

RESEARCH ARTICLE

Genome Based Phylogeny and Comparative Genomic Analysis of Intra-Mammary Pathogenic *Escherichia coli*

Vincent P. Richards^{1#a}, Tristan Lefébure^{1#b}, Paulina D. Pavinski Bitar¹, Belgin Dogan², Kenneth W. Simpson², Ynte H. Schukken¹, Michael J. Stanhope^{1*}

1 Department of Population Medicine and Diagnostic Sciences, College of Veterinary Medicine, Cornell University, Ithaca, New York, United States of America, **2** Department of Clinical Sciences, College of Veterinary Medicine, Cornell University, Ithaca, New York, United States of America

^{#a} Current address: Department of Biological Sciences, College of Agriculture, Forestry and Life Sciences, Clemson University, Clemson, South Carolina, United States of America

^{#b} Current address: Université de Lyon; Centre national de la recherche scientifique, Laboratoire d'Ecologie des Hydrosystèmes Naturels et Anthropisés; Université Lyon 1, Villeurbanne, France

* mjs297@cornell.edu



OPEN ACCESS

Citation: Richards VP, Lefébure T, Pavinski Bitar PD, Dogan B, Simpson KW, Schukken YH, et al. (2015) Genome Based Phylogeny and Comparative Genomic Analysis of Intra-Mammary Pathogenic *Escherichia coli*. PLoS ONE 10(3): e0119799. doi:10.1371/journal.pone.0119799

Academic Editor: Gabriel Moreno-Hagelsieb, Wilfrid Laurier University, CANADA

Received: October 8, 2014

Accepted: January 22, 2015

Published: March 25, 2015

Copyright: © 2015 Richards et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Whole genome shotgun sequences are deposited at GenBank under accession numbers AHHK00000000, NZ_AHHL00000000, NZ_AHHM00000000, and AHHN00000000.

Funding: This work was financially supported by The New York State Foundation for Science, Technology and Innovation (grant C040124) <http://esd.ny.gov/nystar/>. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Escherichia coli is an important cause of bovine mastitis and can cause both severe inflammation with a short-term transient infection, as well as less severe, but more chronic inflammation and infection persistence. *E. coli* is a highly diverse organism that has been classified into a number of different pathotypes or pathovars, and mammary pathogenic *E. coli* (MPEC) has been proposed as a new such pathotype. The purpose of this study was to use genome sequence data derived from both transient and persistent MPEC isolates (two isolates of each phenotype) to construct a genome-based phylogeny that places MPEC in its phylogenetic context with other *E. coli* pathovars. A subsidiary goal was to conduct comparative genomic analyses of these MPEC isolates with other *E. coli* pathovars to provide a preliminary perspective on loci that might be correlated with the MPEC phenotype. Both concatenated and consensus tree phylogenies did not support MPEC monophyly or the monophyly of either transient or persistent phenotypes. Three of the MPEC isolates (ECA-727, ECC-Z, and ECA-O157) originated from within the predominately commensal clade of *E. coli*, referred to as phylogroup A. The fourth MPEC isolate, of the persistent phenotype (ECC-1470), was sister group to an isolate of ETEC, falling within the *E. coli* B1 clade. This suggests that the MPEC phenotype has arisen on numerous independent occasions and that this has often, although not invariably, occurred from commensal ancestry. Examination of the genes present in the MPEC strains relative to the commensal strains identified a consistent presence of the type VI secretion system (T6SS) in the MPEC strains, with only occasional representation in commensal strains, suggesting that T6SS may be associated with MPEC pathogenesis and/or as an inter-bacterial competitive attribute and therefore could represent a useful target to explore for the development of MPEC specific inhibitors.

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

Introduction

Dairy cattle mastitis, or inflammation of the mammary gland, is the dominant bovine health disorder affecting milk production worldwide, with reported annual losses exceeding two billion dollars in the United States dairy industry [1–3]. A variety of both gram positive and negative species of bacteria, including *Escherichia coli* [4], are linked to bovine mastitis. *E. coli* can cause both severe inflammation with a short-term transient infection, as well as less severe but more chronic inflammation and infection persistence [5, 6]. *E. coli* is a highly diverse organism that has been classified into a number of different pathotypes or pathovars, including adherent-invasive *E. coli* (AIEC), avian pathogenic *E. coli* (APEC), enteroaggregative *E. coli* (EAEC), enterohemorrhagic *E. coli* (EHEC), enteropathogenic *E. coli* (EPEC), enterotoxigenic *E. coli* (ETEC), extra-intestinal pathogenic *E. coli* (ExPEC), and uropathogenic *E. coli* (UPEC). Mammary pathogenic *E. coli* (MPEC) has been proposed as a new *E. coli* pathotype [4]. Distinct subspecies or phylogroups of *E. coli* have long been acknowledged [7–10] and strains of the different phylogroups appear to differ in their ecological niches, and disease characteristics. At present, there is no published *E. coli* phylogeny that includes MPEC isolates in a comparative dataset that encompasses representative isolates of the various *E. coli* pathovars. The only published sequence based phylogeny regarding MPEC that we are aware of, uses MLST data and reports a single most parsimonious tree of a set of 165 MP trees and includes only one non-MPEC isolate in the analysis [11]. There are studies involving MPEC isolates that employ a PCR based test [12] to assess *E. coli* phylogroup membership, however, this is not a phylogenetic analysis, and provides no information on relative relationships of the isolates within each phylogroup. Furthermore, this PCR based test has been shown to provide unreliable designations for certain phylogroups [13], and in particular, that group (phylogroup A) to which MPEC isolates are most typically assigned.

We have developed a collection of *E. coli* isolates that can be divided into two distinct groups: those isolated from intramammary infections (IMI) displaying transient infection characteristics (see for examples [14, 15]) and those isolated from IMI associated with persistently infected mammary glands (see for examples [16]). The purpose of this study is to use genome sequence data derived from both transient and persistent MPEC isolates to construct a genome-based phylogeny that places MPEC in its phylogenetic context with other *E. coli* pathovars, and to evaluate whether there is possibly a different evolutionary history underlying transient and persistent phenotypes. A subsidiary goal is to conduct comparative genomic analyses of these MPEC isolates with other *E. coli* pathovars to provide a preliminary perspective on loci that might be correlated with the MPEC phenotype.

