolution of *E. coli* [63].

Based on their distribution in *E. coli*, we can divide the CU fimbrial types into three groups: core-associated, clade-specific, and sporadic fimbriae. The core-associated Type 1, Yad, Yeh, Yfc and Mat (Ecp) fimbriae were conserved in the vast majority of *E. coli* strains, suggesting their presence in an *E. coli* common ancestor. These genes were present as intact or disrupted operons at the same genetic locus in almost all strains examined, with only the *yfc* cluster intact in all genomes. The *E. coli mat (ecp)* fimbrial genes are also highly conserved in *Klebsiella pneumoniae* genomes but do not share the same syntenic location. The F9 and Ybg fimbrial gene clusters could also be considered as part of the core-associated group, however these loci are not intact in many strains.

E. coli population genetics have identified five major monophyletic clades (phylogroups A, B1, B2, D and E) [49]. Although these phylogroups do not correlate directly with virulence, some inferences can be made; for example ExPEC strains mainly belong to phylogroups B2 and D, whereas EHEC strains are associated with phylogroups B1 and E. The number of CU fimbrial gene clusters identified from strains in each phylogroup varied as follows: A (n = 8 strains), average of 12 (total) and 9 (intact) CU fimbriae per strain; B1 (n = 7 strains), average of 15 (total) and 13 (intact) CU fimbriae per strain; B2 (n = 11 strains), average of 11 (total) and 9 (intact) CU fimbriae per strain; D (n = 4 strains), average of 13 (total) and 11 (intact) CU fimbriae per strain; E (n = 5 strains), average of 14 (total) and 12 (intact) CU fimbriae per strain. Clade-specific fimbriae were associated with one or more *E. coli* phylogroups. An example can be observed in the case of Yqi/Yqi-like fimbriae and Auf/Ycb fimbriae. These fimbrial types occupy various locus positions on the bacterial genome and are closely related but mutually exclusive. The Auf and Yqi operons were common to the B2 phylogroup, while the Ycb and Yqi-like operons were associated with the A, B1, D and E phylogroups. The CU fimbrial profile of *E. fergusonii* is most similar to the *E. coli* B2 phylogroup, which exhibits the most ancient divergence from the A, B1, D and E phylogroups [64].

Sporadic fimbriae located on the chromosome (e.g. Yhc, K99, ECSF_0165) may represent remnants of ancient CU fimbrial gene clusters lost in the majority of strains, or, as in the case of PAI-

associated fimbriae, genes that were acquired more recently. Further analysis of these fimbrial gene clusters in a larger genome dataset is required before additional conclusions can be drawn on their prevalence in the *E. coli* pan-genome. This group of fimbriae also includes plasmid-borne gene clusters, which by definition are more likely to be associated with horizontal gene transfer.

Comparative analysis of *Escherichia* and *Salmonella* CU fimbrial gene clusters

To gain a broader insight into the evolution of CU fimbriae in *Escherichia*, the *Salmonella* pan-genome (NCBI database) was investigated for the presence of the 38 fimbrial types identified in our study. *Salmonella* and *Escherichia* diverged from a common ancestor approximately 100 million years ago [65]. Nevertheless, we identified six CU fimbrial types which were conserved in both genera; Yad, Yeh, Sfm, Lpf, Lpf-like and EFER_1138 (corresponding to Sta, Stc, Fim, Stg, Lpf and SARI_01025 in *Salmonella*, respectively) [50,66,67]. These CU fimbrial gene clusters clade together according to usher phylogeny and occupy an identical locus position relative to the MG1655 genome (data not shown), indicating that they are ancient and were present in the common ancestor of the two genera (Figure 5). Yeh fimbriae also occupy the same locus in *Citrobacter* [68]. Although the Yad and Yeh fimbriae are highly conserved in extant *E. coli* strains (intact in 94% and 92% of our strain database, respectively), other ancestral CU fimbrial gene clusters have been lost in one or several of the *E. coli* phylogroups. For example, Sfm fimbriae (annotated as type 1 fimbriae in *Salmonella*) were present in all *E. coli* phylogroups except for B2 and *E. fergusonii*, suggesting that the *sfm* gene cluster was acquired by an ancient ancestor of these genera and later separately lost by the *E. coli* B2 and *E. fergusonii* phylogroup progenitors. Phylogenetic analysis of the Sfm usher amino acid sequences of *Escherichia*, *Salmonella*, *Citrobacter* and *Enterobacter* supports this hypothesis (data not shown). EFER_1138 is conserved in *E. fergusonii*, *Salmonella*, *Citrobacter*, *Enterobacter* and *Cronobacter*, however no remnants of this archaic fimbrial gene cluster were detected in *E. coli*.

