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. 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].
| Phylogroup | Total CU Operons | Intact CU Operons | Reference |
|------------|------------------|-------------------|-----------|
| UPEC       |                  |                   |           |
| CFT073     | 12               | 11                | Welch et al. 2002 [41] |
| 536        | 14               | 12                | Hochhut et al. 2006 [70] |
| F11        | 13               | 11                | Rasko et al. 2006 [71] |
| UT189      | 12               | 10                | Chen et al. 2006 [72] |
| EC958      | 10               | 9                 | Totsika et al. 2011 [73] |
| UMN026     | 12               | 12                | Touchon et al. 2009 [62] |
| IA139      | 14               | 11                | Touchon et al. 2009 [62] |
| ABU        |                  |                   |           |
| 83972      | 12               | 9                 | Zdziarski et al. 2010 [74] |
| NMEC       |                  |                   |           |
| S88        | 10               | 8                 | Touchon et al. 2006 [62] |
| APEC       |                  |                   |           |
| APEC01     | 10               | 9                 | Johnson et al. 2006 [75] |
| EAEC       |                  |                   |           |
| 55989      | 16               | 14                | Touchon et al. 2009 [62] |
| 042        | 12               | 11                | Chaudhuri et al. 2010 [76] |
| EPEC       |                  |                   |           |
| O127:H6 E2348/69 | 9               | 8                 | Iguchi et al. 2009 [77] |
| O55:H7 CB9615 | 14             | 12                | Zhou et al. 2010 [78] |
| ETEC       |                  |                   |           |
| O78:H11 H10407 | 13             | 8                 | Crossman et al. 2010 [79] |
| E24377A    | 15               | 12                | Rasko et al. 2006 [71] |
| EHEC       |                  |                   |           |
| O26:H11 11368 | 15             | 14                | Ogura et al. 2009 [80] |
| O103:H2 12009 | 14             | 10                | Ogura et al. 2009 [80] |
| O111:H-11128 | 14             | 12                | Ogura et al. 2009 [80] |
| O157:H7 EDL933 | 14             | 12                | Perna et al. 2001 [81] |
| O157:H7 Sakai | 14             | 12                | Hayashi et al. 2001 [82] |
| O157:H7 EC4115 | 14             | 11                | Epping et al. 2011 [83] |
| O157:H7 TW14359 | 14             | 11                | Kulasekara et al. 2009 [84] |
| Commensal  |                  |                   |           |
| ATCC 8739  | 11               | 9                 | Joint Genome Institute [85] |
| IAI1       | 15               | 14                | Touchon et al. 2009 [62] |
| ED1a       | 10               | 5                 | Touchon et al. 2009 [62] |
| HS         | 13               | 9                 | Rasko et al. 2006 [71] |
| SE11       | 17               | 16                | Oshima et al. 2008 [86] |
| SE15       | 9                | 8                 | Toh et al. 2010 [87] |
| Environmental |              |                   |           |
| SMS-3-5    | 12               | 10                | Fricke et al. 2008 [88] |
| Laboratory |                  |                   |           |
| BL21(DE3)  | 12               | 9                 | Jeong et al. 2009 [89] |
| B REL606   | 12               | 9                 | Jeong et al. 2009 [89] |
| K-12 MG1655 | 12              | 9                 | Blattner et al. 1997 [45] |
| K-12 DH10β | 11               | 8                 | Durfee et al. 2008 [90] |
Table 1. Cont.

| E. coli | Phylogroup | Total CU Operons | Intact CU Operons | Reference |
|---------|------------|------------------|-------------------|-----------|
| K-12 BW2952 | A | 11 | 8 | Ferenci et al. 2009 [91] |

| E. fergusonii | Total CU Operons | Intact CU Operons | Reference |
|---------------|------------------|-------------------|-----------|
| Commensal | - | 7 | 7 | Touchon et al. 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

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/ll), 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, π, κ, 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

*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.
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. In this study, CU fimbriae phylogeny was 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. fergusonnii*) 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 coli*. 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
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 4: Distribution of CU fimbrial gene clusters among *E. coli* pathotypes.
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 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 the 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 phylogenograms [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 (\(\text{adk}, \text{fumC}, \text{gyrB}, \text{icd}, \text{mdh}, \text{purA}, \text{recA}\)) 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.