Materials and Methods

Genome sequencing and annotation of MPEC strains

Two *E. coli* strains associated with transient IMI, and two strains associated with persistent IMI were evaluated. The strains from transient infections were ECA-727 [17] and ECA-O157 [18]. The strains ECA-727 and ECA-O157 have been repeatedly shown to cause only brief transient IMI following an intramammary challenge. The *E. coli* isolates associated with persistent IMI, strain ECC-Z [19] and strain ECC-1470 [5], were isolated from persistently infected cows. Presence of a persistent IMI was confirmed by pulsed-field gel electrophoresis (PFGE) typing of isolates collected over time from the same infected mammary quarter. All four strains have also been used in phenotypic studies of infection characteristics in mammary epithelial cells [5, 6].

We determined the genome sequences of the four MPEC strains isolated from transient (ECA-O157 and ECA-727) and persistent (ECC-1470 and ECC-Z) bovine mastitis. Roche/454

pyrosequencing, involving single or paired-end reads from the FLX sequencer, was used to determine the sequence of the four MPEC genomes. The genome sequence for ECA-727 did not involve paired end data. The sequences were assembled (*De novo* assembly with Newbler Software) into 1173, 539, 994, and 24 contigs for ECA-0157, ECA-727, ECC-1470, and ECC-Z, respectively. Genome annotation for the four MPEC strains was done by the National Center for Biotechnology Information (NCBI) Prokaryotic Genomes Automatic Annotation Pipeline (PGAAP) at <http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html>. The whole genome shotgun projects for ECA-0157, ECA-727, ECC-1470, and ECC-Z have been deposited at DDBJ/EMBL/GenBank under the accession numbers AHHK00000000, NZ_AHHL00000000, NZ_AHHM00000000, and AHHN00000000, respectively. We compared these four MPEC genome sequences to 56 complete *E. coli* genome sequences representing a combination of commensal isolates and the following ten pathotypes or pathovars: ABU (Asymptomatic Bacteriuria), AIEC, APEC, EAEC, EHEC, EPEC, ETEC, ExPEC, NMEC (Neonatal Meningitis), and UPEC. Also included were nine *Shigella* sp. isolates representing four different species. All of these 65 sequences were obtained from NCBI; their strain names, accession numbers, genome lengths and respective pathotypes are presented in [S1 Table](#) in the supporting information.

Gene clustering and phylogenetic analyses

Homologous genes among the 69 genomes analyzed were delineated using the MCL algorithm [20] as implemented in the MCLBLASTLINE pipeline (available at <http://micans.org/mcl>). The pipeline uses Markov clustering (MCL) to assign genes to homologous clusters based on a BLASTp search between all pairs of protein sequences using an *E* value cut-off of 1e-5. The MCL algorithm was implemented using an inflation parameter of 1.8. Simulations have shown this value to be generally robust to false positives and negatives [21]. Nucleotide sequences corresponding to each MCL gene cluster were aligned using Probalign v1.1 [22]. For the phylogenetic analyses we selected those MCL gene clusters that were shared among all genomes and contained only single gene copies for each (the core set) (984 clusters). Recombination for genes in these clusters was assessed using a combination of three methods: (i) the Pairwise Homoplasy Index (PHI), (ii) the Neighbor Similarity Score (NSS), and (iii) Maximum χ^2 . We excluded any cluster that showed evidence of recombination for any two of the three methods. Using this approach, we excluded 546 clusters (438 remained).

We utilized two approaches to phylogenetic reconstruction. The first was a gene tree consensus approach. Here, phylogenetic trees for each MCL gene cluster alignment were obtained using maximum likelihood (PhyML v.3.0), and a consensus of these gene trees obtained using the Triple Construction Method as implemented in the program Triplec [23]. This procedure is based on the observation that the most probable three-taxon tree consistently matches the species tree [24]. The method searches all input trees for the most frequent of the three possible rooted triplets for each set of three taxa. Once found, the set of rooted triples are joined to form the consensus tree using the quartet puzzling heuristic [25]. The method has been shown to outperform majority-rule and greedy consensus methods [26]. The second approach was to generate an alignment of single nucleotide polymorphism (SNP) genotypes for each isolate by concatenating the aligned non-recombinant core gene clusters (16,062 SNPs). Phylogenetic analysis was again performed using maximum likelihood (PhyML v.3.0) [27]. Analyses were performed using the GTR+G substitution model, which was determined to be the best fit for the data using the Akaike Information Criterion (AIC) in MODELTEST [28]. Node support was obtained using non-parametric bootstrapping (500 replicates).

Table 1. Basic genome statistics for the four MPEC strains.

Strain	Genome size (bp)	CDS	% GC	tRNAs
ECA-O157	4,654,138	5,179	50.7	47
ECA-727	4,859,863	4,901	50.9	50
ECC-1470	4,739,402	5,142	50.7	56
ECC-Z	4,911,586	4,679	53.0	69

doi:10.1371/journal.pone.0119799.t001

Gene Ontology (GO) terms were assigned to all genomes using Blast2GO v.2.5.0 [29]. Relative enrichment (over representation) of GO terms among lineages was assessed using Fisher exact tests. The test was performed using the Gossip statistical package [30] implemented within Blast2GO. The false discovery rate (FDR) procedure of Benjamini and Hochberg [31] was used to correct for multiple hypothesis testing ($FDR = 0.05$).

Results and Discussion

Basic genome statistics for the four sequenced MPEC strains are summarized in [Table 1](#).