Conclusions

CU fimbriae are cell surface-located organelles produced by many Gram-negative bacteria. These fimbriae have been best studied in *E. coli*, where they contribute to adherence, colonisation, tissue tropism and biofilm formation. The generic CU fimbrial gene cluster comprises at least four genes, encoding a chaperone, usher, major subunit and adhesin. In this study, we identified 38 CU fimbrial types from a comprehensive genome and plasmid dataset that represents a diverse array of *E. coli* strains. The majority of these fimbrial types belonged to the γ clade based on usher phylogeny, however representatives from all other previously defined clades were identified. Most of the CU fimbrial gene clusters were located in syntenic locus positions on the different *E. coli* chromosomes, and these were mapped to various locations relative to the *E. coli* K12 MG1655 reference genome. Less common CU fimbrial gene clusters were often associated with PAIs or plasmid-borne. A group of core-associated *E. coli* CU fimbriae were defined, yet interestingly few of these fimbriae have been properly characterised. The diversity of CU fimbrial gene clusters identified in this study highlights several deficiencies in our knowledge of these structural organelles. While some CU fimbriae such as type 1 and P fimbriae have been comprehensively studied, little is known about the regulation and function of many other CU fimbrial types, some of which are cryptic in nature. This study provides a framework for the effective characterisation and

functional analysis of the complete subset of *E. coli* CU fimbriae, and should enable comprehensive typing of these fimbriae based on their chromosome location and evolutionary history.

Supporting Information

Table S1 Plasmids analysed in this study.
(DOCX)

Table S2 Fimbriae identified in *Escherichia* genomes and plasmids.
(XLSX)