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

Chaperone-Usher Fimbriae of Escherichia coli

PLOS ONE | www.plosone.org 7 January 2013 | Volume 8 | Issue 1 | e52835
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 polyclustronic gene clusters containing at least an usher and a chaperone encoding sequence, and flanked by one or more genes encoding fimbral 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 438 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 fimbral 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 (γ1, γ2, γ3, γ4 and γ5), 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 encompasses 24 CU fimbrial types across five sub-clades, with the best-characterised fimbriae represented by type 1 fimbiae. The π clade contained 6 CU fimbiae, including the well-characterised P fimbiae 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 fimbiae as members of an alternate CU pathway [52].

The major CU 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 fimbral 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 fimbiae), 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 fimbral protein domains (Figure 3). In most instances, the genetic structure of operons belonging to the same fimbral type was conserved. The exceptions were Lpf and K88 fimbiae, 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 fimbral subunit protein (COG3339 domain) at its 3’-end. The amino acid sequence of this fimbral 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 fimbral types were common to most pathotypes: type 1, Yad, Yeh, Yfc and Mat (Ecp) fimbiae. Type 1 fimbiae, as discussed above, represent the most well characterised CU fimbiae and mediate binding to α-D-mannosylated receptors. The yad, yeh and yfj CU fimbral genes encode for functional but cryptic surface organelles, and thus their precise role in colonisation remains to be determined [35]. Recently, Yad fimbiae 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) fimbiae 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) fimbiae 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 fimbiae genes were only found in ExPEC strains. P and F1C/S fimbiae are associated with colonisation of the urinary tract. F1C fimbiae 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 fimbiae recognize α-sialyl-2,3-galactose receptors present on the surface of host glycoproteins [58]. Pix fimbiae, although functionally characterised from an E. coli strain isolated from the urinary tract, do not bind to receptor targets recognized by other UPEC fimbiae [59]. The function of F17-like fimbiae has not been characterised.

Other examples of pathotype association were also apparent. CS1-CFA/I fimbiae, which contribute to intestinal colonisation [60], were strongly associated with ETEC. The Lpf and Lpl-like fimbral types were predominantly associated with diarrheagenic E. coli strains, although there were some exceptions for Lpf fimbiae as they were also detected in the UPEC strains UMN026 and IA139. Similarly, Ygb fimbiae were predominantly found in commensal and diarrheagenic E. coli strains (except for UPEC strain UMN026) and type 3-like fimbiae were only found in EHEC and EPEC strains. In this dataset, AAF fimbiae were only present in EAEC. AFA fimbiae, 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, ECSE-4008, ECSEF_0165, K99, 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, and D and E) [49]. Although these phylogroups do not correlate directly with virulence, some inferences can be made; for example ExPEC strains mainly associate group, however these loci are not intact in many strains.associate 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 chromosomal location and evolutionary history.