Escherichia coli strains have been classified into an array of different phylogenetic groups, often referred to as phylogroups or phylotypes (A, B1, B2, C, D, E, F) [7–10], as well as a number of pathotypes (or pathovars), reflecting their different features of pathogenesis [32]. The Clermont genotyping scheme [12] has been used in several studies to assess the phylogroups of MPEC isolates and typically a majority, or at least a significant proportion, fall within phylogroup A, which includes predominately commensal isolates. For example, in a study from Iran about 45% of the isolates were classified as phylogroup A [33] and in another study from Finland about 83% were phylogroup A [34]. Recently, a study from China found 36% of clinical and subclinical bovine mastitis isolates were from group A, 58% group B1, and 6% group D [35]. Gordon et al. [13] compared triplex phylogroup determinations with MLST genotype data for a large number of isolates and determined that although overall the Clermont test correctly identified phylogroups 80–85% of the time, this differed depending on the particular phylogroup in question. Specifically, phylogroup A was seldom assigned correctly. This appears to be because the Clermont genotype of phylogroup A is the absence of all three PCR products, and because of potential mismatch between primer and template, absence of a PCR product in any diagnostic test is typically more unreliable than its presence. Gordon et al. suggested that strains failing to yield any PCR product are seldom members of phylogroup A and strains with such a genotype should not be assigned to a phylogroup. Therefore, Clermont determinations of MPEC as typically phylogroup A may not have been accurate.

In the present study, we used a set of 438 orthologous loci to infer the evolutionary history of 60 *E. coli* strains, and nine strains of *Shigella*. This involved data for 56 complete genomes of *E. coli* taken from GenBank representing all the major phylotypes (with the exception of group C for which there were no complete genome sequences at the time this study was conducted) and pathovars, plus draft genome sequences of our four MPEC isolates. The comparative genome sequence data were first screened for putative recombinant loci, which were excluded in the phylogenetic reconstructions and therefore these orthologous loci should represent our best source of data to estimate the true evolutionary history of these set of isolates. We reconstructed the phylogeny using both a concatenated dataset that included all 438 of these non-recombinant orthologs (16,062 SNPs) ([Fig. 1](#)), and a consensus of the set of gene trees ([Fig. 2](#)). Both resulting trees were highly congruent and in agreement with earlier *E. coli* phylogenies (e.g. [36]), provided convincing support for the monophyly of the various phylogroups. Both the concatenated and the consensus trees did not support MPEC monophyly or the monophyly

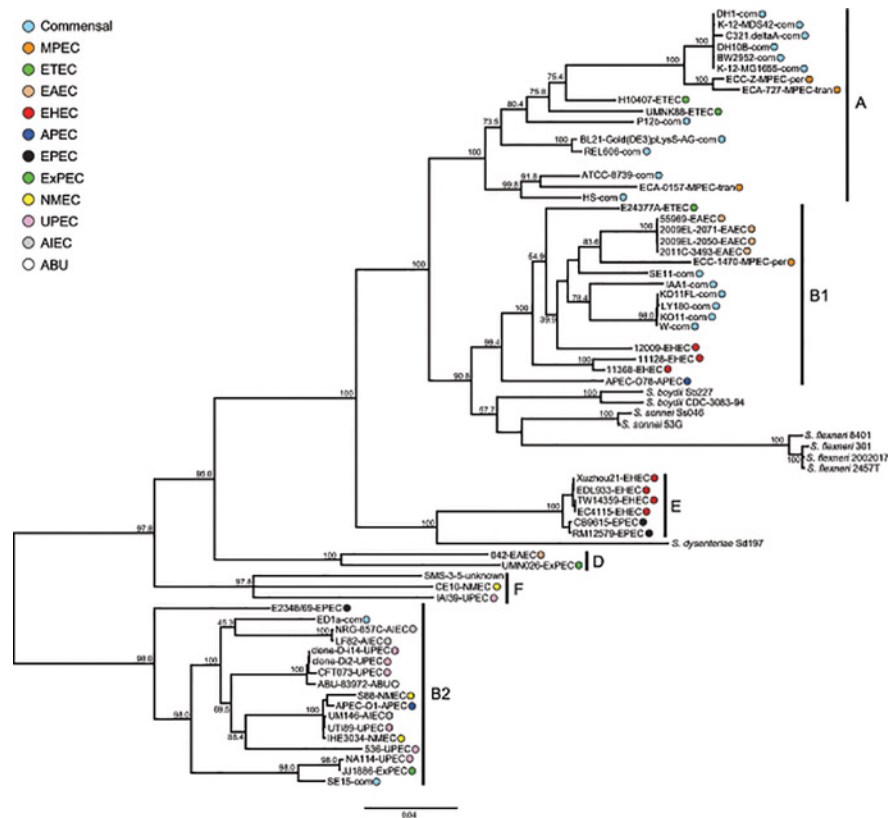


Fig 1. Maximum likelihood phylogeny generated from 16,062 SNPs. Bootstrap support values are shown over branches. Branch lengths are drawn proportion to the amount of sequence change. Phylogroup membership is indicated with a horizontal bar and the corresponding letter. Pathotypes of each of the *E. coli* strains is indicated with a color code and their lettered abbreviation adjacent to the strain names of each; commensal strains are abbreviated as “com”.

doi:10.1371/journal.pone.0119799.g001

of either transient or persistent phenotypes. Three of these isolates (ECA-727, ECC-Z, and ECA-0157) originated from within the predominately commensal clade, referred to as phylogroup A. This classification is in accordance with our previous work, based on triplex PCR typing [5]. In the consensus phylogeny, ECC-Z, and ECA-727 were imbedded within a clade of six other exclusively commensal isolates. In the concatenated phylogeny these two MPEC isolates are a sister group to a clade of these same six commensal strains, which in turn are highly similar, as evidenced by the lack of branch length in Fig. 1. The branch length separating this monophyletic group of MPEC/commensal strains is the longest in phylogroup A, suggesting this MPEC/commensal group is quite distinct from the rest of the phylogroup. The phylogenetic position of ECA-O157 also suggests its origins are from commensal strains. The fourth MPEC isolate, of the persistent phenotype (ECC-1470), was sister group to an isolate of ETEC (E24377A) in the consensus phylogeny, falling within the B1 clade. This isolate was previously identified as belonging to phylogroup A [5], however, this classification was based on the absence of bands in a triplex PCR, which is more likely to be erroneous. In addition, none of the 438 individual ML trees that comprise the ultimate consensus tree supported the monophyly of MPEC. This then strongly suggests that the MPEC phenotype has arisen on at least three independent occasions and furthermore, that at least some of the time, this occurred from commensal ancestry. The fact that MPEC strains have a history that includes independent origin from commensal ancestry suggests that these isolates have acquired the loci underlying their