References

- Proft T, Baker EN (2009) Pili in Gram-negative and Gram-positive bacteria – structure, assembly and their role in disease. *Cell Mol Life Sci* 66: 613–635.
- Kline KA, Falker S, Dahlberg S, Normark S, Henriques-Normark B (2009) Bacterial adhesins in host-microbe interactions. *Cell Host Microbe* 5: 580–592.
- Nuccio SP, Baumler AJ (2007) Evolution of the chaperone/usher assembly pathway: fimbrial classification goes Greek. *Microbiol Mol Biol Rev* 71: 551–575.
- Sauer FG, Remaut H, Hultgren SJ, Waksman G (2004) Fiber assembly by the chaperone-usher pathway. *Biochim Biophys Acta* 1694: 259–267.
- Sauer FG, Futterer K, Pinkner JS, Dodson KW, Hultgren SJ, et al. (1999) Structural basis of chaperone function and pilus biogenesis. *Science* 285: 1058–1061.
- Zavialov A, Berglund J, Knight SD (2003) Overexpression, purification, crystallization and preliminary X-ray diffraction analysis of the F1 antigen Caf1M-Caf1 chaperone-subunit pre-assembly complex from *Yersinia pestis*. *Acta Crystallogr D Biol Crystallogr* 59: 359–362.
- Zavialov AV, Tischenko VM, Fooks LJ, Brandsdal BO, Aqvist J, et al. (2005) Resolving the energy paradox of chaperone/usher-mediated fibre assembly. *Biochem J* 389: 685–694.
- Zavialov AV, Knight SD (2007) A novel self-capping mechanism controls aggregation of periplasmic chaperone Caf1M. *Mol Microbiol* 64: 153–164.
- Vergers D, Bullitt E, Hultgren SJ, Waksman G (2007) Crystal structure of the P pilus rod subunit PapA. *PLoS Pathog* 3: e73.
- Salih O, Remaut H, Waksman G, Orlova EV (2008) Structural analysis of the Saf pilus by electron microscopy and image processing. *J Mol Biol* 379: 174–187.
- Waksman G, Hultgren SJ (2009) Structural biology of the chaperone-usher pathway of pilus biogenesis. *Nat Rev Microbiol* 7: 765–774.
- Klemm P (1986) Two regulatory fim genes, fimB and fimE, control the phase variation of type 1 fimbriae in *Escherichia coli*. *EMBO J* 5: 1389–1393.
- Mu XQ, Bullitt E (2006) Structure and assembly of P-pili: a protruding hinge region used for assembly of a bacterial adhesion filament. *Proc Natl Acad Sci U S A* 103: 9861–9866.
- Kuehn MJ, Heuser J, Normark S, Hultgren SJ (1992) P pili in uropathogenic *E. coli* are composite fibres with distinct fibrillar adhesive tips. *Nature* 356: 252–255.
- Holden N, Totsika M, Dixon L, Catherwood K, Gally DL (2007) Regulation of P-fimbrial phase variation frequencies in *Escherichia coli* CFT073. *Infect Immun* 75: 3325–3334.
- Hahn E, Wild P, Hermans U, Sebbel P, Glockshuber R, et al. (2002) Exploring the 3D molecular architecture of *Escherichia coli* type 1 pili. *J Mol Biol* 323: 845–857.
- Totsika M, Beatson SA, Holden N, Gally DL (2008) Regulatory interplay between pap operons in uropathogenic *Escherichia coli*. *Mol Microbiol* 67: 996–1011.
- Jones CH, Pinkner JS, Roth R, Heuser J, Nicholes AV, et al. (1995) FimH adhesin of type 1 pili is assembled into a fibrillar tip structure in the Enterobacteriaceae. *Proc Natl Acad Sci U S A* 92: 2081–2085.
- Choudhury D, Thompson A, Stojanoff V, Langermann S, Pinkner J, et al. (1999) X-ray structure of the FimC-FimH chaperone-adhesin complex from uropathogenic *Escherichia coli*. *Science* 285: 1061–1066.
- Wu XR, Sun TT, Medina JJ (1996) In vitro binding of type 1-fimbriated *Escherichia coli* to uroplakins Ia and Ib: relation to urinary tract infections. *Proc Natl Acad Sci U S A* 93: 9630–9635.
- Connell I, Agace W, Klemm P, Schembri M, Marild S, et al. (1996) Type 1 fimbrial expression enhances *Escherichia coli* virulence for the urinary tract. *Proc Natl Acad Sci U S A* 93: 9827–9832.
- Anderson GG, Palermo JJ, Schilling JD, Roth R, Heuser J, et al. (2003) Intracellular bacterial biofilm-like pods in urinary tract infections. *Science* 301: 105–107.
- Mulvey MA, Lopez-Boado YS, Wilson CL, Roth R, Parks WC, et al. (1998) Induction and evasion of host defenses by type 1-piliated uropathogenic *Escherichia coli*. *Science* 282: 1494–1497.