**Supporting Information**

**Table S1 Plasmids analysed in this study.**

**Table S2 Fimbriae identified in *Escherichia coli* genomes and plasmids.**

**XLSX**

**References**

1. 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.
2. Kline KA, Falkner S, Dahlberg S, Normark S, Henrique-Normark B (2009) Bacterial adhesins in host-microbe interactions. Cell Host Microbe 5: 590–592.
3. Nuccio SP, Baumann AJ (2007) Evolution of the chaperone/usher assembly pathway: fimbral classification goes Greek. Microbiol Mol Biol Rev 71: 551–575.
4. Sauer FG, Rernant H, Hultgren SJ, Waksman G (2004) Fiber assembly by the chaperone-usher pathway. Biochim Biophys Acta 1694: 259–267.
5. Sauer FG, Futterer K, Pinkner JS, Dodson KW, Hultgren SJ, et al. (1999) Structural basis of chaperone function and pilus biogenesis. Science 283: 1036–1040.
6. Zavialov A, Berglund J, Knight SD (2003) Overexpression, purification, crystallization and preliminary X-ray diffraction analysis of the F1 antigen Café1M-Caf1 chaperone-subunit pre-assembly complex from Yersinia pestis. Acta Crystallogr D Biol Crystallogr 59: 359–362.
7. Zavialov AV, Tischken VM, Fooks LJ, Brandalsdal BO, Aqvist J, et al. (2005) Resolving the energy paradox of chaperone-usher-mediated fibre assembly. Biochem J 389: 665–694.
8. Zavialov AV, Knight SD (2007) A novel self-capping mechanism controls aggregation of periplasmic chaperone Café1M. Mol Microbiol 64: 153–164.
9. Verger D, Bullitt E, Hultgren SJ, Waksman G (2007) Crystal structure of the C plus rod subunit PapA. PLoS Pathog 3: e73.
10. Saleh O, Rernant H, Waksman G, Orlova EV (2008) Structural analysis of the Caf plus rods by electron microscopy and image processing. J Mol Biol 379: 174–187.
11. Waksman G, Hultgren SJ (2009) Structural biology of the chaperone-usher pathway of pilus biogenesis. Nat Rev Microbiol 7: 765–774.
12. 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.
13. Mu XQ, Bullitt E (2006) Structure and assembly of P-pili: a protruding hinge between pap operons in *Uropathogenic Escherichia coli*. Mol Microbiol 67: 996–1007.
14. Holden N, Totsika M, Dixon L, Catherwood K, Gally DL (2007) Regulation of pilus biogenesis. FEBS J 272: 5101–5109.
15. 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.
16. Sokellari 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: 12928–12932.
17. Kuehn MJ, Heuser J, Normark S, Hultgren SJ (1992) P pilus in *Uropathogenic* *Escherichia coli*. Science 282: 1494–1497.
18. Roberts JA, Marklund BI, Iver D, Haslam D, Kaack MB, et al. (1994) The Galαβ1-4Gal-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: 11089–11093.
19. 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.
20. Savarino SJ, Fox P, Deng Y, Nataro JP (1994) Identification and characterization of a gene cluster mediating enterogauggerative *Escherichia coli* agglutinative adherence fimbria I biogenesis. J Bacteriol 176: 4499–4507.
21. Roberts JA, Marklund BI, Iver D, Haslam D, Kaack MB, et al. (1994) The Galαβ1-4Gal-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: 11089–11093.
22. Nagaswami M, Ginsburg V, Krivan HC (1989) *Escherichia coli* K99 binds to N-glycolylidolapargaloboside and N-glycolylGM3 found in piglet small intestine. Arch Biochem Biophys 280: 391–397.
23. Bolker D, Willemsen PT, Simons LHJ, van Zijlveeld GF, de Graaf FK (1992) Characterization of the antigenic and adhesiv properties of FaeG, the major subunit of K88 fimbriae. Mol Microbiol 6: 247–255.
24. Kornberg MD, Bodoury M, 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.
25. Chaudhuri RR, Loman NJ, Snyder LA, Bailey CM, Stekel DJ, et al. (2008) Protein database searches using compositionally adjusted substitution matrices. FEBS J 272: 5101–5109.
26. 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.
27. Welch RA, Borland V, Phanakit 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.