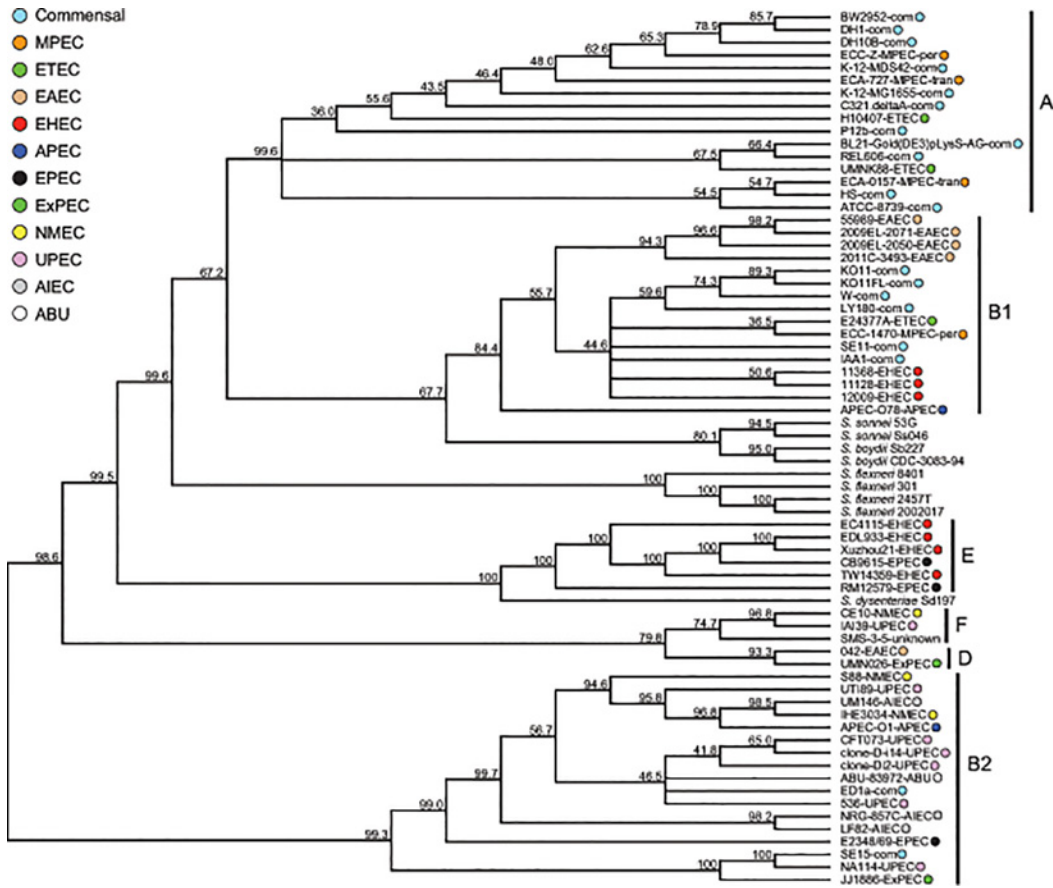


Fig 2. Consensus of 438 maximum likelihood gene phylogenies. Proportion of genes supporting groups shown on branches. Phylogroup and pathotype information same as in Fig. 1.

doi:10.1371/journal.pone.0119799.g002

pathogenic properties via lateral gene transfer (LGT). A comparison of the gene content of our two higher quality MPEC genome sequences (ECC-Z and ECA-727) relative to the 11 commensal strains that comprise the phylogroup A clade of our dataset, indicated that the MPEC strains are enriched, or over-represented, for the Genome Ontology (GO) categories of conjugation, and viral capsid assembly (FDR corrected P values of 3.26E-03 and 8.29E-04, respectively). They were under-represented for the categories of transposase activity, DNA mediated transposition, transposition, and DNA recombination (FDR corrected P values of 5.31E-19, 6.80E-18, 6.06E-13, 1.14E-06, respectively).

Earlier studies involving PCR surveys or DNA microarrays of *E. coli* virulence genes, and including many more MPEC isolates than are included here, have indicated that the presence and combinations of *E. coli* virulence loci in MPEC strains varies greatly (e.g. [11, 33, 34, 37]). Dogan et al. [37] showed that both transient and persistent MPEC isolates contained genes encoding type II, IV and VI secretion systems, long polar fimbriae (lpfA) and iron acquisition as potential pathogenicity loci. Since our isolate sample is limited, combined with the relatively low sequence coverage for two of our strains, we suggest there is little to be gained here by cataloging the presence and absence of virulence loci. Of somewhat greater possible interest, since our study does represent one of the first comparative genomic datasets involving MPEC strains, are the genes that are uniquely present in all four MPEC strains, or at least three of the four MPEC strains, and/or loci that are uniquely present in each of the transient or persistent

strains (presented in [S2](#) and [S3 Table](#) respectively). There was only one hypothetical protein unique to all four MPEC strains. Loci uniquely present in at least 3 MPEC strains included glutamate decarboxylase, multidrug efflux system protein MdtE, RhsC core protein with extension, a sensor protein, phosphoenolpyruvate synthase, an oxidoreductase, transcription-repair coupling factor, putative hydrolase, carboxymethylenebutenolidase and a few hypothetical proteins ([S2 Table](#)). The genes uniquely shared in the two transient MPEC strains included a set of 36 loci, whereas the genes uniquely present in the two persistent strains was only 11, and in this latter case, nearly all of which were hypothetical proteins ([S3 Table](#)). This transient and persistent comparison, however, should be regarded with a degree of skepticism because of the small sample size representing the two phenotypes and because of the relatively lower sequence coverage for two of the isolates (one persistent and one transient), which will in particular, make the assessment of absent loci somewhat tenuous.