- Bergsten G, Samuelsson M, Wullt B, Leijonhufvud I, Fischer H, et al. (2004) PapG-dependent adherence breaks mucosal inertia and triggers the innate host response. *J Infect Dis* 189: 1734–1742.
- Roberts JA, Kaack MB, Baskin G, Chapman MR, Hunstad DA, et al. (2004) Antibody responses and protection from pyelonephritis following vaccination with purified *Escherichia coli* PapDG protein. *J Urol* 171: 1682–1685.
- Roberts JA, Marklund BI, Ilver D, Haslam D, Kaack MB, et al. (1994) The Gal(alpha 1-4)Gal-specific tip adhesin of *Escherichia coli* P-fimbriae is needed for pyelonephritis to occur in the normal urinary tract. *Proc Natl Acad Sci U S A* 91: 11889–11893.
- Hedges SR, Agace WW, Svanborg C (1995) Epithelial cytokine responses and mucosal cytokine networks. *Trends Microbiol* 3: 266–270.
- Khan AS, Kniep B, Oelschlaeger TA, Van Die I, Korhonen T, et al. (2000) Receptor structure for F1C fimbriae of uropathogenic *Escherichia coli*. *Infect Immun* 68: 3541–3547.
- Korhonen TK, Parkkinen J, Hacker J, Finne J, Pere A, et al. (1986) Binding of *Escherichia coli* S fimbriae to human kidney epithelium. *Infect Immun* 54: 322–327.
- Savarino SJ, Fox P, Deng Y, Nataro JP (1994) Identification and characterization of a gene cluster mediating enteroaggregative *Escherichia coli* aggregative adherence fimbria I biogenesis. *J Bacteriol* 176: 4949–4957.
- Jordan DM, Cornick N, Torres AG, Dean-Nystrom EA, Kaper JB, et al. (2004) Long polar fimbriae contribute to colonization by *Escherichia coli* O157:H7 in vivo. *Infect Immun* 72: 6168–6171.
- Sakellaris H, Munson GP, Scott JR (1999) A conserved residue in the tip proteins of CS1 and CFA/I pili of enterotoxigenic *Escherichia coli* that is essential for adherence. *Proc Natl Acad Sci U S A* 96: 12828–12832.
- Kyogashima M, Ginsburg V, Krivan HC (1989) *Escherichia coli* K99 binds to N-glycolylsialoparagloboside and N-glycolyl-GM3 found in piglet small intestine. *Arch Biochem Biophys* 270: 391–397.
- Bakker D, Willemsen PT, Simons LH, van Zijderveld FG, de Graaf FK (1992) Characterization of the antigenic and adhesive properties of FaeG, the major subunit of K88 fimbriae. *Mol Microbiol* 6: 247–255.
- Korea CG, Badouraly R, Prevost MC, Ghigo JM, Beloin C (2010) *Escherichia coli* K-12 possesses multiple cryptic but functional chaperone-usher fimbriae with distinct surface specificities. *Environ Microbiol* 12: 1957–1977.
- Ottow JC (1975) Ecology, physiology, and genetics of fimbriae and pili. *Annu Rev Microbiol* 29: 79–108.
- Gaastra W, Svennerholm AM (1996) Colonization factors of human enterotoxigenic *Escherichia coli* (ETEC). *Trends Microbiol* 4: 444–452.
- Orskov I, Orskov F (1990) Serologic classification of fimbriae. *Curr Top Microbiol Immunol* 151: 71–90.
- Altschul SF, Wootton JC, Gertz EM, Agarwala R, Morgulis A, et al. (2005) Protein database searches using compositionally adjusted substitution matrices. *FEBS J* 272: 5101–5109.
- Jain E, Bairoch A, Duvaud S, Phan I, Redaschi N, et al. (2009) Infrastructure for the life sciences: design and implementation of the UniProt website. *BMC Bioinformatics* 10: 136.
- Welch RA, Burland V, Plunkett G 3rd, Redford P, Roesch P, et al. (2002) Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc Natl Acad Sci U S A* 99: 17020–17024.
- Luo C, Hu GQ, Zhu H (2009) Genome reannotation of *Escherichia coli* CFT073 with new insights into virulence. *BMC Genomics* 10: 552.
- Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, et al. (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res* 39: D225–229.
- Chaudhuri RR, Loman NJ, Snyder LA, Bailey CM, Stekel DJ, et al. (2008) xBASE2: a comprehensive resource for comparative bacterial genomics. *Nucleic Acids Res* 36: D543–546.
- Blattner FR, Plunkett G 3rd, Bloch CA, Perna NT, Burland V, et al. (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* 277: 1453–1462.
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947–2948.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28: 2731–2739.
- Letunic I, Bork P (2007) Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 23: 127–128.

