Infection and Immunity

Sequencing and Functional Annotation of Avian Pathogenic Escherichia coli Serogroup O78 Strains Reveal the Evolution of E. coli Lineages Pathogenic for Poultry via Distinct Mechanisms

Francis Dziva, Heidi Hauser, Thomas R. Connor, Pauline M. van Diemen, Graham Prescott, Gemma C. Langridge, Sabine Eckert, Roy R. Chaudhuri, Christa Ewers, Melha Mellata, Suman Mukhopadhyay, Roy Curtiss III, Gordon Dougan, Lothar H. Wieler, Nicholas R. Thomson, Derek J. Pickard and Mark P. Stevens
Infect. Immun. **2013, 81(3):838. DOI: 10.1128/IAI.00585-12. Published Ahead of Print 28 December 2012.**

IAI Journals.ASM.org

Sequencing and Functional Annotation of Avian Pathogenic *Escherichia coli* Serogroup O78 Strains Reveal the Evolution of *E. coli* Lineages Pathogenic for Poultry via Distinct Mechanisms

Francis Dziva,[a]* Heidi Hauser,[b] Thomas R. Connor,[b] Pauline M. van Diemen,[a]* Graham Prescott,[a] Gemma C. Langridge,[b] Sabine Eckert,[b]* Roy R. Chaudhuri,[c] Christa Ewers,[d]* Melha Mellata,[e] Suman Mukhopadhyay,[e]* Roy Curtiss III,[e] Gordon Dougan,[b] Lothar H. Wieler,[d] Nicholas R. Thomson,[b] Derek J. Pickard,[b] Mark P. Stevens[f]

Enteric Bacterial Pathogens Laboratory, Institute for Animal Health, Compton, Berkshire, United Kingdom[a]; Pathogen Genomics, The Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom[b]; Department of Veterinary Medicine, University of Cambridge, Cambridge, United Kingdom[c]; Veterinary Faculty, Free University Berlin, Berlin, Germany[d]; The Biodesign Institute and School of Life Sciences, Arizona State University, Tempe, Arizona, USA[e]; Division of Infection and Immunity, The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Midlothian, United Kingdom[f]

Avian pathogenic *Escherichia coli* (APEC) causes respiratory and systemic disease in poultry. Sequencing of a multilocus sequence type 95 (ST95) serogroup O1 strain previously indicated that APEC resembles *E. coli* causing extraintestinal human diseases. We sequenced the genomes of two strains of another dominant APEC lineage (ST23 serogroup O78 strains χ7122 and IMT2125) and compared them to each other and to the reannotated APEC O1 sequence. For comparison, we also sequenced a human enterotoxigenic *E. coli* (ETEC) strain of the same ST23 serogroup O78 lineage. Phylogenetic analysis indicated that the APEC O78 strains were more closely related to human ST23 ETEC than to APEC O1, indicating that separation of pathotypes on the basis of their extraintestinal or diarrheagenic nature is not supported by their phylogeny. The accessory genome of APEC ST23 strains exhibited limited conservation of APEC O1 genomic islands and a distinct repertoire of virulence-associated loci. In light of this diversity, we surveyed the phenotype of 2,185 signature-tagged transposon mutants of χ7122 following intra-air sac inoculation of turkeys. This procedure identified novel APEC ST23 genes that play strain- and tissue-specific roles during infection. For example, genes mediating group 4 capsule synthesis were required for the virulence of χ7122 and were conserved in IMT2125 but absent from APEC O1. Our data reveal the genetic diversity of *E. coli* strains adapted to cause the same avian disease and indicate that the core genome of the ST23 lineage serves as a chassis for the evolution of *E. coli* strains adapted to cause avian or human disease via acquisition of distinct virulence genes.

Avian pathogenic *Escherichia coli* (APEC) imposes substantial economic and welfare costs on poultry producers worldwide. Respiratory infections typically involve inflammation of the air sacs and lung and may spread to visceral organs, causing perihepatitis, pericarditis, peritonitis, salpingitis, and sepsis (1). A need exists for effective cross-protective vaccines to control APEC in poultry, since autologous bacterins confer limited serotype-specific protection, and control via antibiotics is hindered by resistance and restrictions on prophylaxis. Diverse serotypes are associated with disease, and the molecular mechanisms underlying mucosal colonization and systemic translocation are ill defined. Serogroup O1, O2, and O78 strains are frequently isolated from diseased poultry and mostly belong to multilocus sequence types 95 and 23 (ST95 and ST23) (http://mlst.ucc.ie/mlst/dbs/Ecoli), as evidenced by recent surveys in chickens (2) and turkeys (3). The genetic traits that define the APEC pathotype and the extent of inter- and intra-ST and serogroup diversity are incompletely understood.