Examination of the genes present in the MPEC strains relative to the commensal strains identified a consistent presence of the type VI secretion system (T6SS) in the MPEC strains, with more occasional representation in commensal strains (36.8% of strains). The type VI secretion system is the most recently described [[38](#), [39](#)] of the gram-negative bacterial secretion systems and is present in a diverse array of species. T6SSs inject effector proteins into both eukaryotic and prokaryotic target cells using a bacteriophage-like cell-puncturing device [[40](#)]. Targeting eukaryotic cells is part of the infection process and targeting prokaryotic cells is part of inter-bacterial antagonism and competition for the colonization of their environment. The effector proteins of these secretion systems are often critical for virulence; for example, the loss of the T3SS in *Yersinia pestis* is sufficient to render the bacteria completely avirulent [[41](#)].

The T6SS has been reported in commensal *E. coli*, including for example, strains W and HS included in our analyses [[42](#)]. It has also been identified in *E. coli* strains isolated from the ileum of patients with inflammatory bowel disease [[43](#)] including *E. coli* with an adherent-invasive (AIEC) pathotype. PCR based analysis of *E. coli* associated with mastitis, with similar adherent and invasive activity to AIEC, identified the T6SS *hcp* gene, a T6SS effector protein, in 88% of transient and 67% of persistent strains [[37](#)]. We conducted a BLASTn search including our MPEC isolates and the complete genome sequences included in our phylogeny, using the T6SS from strain W, which is referred to as the “enterohaemorrhagic” *E. coli* type six secretion system cluster [[42](#), [44](#)]. We found T6SS in all four of our MPEC strains, as well as 31 additional *E. coli* strains from the total of 69 strains included in [Figs. 1](#) and [2](#). T6SS was present in only one of the commensal strains from phylogroup A and this was strain HS. For reference, [S4 Table](#) shows the locus_tags and gene product descriptions for the T6SS from strain W and the MPEC strain ECC-Z. For MPEC strain ECA-727, the T6SS spanned three contigs, with the majority of genes (15) present on one contig. We performed a global alignment of the genes from this contig and the corresponding orthologs from each of the 31 additional *E. coli* and the MPEC strains. MPEC strain ECA-O157 was excluded because of fragmentation across multiple different contigs. Using this alignment, we performed a ML phylogenetic analysis using PhyML (see [methods](#)). The GTR+G substitution model was determined as the best fit for the data and node support was obtained using non-parametric bootstrapping (500 replicates). The resulting phylogeny is presented in [Fig. 3](#). The 5 additional commensal strains that had T6SS came from phylogroup B1. The fact that phylogroup A commensals in our data set do not tend to have T6SS, and that three of our four MPEC isolates originated from within phylogroup A, suggests that the MPEC isolates likely obtained T6SS via LGT. An examination of the loci up and downstream of the T6SS indicates the presence of various transposase and/or phage genes, in both our MPEC isolates as well as other isolates carrying T6SS. The T6SS phylogeny supports a clear dichotomy with two very different T6SS clades ([Fig. 3](#)) and these two groups tend to include isolates of distinct phylotypes and pathotypes. One of the T6SS groups includes