Acknowledgments

We would like to thank Nabil-Fareed Alikhan for his assistance with the circular comparison of genetic regions.

Author Contributions

Conceived and designed the experiments: DJW SAB MT NKP MAS. Performed the experiments: DJW. Analyzed the data: DJW SAB MAS. Contributed reagents/materials/analysis tools: DJW SAB MT NKP. Wrote the paper: DJW SAB MAS.

49. Wirth T, Falush D, Lan R, Colles F, Mensa P, et al. (2006) Sex and virulence in *Escherichia coli*: an evolutionary perspective. *Mol Microbiol* 60: 1136–1151.
50. McClelland M, Sanderson KE, Spieth J, Clifton SW, Latreille P, et al. (2001) Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature* 413: 852–856.
51. Kuo CH, Ochman H (2010) The extinction dynamics of bacterial pseudogenes. *PLoS Genet* 6.
52. Soto GE, Hultgren SJ (1999) Bacterial adhesins: common themes and variations in architecture and assembly. *J Bacteriol* 181: 1059–1071.
53. Spurbeck RR, Stapleton AE, Johnson JR, Walk ST, Hooton TM, et al. (2011) Fimbrial profiles predict virulence of uropathogenic *Escherichia coli* strains: contribution of *ygi* and *yad* fimbriae. *Infect Immun* 79: 4753–4763.
54. Pouttu R, Westerlund-Wikstrom B, Lang H, Alsti K, Virkola R, et al. (2001) *matB*, a common fimbriin gene of *Escherichia coli*, expressed in a genetically conserved, virulent clonal group. *J Bacteriol* 183: 4727–4736.
55. Rendon MA, Saldana Z, Erdem AL, Monteiro-Neto V, Vazquez A, et al. (2007) Commensal and pathogenic *Escherichia coli* use a common pilus adherence factor for epithelial cell colonization. *Proc Natl Acad Sci U S A* 104: 10637–10642.
56. Saldana Z, Erdem AL, Schuller S, Okeke IN, Lucas M, et al. (2009) The *Escherichia coli* common pilus and the bundle-forming pilus act in concert during the formation of localized adherence by enteropathogenic *E. coli*. *J Bacteriol* 191: 3451–3461.
57. Backhed F, Alsen B, Roche N, Angstrom J, von Euler A, et al. (2002) Identification of target tissue glycosphingolipid receptors for uropathogenic, F1C-fimbriated *Escherichia coli* and its role in mucosal inflammation. *J Biol Chem* 277: 18198–18205.
58. Parkkinen J, Rogers GN, Korhonen T, Dahr W, Finne J (1986) Identification of the O-linked sialyloligosaccharides of glycoporphin A as the erythrocyte receptors for S-fimbriated *Escherichia coli*. *Infect Immun* 54: 37–42.
59. Lugerling A, Benz I, Knochenhauer S, Ruffing M, Schmidt MA (2003) The Pix pilus adhesin of the uropathogenic *Escherichia coli* strain X2194 (O2: K(-): H6) is related to Pap pili but exhibits a truncated regulatory region. *Microbiology* 149: 1387–1397.
60. Sakellaris H, Balding DP, Scott JR (1996) Assembly proteins of CS1 pili of enterotoxigenic *Escherichia coli*. *Mol Microbiol* 21: 529–541.
61. Le Bouguenec C, Garcia MI, Quin V, Desperrier JM, Gounon P, et al. (1993) Characterization of plasmid-borne *afa-3* gene clusters encoding afimbrial adhesins expressed by *Escherichia coli* strains associated with intestinal or urinary tract infections. *Infect Immun* 61: 5106–5114.
62. Touchon M, Hoede C, Tenailon O, Barbe V, Baeriswyl S, et al. (2009) Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet* 5: e1000344.
63. Zhaxybayeva O, Nesbo CL, Doolittle WF (2007) Systematic overestimation of gene gain through false diagnosis of gene absence. *Genome Biol* 8: 402.
64. Sims GE, Kim SH (2011) Whole-genome phylogeny of *Escherichia coli*/Shigella group by feature frequency profiles (FFPs). *Proc Natl Acad Sci U S A* 108: 8329–8334.
65. Ochman H, Wilson AC (1987) Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. *J Mol Evol* 26: 74–86.
66. Yue M, Rankin SC, Blanchet RT, Nulton JD, Edwards RA, et al. (2012) Diversification of the salmonella fimbriae: a model of macro- and microevolution. *PLoS One* 7: e38596.
67. Townsend SM, Kramer NE, Edwards R, Baker S, Hamlin N, et al. (2001) *Salmonella enterica* serovar Typhi possesses a unique repertoire of fimbrial gene sequences. *Infect Immun* 69: 2894–2901.
68. Petty NK, Bulgin R, Crepin VF, Cerdeno-Tarraga AM, Schroeder GN, et al. (2010) The *Citrobacter rodentium* genome sequence reveals convergent evolution with human pathogenic *Escherichia coli*. *J Bacteriol* 192: 525–538.
69. Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA (2011) BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 12: 402.