Sequencing of the complete genome of a ST95 strain of APEC serotype O1:K1:H7 (APEC O1) revealed that it is closely related to extraintestinal pathogenic *E. coli* (ExPEC) strains associated with human urinary tract infections (4). This is further evident from multilocus sequence typing and the extensive conservation of virulence-associated loci in APEC and human ExPEC (4–13) and indicates that APEC found in poultry may pose a threat of zoonosis. Indeed, a subset of APEC ST95 serogroup O18 isolates produced pathology comparable to that of human neonatal meningitis-associated *E. coli* (NMEC) in a rat model of meningitis, and reciprocally, NMEC O18 isolates caused systemic disease in chickens (14). Analysis of the genomes of ExPEC and nonpathogenic *E. coli* isolates recently identified conserved ExPEC-specific antigens, some of which were protective in a murine model of sepsis and may be useful in controlling a range of ExPEC infections (12).

Received 5 November 2012 **Returned for modification** 12 December 2012
Accepted 19 December 2012

Published ahead of print 28 December 2012

Editor: J. B. Bliska

Address correspondence to Mark P. Stevens, Mark.Stevens@roslin.ed.ac.uk.

* Present address: Francis Dziva, School of Veterinary Medicine, University of the West Indies, St Augustine, Trinidad and Tobago; Pauline M. van Diemen, Jenner Institute, Nuffield Department of Clinical Medicine, University of Oxford, Oxford, United Kingdom; Sabine Eckert, Oxford Nanopore Technologies, Edmund Cartwright House, Oxford, United Kingdom; Christa Ewers, Fachtierärztin für Mikrobiologie, Institut für Hygiene und Infektionskrankheiten der Tiere, Justus-Liebig-Universität, Giessen, Germany; Suman Mukhopadhyay, Division of Microbiology and Infectious Diseases, National Institute of Allergy and Infectious Diseases, Bethesda, Maryland, USA.

H.H. and T.R.C. contributed equally to this article.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/IAI.00585-12.

doi:10.1128/IAI.00585-12

Analysis of the sequences and functions of APEC genes in natural hosts may therefore inform the design of strategies to control various types of ExPEC.

Current understanding of the genetic basis of the virulence of APEC in poultry derives mostly from studies with defined or random mutants. Signature-tagged transposon mutagenesis (STM) assigned roles to 28 genes of an APEC ST95 O2:K1:H5 strain following intratracheal inoculation of chickens and recovery of mutants from the spleen (15), including genes associated with the production of proven APEC virulence factors, such as the capsule (16), lipopolysaccharide (16, 17), SitABCD metal transporter (18), and YjjQ regulator (19). Analysis of the same APEC O2:K1:H5 mutant library in a low-dose model of chicken lung colonization identified a novel fimbrial locus, encoding ExPEC adhesin I (*yqi*), which influences virulence (20, 21). A number of other virulence factors have been identified by using defined mutants in poultry models, including the autotransported proteins AatA (22), Tsh (23, 24), and Vat (25), fimbrial gene clusters *pap* (26), *fim* (27, 28), *stg* (29), and *csg* (28), a degenerate *inv*/*mxi-spa*-like type III secretion locus termed ETT2 (30), and factors mediating iron uptake (31–33) and phosphate transport (34). Flagellin contributes to persistence and invasion after oral dosing of chicks with APEC O78 (28), and ExPEC invasins (Ibe proteins) also facilitate APEC invasion (35, 36). Indirect effects on the expression of virulence-associated fimbriae have been reported for mutants with a constitutively activated *pho* regulon (37) and those lacking the *pst* system (38), *ibeA*, *ibeB*, and *ibeT* (36, 39), the *barA-uvrY*-encoded two-component sensory system (40), and a type VI protein secretion system (41). These reports may partially explain the attenuation of such mutants in chickens and indicate that the effect of mutations may propagate through complex networks.

Plasmids contribute significantly to APEC biology. APEC ST23 O78 strain χ7122 has four plasmids, and curing has revealed their individual and collective contributions to pathogenesis and fitness-associated phenotypes (42, 43). Loss of all three of the major plasmids, and of the ColV pChi7122-1 plasmid in particular, severely attenuated the virulence of χ7122 in chickens (42). pChi7122-1 (formerly pAPEC-1) harbors a ca. 80-kb region enriched in virulence-associated loci, including four iron acquisition systems (*iutA*, *iucABCD*, *sitABCD*, *iroBCDN*), *tsh*, *iss*, and other candidate virulence genes (44). This region also appears to contribute to air sac colonization and the induction of pathology by an O nontypeable:H28 APEC field isolate in poultry (45). Of the other plasmids found in χ7122, pChi7122-2 is an IncFII plasmid that encodes an ABC iron transport system (*eitABCD*), whereas pChi7122-3 (IncI2) carries a putative type IV fimbrial locus (43). A further ColE2-like cryptic plasmid exists in χ7122; however, its role is ill defined.