28. Lu G, Guo GQ, Zhu H (2009) Genome reannotation of *Escherichia coli* CFT073 with new insights into virulence. BMC Genomics 10: 552.
29. Schutl SE, Woostec JG, Gerz EM, Agarwala R, Mouglistu A, et al. (2005) Protein database searches using compositionally adjusted substitution matrices. FEBS J 272: 5101–5109.
30. CFT073 with new insights into virulence. BMC Genomics 10: 552.
31. 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.
32. Sokellari 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: 12928–12932.
33. Kornberg MD, Bodoury M, 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.
34. Chaudhuri RR, Loman NJ, Snyder LA, Bailey CM, Stekel DJ, et al. (2008) Protein database searches using compositionally adjusted substitution matrices. FEBS J 272: 5101–5109.
35. 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.
36. Welch RA, Borland V, Phanakit 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.
37. Lu G, Guo GQ, Zhu H (2009) Genome reannotation of *Escherichia coli* CFT073 with new insights into virulence. BMC Genomics 10: 552.
38. Schutl SE, Woostec JG, Gerz EM, Agarwala R, Mouglistu A, et al. (2005) Protein database searches using compositionally adjusted substitution matrices. FEBS J 272: 5101–5109.
58. Parkkinen J, Rogers GN, Korhonen T, Dahr W, Finne J (1986) Identification of
57. Backhed F, Alsen B, Roche N, Angstrom J, von Euler A, et al. (2002)
54. Pouttu R, Westerlund-Wikstrom B, Lang H, Alsti K, Virkola R, et al. (2001)
52. Soto GE, Hultgren SJ (1999) Bacterial adhesins: common themes and variations
51. Kuo CH, Ochman H (2010) The extinction dynamics of bacterial pseudogenes.
50. McClelland M, Sanderson KE, Spieth J, Clifton SW, Latreille P, et al. (2001)
49. Wirth T, Falush D, Falush D, Lan R, Colles F, Mensa P, et al. (2006) Sex and virulence in
70. Hochhut B, Wilde C, Balling G, Middendorf B, Dobrindt U, et al. (2006) Role of
68. Petty NK, Bulgin R, Crepin VF, Cerdeno-Tarraga AM, Schroeder GN, et al.
65. Ochman H, Wilson AC (1987) Evolution in bacteria: evidence for a universal
61. Le Bouguenec C, Garcia MI, Ouin V, Desperrier JM, Gounon P, et al. (1993)
60. Sakellaris H, Balding DP, Scott JR (1996) Assembly proteins of CS1 pili of
43. Westerman LR, Hepper P, Collin RS, Homewood JA, Romanowski J, et al. (2010) The
42. Ochman H, Wilson AC (1987) Evolution in bacteria: evidence for a universal
41. Leitner T, Cargill M, Tishkoff SA, Eichler EE, Niswander KA, et al. (2002) Evolutionary
38. Perna NT, Plunkett G 3rd, Burland V, Mau B, Glasner JD, et al. (2001) Genome
37. Crossman LC, Chaudhuri RR, Beattie SA, Wills TJ, Daveson M, et al. (2010) A commonly used gene b panel complete genome sequence of the prototypical enterotraggregative Escherichia coli strain O127:H6. J Bacteriol 191: 347–354.
36. Zhou Z, Li X, Lai B, Beutin L, Xu J, et al. (2010) Derivation of Escherichia coli O157:H7 from its O53:H7 precursor. PLoS One 5: e6700.
35. Parkkinen J, Rogers GN, Korhonen T, Dahr W, Finne J (2002) Identification of
34. Pouttu R, Westerlund-Wikstrom B, Lang H, Alsti K, Virkola R, et al. (2001)
33. Backhed F, Westerlund-Wikstrom B, Lang H, Alsti K, Virkola R, et al. (2001)
32. Ochman H, Wilson AC (1987) Evolution in bacteria: evidence for a universal
31. Leitner T, Cargill M, Tishkoff SA, Eichler EE, Niswander KA, et al. (2002) Evolutionary
30. Perna NT, Plunkett G 3rd, Burland V, Mau B, Glasner JD, et al. (2001) Genome
29. Crossman LC, Chaudhuri RR, Beattie SA, Wills TJ, Daveson M, et al. (2010) A commonly used gene b panel complete genome sequence of the prototypical enterotraggregative Escherichia coli strain O127:H6. J Bacteriol 191: 347–354.
28. Zhou Z, Li X, Lai B, Beutin L, Xu J, et al. (2010) Derivation of Escherichia coli O157:H7 from its O53:H7 precursor. PLoS One 5: e6700.
27. Parkkinen J, Rogers GN, Korhonen T, Dahr W, Finne J (2002) Identification of