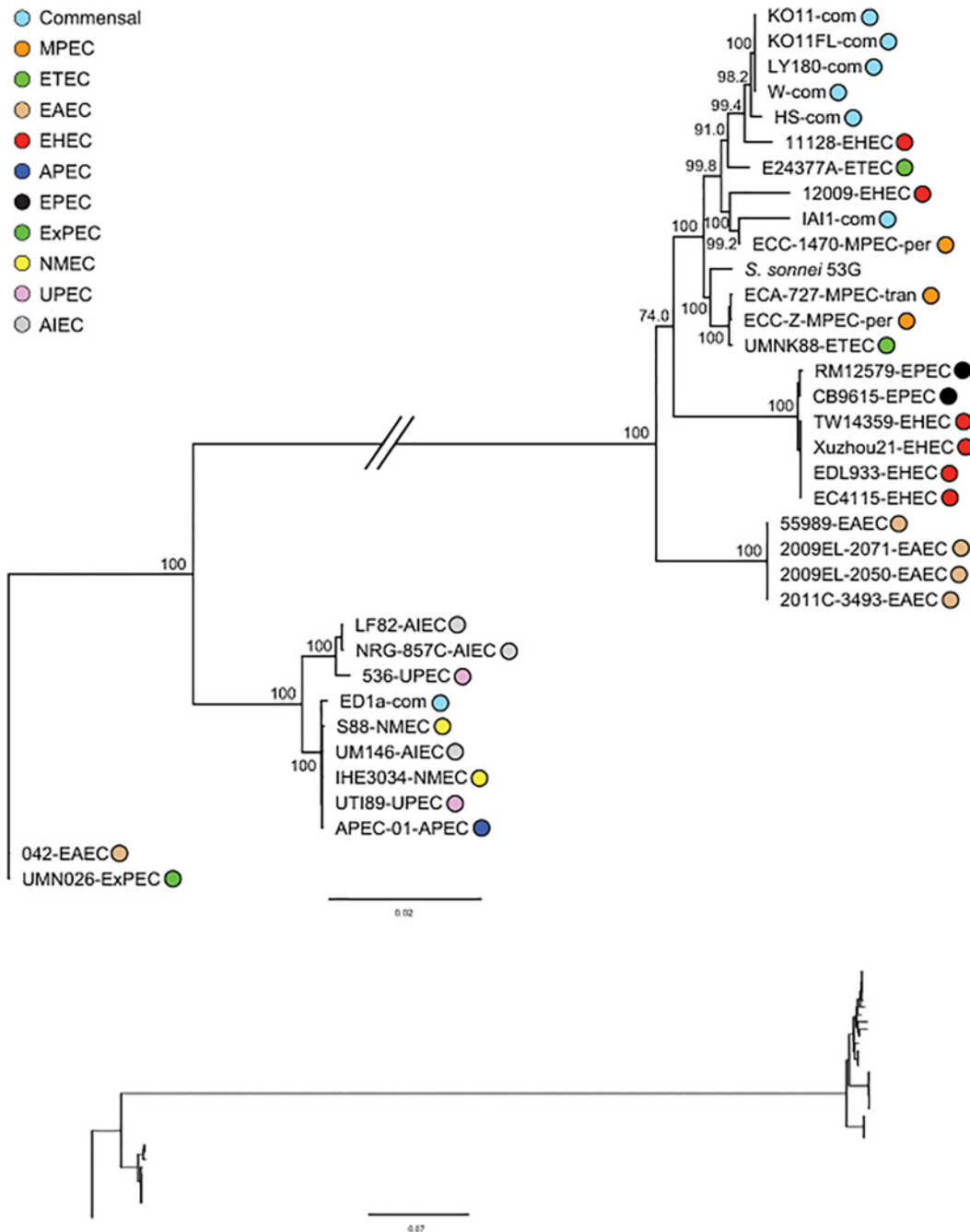


Fig 3. Maximum likelihood phylogeny for type six secretion system. Bootstrap support values are shown over branches. Branch lengths drawn proportion to the amount of sequence change. Phylogroup and pathotype information same as in Fig. 1.

doi:10.1371/journal.pone.0119799.g003

representatives from predominately phylogroup B2 and the other from largely phylogroup B1, with a few phylogroup A representatives. Two of our MPEC strains share a close sister group relationship of their T6SS with a porcine enterotoxigenic *E. coli* (UMNK88-ETEC). The T6SS sequence similarity is very high between the ETEC and these two MPEC strains (ECA-727; ECC-Z). For example, after exclusion of a 1,619bp indel within the IcmF-like protein for strain ECC-Z, identity for the 15 genes used in the phylogeny was 99.9%. This high sequence identity

occurred despite the fact these strains are not closely related in the strain phylogeny (Figs. 1 and 2), suggesting a possible LGT event involving the MPEC strains and this porcine ETEC or one closely related to it; these two MPEC strains include one persistent and one transient, with the T6SS for the other persistent strain (ECC-1470) not particularly closely related to either of these. Examination of the loci up and downstream of the T6SS in ECC-Z (isolate with the longest contig for that region), indicate a close synteny between the MPEC and ETEC strains, including transposase. This evolutionary picture then, suggests a T6SS LGT event involving an ETEC strain, possibly UMNK88, or a closely related strain to UMNK88. Furthermore, these data suggest that T6SS may well be associated with MPEC strains, either as a specific feature of pathogenesis and/or as an inter-bacterial competitive attribute, but possibly not specifically related to differences between transient and persistent phenotypes. Further research on the impact of T6SS on pathogenesis of bovine mastitis in challenge models comparing T6SS mutants and wild types appears indicated.

Supporting Information

S1 Table. Strain names, accession numbers, genome sequence length and pathotype for each isolate included in this study, as well as sequencing statistics for the four MPEC strains.

(XLSX)

S2 Table. Genes that are uniquely present in at least three of the four MPEC strains in comparisons to the other genomes in our dataset.

(XLSX)

S3 Table. Genes that are uniquely present in the transient and persistent MPEC strains in comparisons to the other genomes in our dataset.

(XLSX)

S4 Table. Locus tags, gene product descriptions, and gene sequence identities for the T6SS from strain W and the MPEC strain ECC-Z.

(XLSX)

Author Contributions

Conceived and designed the experiments: KWS YHS MJS. Performed the experiments: PDPB BD. Analyzed the data: VPR TL. Wrote the paper: VPR MJS. Manuscript Revision: TL PDPB BD KWS YHS.

References

1. Keefe GP. Streptococcus agalactiae mastitis: a review. *Can Vet J.* 1997; 38(7):429–37. PMID: [9220132](#)
2. Wilson DJ, Gonzalez RN, Case KL, Garrison LL, Grohn YT. Comparison of seven antibiotic treatments with no treatment for bacteriological efficacy against bovine mastitis pathogens. *J Dairy Sci.* 1999; 82(8):1664–70. PMID: [10480091](#)
3. Katholm J, Rattenborg E, editors. The surveillance program of Streptococcus agalactiae in Danish dairy herds 1989–2008. Mastitis Research Into Practice: Proceedings of the 5th IDF Mastitis Conference, 21–24 March, 2010, Christchurch New Zealand; 2010: Dunmore Pub.
4. Shpigel NY, Elazar S, Rosenshine I. Mammary pathogenic *Escherichia coli*. *Curr Opin Microbiol.* 2008; 11(1):60–5. doi: [10.1016/j.mib.2008.01.004](#) PMID: [18291708](#)
5. Dogan B, Klaessig S, Rishniw M, Almeida RA, Oliver SP, Simpson K, et al. Adherent and invasive *Escherichia coli* are associated with persistent bovine mastitis. *Vet Microbiol.* 2006; 116(4):270–82. PMID: [16787715](#)