Although researchers have begun to unravel the roles of specific virulence factors of APEC, no studies have yet compared the complete genomes of APEC strains belonging to dominant serogroups and sequence types. We therefore sequenced the genomes of two ST23 serogroup O78 strains belonging to EcoR group B1 (χ7122 and IMT2125) and compared them to each other and to the reannotated sequence of APEC O1 (ST95 EcoR group B2) (4). Strain χ7122 has been used extensively as a model APEC strain, for example, to identify *in vivo*-transcribed genes (31) and APEC-specific loci (17) and to define the roles of plasmids (42, 43) and specific loci (16–18, 24, 31–34, 37, 40, 46). We also sequenced a ST23 serogroup O78 strain (E85) belonging to a distinct pathotype associated with human diarrheal illness (enterotoxigenic *E. coli* [ETEC]) in order to explore phylogenetic relationships between lineages and the evolution of virulence in *E. coli*. To unravel the role of APEC O78 genes *in vivo*, we screened a library of random χ7122 transposon mutants in target animals by signature-tagged mutagenesis and identified factors that play strain- and tissue-specific roles.

MATERIALS AND METHODS

Bacterial strains. Avian pathogenic *E. coli* strain χ7122 is a spontaneous nalidixic acid-resistant (*gyrA*) mutant of strain EC1 (O78:H9, ST23; EcoR group B1), originally isolated from the liver of a diseased turkey (47). A defined *rfb* mutant of strain χ7122 (16) was kindly supplied by J. M Fairbrother, University of Montreal. IMT2125 (O78:H9, ST23; EcoR group B1) was isolated from a chicken with fatal airsacculitis in Germany in 1999 and is highly virulent in a chicken infection model (C. Ewers et al., unpublished data). *E. coli* K-12 strain S17-1λ*pir* was used as conjugative donor of transposon-encoding suicide replicons (48), and chemically competent TOP10 cells (Invitrogen, Paisley, United Kingdom) were used for routine cloning. Strains were cultured in Luria-Bertani (LB) medium, supplemented as appropriate with 100 μg/ml ampicillin, 50 μg/ml kanamycin, or 20 μg/ml nalidixic acid.

Animals. Big-5 FLX specific-pathogen-free unvaccinated turkey poults were received from Aviagen Ltd. (Tattenhall, Cheshire, United Kingdom) at the age of 1 day and were housed in floor pens in biosecure accommodations, with access to water and vegetable protein-based feed *ad libitum*. All procedures were conducted according to the requirements of the Animals (Scientific Procedures) Act 1986 (project license 30/2463) with the approval of the local Ethical Review Committee.

Sequencing and annotation. Total genomic DNAs of strains χ7122 and IMT2125 were sequenced using the 454/Roche GS FLX analyzer and 3-kb-insert paired-end libraries. For χ7122, this yielded 764,001 reads with a theoretical sequence coverage of 21-fold, whereas for IMT2125, 419,947 reads were obtained, providing 14-fold coverage. *De novo* assemblies were produced using the 454/Roche Newbler assembly algorithm (49). For χ7122, this yielded 160 contigs, the largest of which was 292,307 bp, with an N50 contig size of 104,791 bp. For IMT2125, *de novo* assembly generated 234 contigs, the largest of which was 155,858 bp, with an N50 contig size of 56,670 bp. Gaps between contigs were closed by directed PCR, and the products were sequenced using ABI 3730 capillary sequencers with BigDye Terminator chemistry. The sequences were manipulated to the standard of an "improved high-quality draft" (50). The sequence of the χ7122 chromosome consists of 4,771,701 bp and is contained in 4 contigs. The order and orientation of each χ7122 contig relative to the others have not been established. A total of 8 sequence gaps exist within the 4 χ7122 contigs, but these are spanned by PCR products and have therefore been joined with N′s (any nucleotide) to maintain the order and orientation of sequence within each contig. The sequence of the IMT2125 chromosome consists of 4,754,148 bp in 15 contigs, the order and orientation of which are not defined. A total of 32 sequence gaps exist in these 15 contigs, but the gaps are joined with N′s, since they were spanned by PCR. The sequences of plasmids in each strain were obtained from the same sequence reads as those used to assemble the chromosomes and were also manipulated to the "improved high-quality draft" standard (50). Four contiguous plasmids were obtained for χ7122; they comprised 103,201 bp, 82,676 bp, 56,676 bp, and 4,300 bp. For IMT2125, four plasmids, ranging in size from ca. 118 kb to 1.6 kb (118,067 bp containing 2 gaps joined with N′s as spanned by PCR; 106,097 bp with 1 joined gap; 4,107 bp; and 1,616 bp), were identified. Artemis (51, 52) was used to collate data and facilitate annotation. Sequences spanning the rRNA operons of *E. coli* K-12 were used to order and join contigs over rRNA loci. Initial coding sequences (CDS) were transferred from *E. coli* K-12 (53) using Annotations Update (a proprietary script available from the authors on request) and were amalgamated with Glimmer3 software predictions for both χ7122 and IMT2125. Prodigal was used to ensure that all coding