26. Petty NK, Bulgin R, Crepin VF, Cerdeno-Tarraga AM, Schroeder GN, et al.
25. Ochman H, Wilson AC (1987) Evolution in bacteria: evidence for a universal
24. Leitner T, Cargill M, Tishkoff SA, Eichler EE, Niswander KA, et al. (2002) Evolutionary
23. Westerman LR, Hepper P, Collin RS, Homewood JA, Romanowski J, et al. (2010) The
22. Ochman H, Wilson AC (1987) Evolution in bacteria: evidence for a universal
21. Yue M, Rankin SG, Blanchet RT, Nulton JD, Edwards RA, et al. (2012) Diversification of the salmonella fimbiae: a model of macro- and microevolution. PLoS One 7: e30936.
20. Townsend SM, Kramer NE, Edwards R, Baker S, Hamlin N, et al. (2001) Salmonella enterica serovar Typhi possesses a unique repertoire of fimbral gene sequences. Infect Immun 69: 2894–2901.
19. Petty NK, Bulgin R, Crepin VF, Cerdeno-Tarraga AM, Schroeder GN, et al. (2010) The Citrobacter rodentium rodentium genome sequence reveals convergent evolution with human pathogenic Escherichia coli. J Bacteriol 192: 525–538.
18. Alikhan NF, Petty NK, Ben Zakour NL, Beattson SA (2011) BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics 12: 402.
17. Hochhut B, Wilde C, Balling R, 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: 584–593.
16. Rasko DA, Rosovitz MJ, Myers GS, Mongodin EF, Fricker WF, et al. (2008) The pangenome structure of Escherichia coli: comparative genomic analysis of E. coli communal and pathogenic isolates. J Bacteriol 190: 6881–6893.
15. Chen SL, Hung CS, Xu J, Regestad CK, Magni 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.
14. Tatsuka M, Beattson SA, Sarkar S, Phan MD, Petty NK, et al. (2011) Insights into a multidrug resistant Escherichia coli pathogen of the globally disseminated UTI1 lineage: genome analysis and virulence mechanisms. PLoS One 6: e26578.
13. Zdziarski J, Bruzko skiewicz E, Walli B, Liesegang B, Biran D, et al. (2010) Host imprints on bacterial genomes: rapid, divergent evolution in individual patients. PLoS Pathog 6: e1000778.
12. Johnson TJ, Johnson SJ, Nolan LK (2006) Complete DNA sequence of a ColB plasmid from avian pathogenic Escherichia coli suggests that it evolved from closely related ColV plasmids. J Bacteriol 188: 5975–5983.
11. Chaudhuri RR, Sehalkia M, Hobman JL, Weber MA, Leyton DL, et al. (2010) Complete genome sequence and comparative metabolic profiling of the prototypical enterotraggregative Escherichia coli strain H42. PLoS One 5: e8601.
10. Iguchi A, Thomson NR, Ogura Y, Saa nders D, Oska t, et al. (2009) Complete genome sequence and comparative genome analysis of enteropathogenic Escherichia coli O127:H6 strain E2348/69. J Bacteriol 191: 345–354.
9. Zhou Z, Li X, Lai B, Beutin L, Xu J, et al. (2010) Derivation of Escherichia coli O157:H7 from its O53:H7 precursor. PLoS One 5: e6700.
8. Grossman LC, Chaudhuri RR, Beattson SA, Wills TJ, Daveson M, et al. (2010) A commonly used gene b panel complete genome sequence of the prototypical enterotraggregative Escherichia coli strain H104: J Bacteriol 192: 5222–5231.
7. Ogura Y, Oska t, Iguchi A, Toh H, Anandhhgani M, et al. (2009) Comparative genomes 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.
6. Verna NT, Plunkett G 3rd, Burland V, Maa B, Blasen JD, et al. (2001) Genome sequence of enterohemorrhagic Escherichia coli O157:H7. Nature 409: 529–533.
5. Hayashi T, Makino K, Ohi mishi 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.
4. Epping M, Mammel MK, Leeclerc JE, Ravel J, Cebula TA (2011) Genomic anatomy of Escherichia coli O157:H7 outbreaks. Proc Natl Acad Sci U S A 108: 20142–20147.
3. Kula sekara 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.
2. Grigoriev IV, Nordberg H, Shahabov I, Aerts A, Cantor M, et al. (2012) The genome portal of the Department of Energy Joint Genome Institute. Nucleic Acids Res 40: 926–932.
1. 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: 373–384.