70. Hochhut B, Wilde C, Balling G, Middendorf B, Dobrindt U, et al. (2006) Role of pathogenicity island-associated integrases in the genome plasticity of uropathogenic *Escherichia coli* strain 536. *Mol Microbiol* 61: 584–595.
71. Rasko DA, Rosovitz MJ, Myers GS, Mongodin EF, Fricke WF, et al. (2008) The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *J Bacteriol* 190: 6881–6893.
72. Chen SL, Hung CS, Xu J, Reigstad CS, Magrini V, et al. (2006) Identification of genes subject to positive selection in uropathogenic strains of *Escherichia coli*: a comparative genomics approach. *Proc Natl Acad Sci U S A* 103: 5977–5982.
73. Totsika M, Beatson SA, Sarkar S, Phan MD, Petty NK, et al. (2011) Insights into a multidrug resistant *Escherichia coli* pathogen of the globally disseminated ST131 lineage: genome analysis and virulence mechanisms. *PLoS One* 6: e26578.
74. Zdziarski J, Brzuszkiewicz E, Wullt B, Liesegang H, Biran D, et al. (2010) Host imprints on bacterial genomes—rapid, divergent evolution in individual patients. *PLoS Pathog* 6: e1001078.
75. Johnson TJ, Johnson SJ, Nolan LK (2006) Complete DNA sequence of a ColBM plasmid from avian pathogenic *Escherichia coli* suggests that it evolved from closely related ColV virulence plasmids. *J Bacteriol* 188: 5975–5983.
76. Chaudhuri RR, Sebahia M, Hobman JL, Webber MA, Leyton DL, et al. (2010) Complete genome sequence and comparative metabolic profiling of the prototypical enteroaggregative *Escherichia coli* strain O42. *PLoS One* 5: e8801.
77. Iguchi A, Thomson NR, Ogura Y, Saunders D, Ooka T, et al. (2009) Complete genome sequence and comparative genome analysis of enteropathogenic *Escherichia coli* O127:H6 strain E2348/69. *J Bacteriol* 191: 347–354.
78. Zhou Z, Li X, Liu B, Beutin L, Xu J, et al. (2010) Derivation of *Escherichia coli* O157:H7 from its O55:H7 precursor. *PLoS One* 5: e8700.
79. Crossman LC, Chaudhuri RR, Beatson SA, Wells TJ, Desvaux M, et al. (2010) A commensal gene bad: complete genome sequence of the prototypical enterotoxigenic *Escherichia coli* strain H10407. *J Bacteriol* 192: 5822–5831.
80. Ogura Y, Ooka T, Iguchi A, Toh H, Asadulghani M, et al. (2009) Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*. *Proc Natl Acad Sci U S A* 106: 17939–17944.
81. Perna NT, Plunkett G 3rd, Burland V, Mau B, Glasner JD, et al. (2001) Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* 409: 529–533.
82. Hayashi T, Makino K, Ohnishi M, Kurokawa K, Ishii K, et al. (2001) Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res* 8: 11–22.
83. Eppinger M, Mammel MK, Leclerc JE, Ravel J, Cebula TA (2011) Genomic anatomy of *Escherichia coli* O157:H7 outbreaks. *Proc Natl Acad Sci U S A* 108: 20142–20147.
84. Kulasekara BR, Jacobs M, Zhou Y, Wu Z, Sims E, et al. (2009) Analysis of the genome of the *Escherichia coli* O157:H7 2006 spinach-associated outbreak isolate indicates candidate genes that may enhance virulence. *Infect Immun* 77: 3713–3721.
85. Grigoriev IV, Nordberg H, Shabalov I, Aerts A, Cantor M, et al. (2012) The genome portal of the Department of Energy Joint Genome Institute. *Nucleic Acids Res* 40: D26–32.
86. Oshima K, Toh H, Ogura Y, Sasamoto H, Morita H, et al. (2008) Complete genome sequence and comparative analysis of the wild-type commensal *Escherichia coli* strain SE11 isolated from a healthy adult. *DNA Res* 15: 375–386.
87. Toh H, Oshima K, Toyoda A, Ogura Y, Ooka T, et al. (2010) Complete genome sequence of the wild-type commensal *Escherichia coli* strain SE15, belonging to phylogenetic group B2. *J Bacteriol* 192: 1165–1166.
88. Fricke WF, Wright MS, Lindell AH, Harkins DM, Baker-Austin C, et al. (2008) Insights into the environmental resistance gene pool from the genome sequence of the multidrug-resistant environmental isolate *Escherichia coli* SMS-3-5. *J Bacteriol* 190: 6779–6794.
89. Jeong H, Barbe V, Lee CH, Vallenet D, Yu DS, et al. (2009) Genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3). *J Mol Biol* 394: 644–652.
90. Durfee T, Nelson R, Baldwin S, Plunkett G 3rd, Burland V, et al. (2008) The complete genome sequence of *Escherichia coli* DH10B: insights into the biology of a laboratory workhorse. *J Bacteriol* 190: 2597–2606.
91. Ferenci T, Zhou Z, Betteridge T, Ren Y, Liu Y, et al. (2009) Genomic sequencing reveals regulatory mutations and recombinational events in the widely used MC4100 lineage of *Escherichia coli* K-12. *J Bacteriol* 191: 4025–4029.