6. Almeida RA, Dogan B, Klaessing S, Schukken YH, Oliver SP. Intracellular fate of strains of *Escherichia coli* isolated from dairy cows with acute or chronic mastitis. *Vet Res Commun*. 2011; 35(2):89–101. doi: [10.1007/s11259-010-9455-5](https://doi.org/10.1007/s11259-010-9455-5) PMID: [21207146](https://pubmed.ncbi.nlm.nih.gov/21207146/)
7. Ochman H, Selander RK. Standard reference strains of *Escherichia coli* from natural populations. *J Bacteriol*. 1984; 157(2):690–3. PMID: [6363394](https://pubmed.ncbi.nlm.nih.gov/6363394/)
8. Herzer PJ, Inouye S, Inouye M, Whittam TS. Phylogenetic distribution of branched RNA-linked multicopy single-stranded DNA among natural isolates of *Escherichia coli*. *J Bacteriol*. 1990; 172(11):6175–81. PMID: [1699928](https://pubmed.ncbi.nlm.nih.gov/1699928/)
9. Desjardins P, Picard B, Kaltenbock B, Elion J, Denamur E. Sex in *Escherichia coli* does not disrupt the clonal structure of the population: evidence from random amplified polymorphic DNA and restriction-fragment-length polymorphism. *J Mol Evol*. 1995; 41(4):440–8. PMID: [7563131](https://pubmed.ncbi.nlm.nih.gov/7563131/)
10. Wirth T, Falush D, Lan R, Colles F, Mensa P, Wieler LH, et al. Sex and virulence in *Escherichia coli*: an evolutionary perspective. *Mol Microbiol*. 2006; 60(5):1136–51. PMID: [16689791](https://pubmed.ncbi.nlm.nih.gov/16689791/)
11. Blum SE, Leitner G. Genotyping and virulence factors assessment of bovine mastitis *Escherichia coli*. *Vet Microbiol*. 2013; 163(3–4):305–12. doi: [10.1016/j.vetmic.2012.12.035](https://doi.org/10.1016/j.vetmic.2012.12.035) PMID: [23465470](https://pubmed.ncbi.nlm.nih.gov/23465470/)
12. Clermont O, Bonacorsi S, Bingen E. Rapid and simple determination of the *Escherichia coli* phylogenetic group. *Appl Environ Microbiol*. 2000; 66(10):4555–8. PMID: [11010916](https://pubmed.ncbi.nlm.nih.gov/11010916/)
13. Gordon DM, Clermont O, Tolley H, Denamur E. Assigning *Escherichia coli* strains to phylogenetic groups: multi-locus sequence typing versus the PCR triplex method. *Environmental microbiology*. 2008; 10(10):2484–96. doi: [10.1111/j.1462-2920.2008.01669.x](https://doi.org/10.1111/j.1462-2920.2008.01669.x) PMID: [18518895](https://pubmed.ncbi.nlm.nih.gov/18518895/)
14. Wilson DJ, Mallard BA, Burton JL, Schukken YH, Grohn YT. Milk and serum J5-specific antibody responses, milk production change, and clinical effects following intramammary *Escherichia coli* challenge for J5 vaccinate and control cows. *Clin Vaccine Immunol*. 2007; 14(6):693–9. PMID: [17460115](https://pubmed.ncbi.nlm.nih.gov/17460115/)
15. Hogan JS, Cannon VB, Smith KL, Rinehart C, Miller S. Effects of adjuvants on safety and efficacy of an *Escherichia coli* J5 bacterin. *J Dairy Sci*. 2005; 88(2):534–42. PMID: [15653518](https://pubmed.ncbi.nlm.nih.gov/15653518/)
16. Dopfer D, Barkema HW, Lam TJ, Schukken YH, Gaastra W. Recurrent clinical mastitis caused by *Escherichia coli* in dairy cows. *J Dairy Sci*. 1999; 82(1):80–5. PMID: [10022009](https://pubmed.ncbi.nlm.nih.gov/10022009/)
17. Hogan JS, Bogacz VL, Aslam M, Smith KL. Efficacy of an *Escherichia coli* J5 bacterin administered to primigravid heifers. *J Dairy Sci*. 1999; 82(5):939–43. PMID: [10342231](https://pubmed.ncbi.nlm.nih.gov/10342231/)
18. Lohuis JA, Schukken YH, Verheijden JH, Brand A, Van Miert AS. Effect of severity of systemic signs during the acute phase of experimentally induced *Escherichia coli* mastitis on milk production losses. *J Dairy Sci*. 1990; 73(2):333–41. PMID: [2184173](https://pubmed.ncbi.nlm.nih.gov/2184173/)
19. Dopfer D, Almeida RA, Lam TJ, Nederbragt H, Oliver SP, Gaastra W. Adhesion and invasion of *Escherichia coli* from single and recurrent clinical cases of bovine mastitis in vitro. *Vet Microbiol*. 2000; 74(4):331–43. PMID: [10831855](https://pubmed.ncbi.nlm.nih.gov/10831855/)
20. van Dongen S. Graph clustering by flow simulation [PhD thesis]: University of Utrecht; 2000.
21. Brohee S, van Helden J. Evaluation of clustering algorithms for protein-protein interaction networks. *BMC Bioinformatics*. 2006; 7:488. PMID: [17087821](https://pubmed.ncbi.nlm.nih.gov/17087821/)
22. Roshan U, Livesay DR. Probalgn: multiple sequence alignment using partition function posterior probabilities. *Bioinformatics*. 2006; 22(22):2715–21. PMID: [16954142](https://pubmed.ncbi.nlm.nih.gov/16954142/)
23. Ewing GB, Ebersberger I, Schmidt HA, von Haeseler A. Rooted triple consensus and anomalous gene trees. *BMC Evol Biol*. 2008; 8:118. doi: [10.1186/1471-2148-8-118](https://doi.org/10.1186/1471-2148-8-118) PMID: [18439266](https://pubmed.ncbi.nlm.nih.gov/18439266/)
24. Degnan JH, Rosenberg NA. Discordance of species trees with their most likely gene trees. *PLoS Genet*. 2006; 2(5):e68. PMID: [16733550](https://pubmed.ncbi.nlm.nih.gov/16733550/)
25. Strimmer K, vonHaeseler A. Quartet puzzling: A quartet maximum-likelihood method for reconstructing tree topologies. *Mol Biol Evol*. 1996; 13(7):964–9.
26. Degnan JH, DeGiorgio M, Bryant D, Rosenberg NA. Properties of consensus methods for inferring species trees from gene trees. *Syst Biol*. 2009; 58(1):35–54. doi: [10.1093/sysbio/syp008](https://doi.org/10.1093/sysbio/syp008) PMID: [20525567](https://pubmed.ncbi.nlm.nih.gov/20525567/)
27. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 2010; 59(3):307–21. doi: [10.1093/sysbio/syq010](https://doi.org/10.1093/sysbio/syq010) PMID: [20525638](https://pubmed.ncbi.nlm.nih.gov/20525638/)
28. Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. *Bioinformatics*. 1998; 14(9):817–8. PMID: [9918953](https://pubmed.ncbi.nlm.nih.gov/9918953/)
29. Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, et al. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res*. 2008; 36(10):3420–35. doi: [10.1093/nar/gkn176](https://doi.org/10.1093/nar/gkn176) PMID: [18445632](https://pubmed.ncbi.nlm.nih.gov/18445632/)
30. Blüthgen N, Brand K, Cajavec B, Swat M, Herzel H, Beule D. Biological profiling of gene groups utilizing Gene Ontology. *Genome Inform*. 2005; 16(1):106–15. PMID: [16362912](https://pubmed.ncbi.nlm.nih.gov/16362912/)

31. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B*. 1995; 57:289–300.
32. Croxen MA, Law RJ, Scholz R, Keeney KM, Wlodarska M, Finlay BB. Recent advances in understanding enteric pathogenic *Escherichia coli*. *Clin Microbiol Rev*. 2013; 26(4):822–80. doi: [10.1128/CMR.00022-13](https://doi.org/10.1128/CMR.00022-13) PMID: [24092857](https://pubmed.ncbi.nlm.nih.gov/24092857/)
33. Ghanbarpour R, Oswald E. Phylogenetic distribution of virulence genes in *Escherichia coli* isolated from bovine mastitis in Iran. *Res Vet Sci*. 2010; 88(1):6–10. doi: [10.1016/j.rvsc.2009.06.003](https://doi.org/10.1016/j.rvsc.2009.06.003) PMID: [19580980](https://pubmed.ncbi.nlm.nih.gov/19580980/)
34. Suojala L, Pohjanvirta T, Simojoki H, Myllyniemi AL, Pitkala A, Pelkonen S, et al. Phylogeny, virulence factors and antimicrobial susceptibility of *Escherichia coli* isolated in clinical bovine mastitis. *Vet Microbiol*. 2011; 147(3–4):383–8. doi: [10.1016/j.vetmic.2010.07.011](https://doi.org/10.1016/j.vetmic.2010.07.011) PMID: [20729012](https://pubmed.ncbi.nlm.nih.gov/20729012/)
35. Liu Y, Liu G, Liu W, Liu Y, Ali T, Chen W, et al. Phylogenetic group, virulence factors and antimicrobial resistance of *Escherichia coli* associated with bovine mastitis. *Res Microbiol*. 2014; 165(4):273–7. doi: [10.1016/j.resmic.2014.03.007](https://doi.org/10.1016/j.resmic.2014.03.007) PMID: [24705087](https://pubmed.ncbi.nlm.nih.gov/24705087/)
36. Kaas RS, Friis C, Ussery DW, Aarestrup FM. Estimating variation within the genes and inferring the phylogeny of 186 sequenced diverse *Escherichia coli* genomes. *BMC Genomics*. 2012; 13:577. doi: [10.1186/1471-2164-13-577](https://doi.org/10.1186/1471-2164-13-577) PMID: [23114024](https://pubmed.ncbi.nlm.nih.gov/23114024/)
37. Dogan B, Rishniw M, Bruant G, Harel J, Schukken YH, Simpson KW. Phylogroup and *lpfA* influence epithelial invasion by mastitis associated *Escherichia coli*. *Vet Microbiol*. 2012; 159(1–2):163–70.
38. Pukatzki S, Ma AT, Sturtevant D, Krastins B, Sarracino D, Nelson WC, et al. Identification of a conserved bacterial protein secretion system in *Vibrio cholerae* using the *Dictyostelium* host model system. *Proc Natl Acad Sci U S A*. 2006; 103(5):1528–33. PMID: [16432199](https://pubmed.ncbi.nlm.nih.gov/16432199/)
39. Mougous JD, Cuff ME, Raunser S, Shen A, Zhou M, Gifford CA, et al. A virulence locus of *Pseudomonas aeruginosa* encodes a protein secretion apparatus. *Science*. 2006; 312(5779):1526–30. PMID: [16763151](https://pubmed.ncbi.nlm.nih.gov/16763151/)
40. Russell AB, Peterson SB, Mougous JD. Type VI secretion system effectors: poisons with a purpose. *Nat Rev Microbiol*. 2014; 12(2):137–48. doi: [10.1038/nrmicro3185](https://doi.org/10.1038/nrmicro3185) PMID: [24384601](https://pubmed.ncbi.nlm.nih.gov/24384601/)
41. Ke Y, Chen Z, Yang R. *Yersinia pestis*: mechanisms of entry into and resistance to the host cell. *Frontiers in cellular and infection microbiology*. 2013; 3:106. doi: [10.3389/fcimb.2013.00106](https://doi.org/10.3389/fcimb.2013.00106) PMID: [24400226](https://pubmed.ncbi.nlm.nih.gov/24400226/)
42. Archer CT, Kim JF, Jeong H, Park JH, Vickers CE, Lee SY, et al. The genome sequence of *E. coli* W (ATCC 9637): comparative genome analysis and an improved genome-scale reconstruction of *E. coli*. *BMC Genomics*. 2011; 12:9. doi: [10.1186/1471-2164-12-9](https://doi.org/10.1186/1471-2164-12-9) PMID: [21208457](https://pubmed.ncbi.nlm.nih.gov/21208457/)
43. Baumgart M, Dogan B, Rishniw M, Weitzman G, Bosworth B, Yantiss R, et al. Culture independent analysis of ileal mucosa reveals a selective increase in invasive *Escherichia coli* of novel phylogeny relative to depletion of Clostridiales in Crohn's disease involving the ileum. *ISME J*. 2007; 1(5):403–18. PMID: [18043660](https://pubmed.ncbi.nlm.nih.gov/18043660/)
44. Petty NK, Bulgin R, Crepin VF, Cerdeno-Tarraga AM, Schroeder GN, Quail MA, et al. The *Citrobacter rodentium* genome sequence reveals convergent evolution with human pathogenic *Escherichia coli*. *J Bacteriol*. 2010; 192(2):525–38. doi: [10.1128/JB.01144-09](https://doi.org/10.1128/JB.01144-09) PMID: [19897651](https://pubmed.ncbi.nlm.nih.gov/19897651/)