regions were identified (54). The gene predictions were further checked by searching all of the predicted genes from all three of the genomes against the raw sequence of each genome in turn using TBLASTN. This identified any homologous genes that were present, but not predicted by Prodigal or Glimmer3, in one or another of the genomes. The same analysis and annotation procedures were also run on the sequenced genome of the APEC O1 strain (GenBank accession number CP000468) (4) in order to correct errors in the original annotation and facilitate comparison. Orthologous proteins in the chromosomes of χ7122, IMT2125, and APEC O1 were identified with OrthoMCL (55), which uses a Markov clustering (MCL) algorithm to identify clusters of orthologous proteins from a reciprocal all-against-all BLASTP search performed on the predicted protein sequences of the sequenced genomes. The OrthoMCL analysis was run in mode 1, with a *P* value cutoff of 1e−5, an MCL inflation parameter of 1.5, and a maximum weight of 316. The Artemis Comparison Tool (56) was used to align TBLASTX comparisons of the chromosomes of χ7122, IMT2125, and the APEC O1 strain.

Chromosome sequences for χ7122 and IMT2125 have been submitted to public domain databases. The sequences of χ7122 plasmids have been described elsewhere (43, 44), and pIMT2125-1, -2, -3, and -4 sequences are available from ftp://ftp.sanger.ac.uk/pub/pathogens/Escherichia/coli/. To facilitate exploration of the annotated chromosomal sequences of χ7122, IMT2125, and APEC O1, we have produced EMBL files that can be viewed in Artemis with color coding of CDS to denote their conservation across the three strains and to highlight strain-specific orthologs. These can be obtained from ftp://ftp.sanger.ac.uk/pub/pathogens/Escherichia/coli/, where instructions for viewing the files can also be found.

Phylogenetic analysis. OrthoMCL was also used for phylogenetic analyses of whole-genome sequences from enterohemorrhagic *E. coli* (EHEC) O103:H2 (strain 12009; accession number AP010958), EHEC O26:H11 (strain 11368; accession number AP010953), EHEC O157:H7 (strain EDL933; accession number AE005174), enteroaggregative *E. coli* (EAEC) O104:H4 (strain 55989; accession number CU928145), enterotoxigenic *E. coli* (ETEC) O139:H28 (strain E24377A; accession number CP000800), *E. coli* K-12 (strain MG1655; accession number U00096), enteropathogenic *E. coli* (EPEC) O127:H6 (strain E2348/69; accession number FM180568), and uropathogenic *E. coli* (UPEC) O18:K1:H7 (strain UTI89; accession number CP000243). In addition, we included data from the genome sequence of strain E85, a human ETEC strain belonging to the same sequence type (ST23) and serogroup (O78) as χ7122 and IMT2125. Strain E85 was isolated in Dhaka, Bangladesh, in 1986 and was kindly supplied by Ann-Mari Svennerholm (Department of Microbiology and Immunology, Institute of Biomedicine at Sahlgrenska Academy, University of Gothenburg, Göteborg, Sweden). Genomic DNA was isolated from *E. coli* E85 by using a Promega (Southampton, United Kingdom) genomic preparation kit, sequenced using standard Illumina/Solexa 78-base paired-end reads to ×100 coverage, and assembled using Velvet as described previously (57). The draft E85 assembly may be obtained from ftp://ftp.sanger.ac.uk/pub/pathogens/Escherichia/coli/. OrthoMCL analysis identified 3,153 proteins that were homologous across all the strains examined. These proteins were extracted and were aligned individually using MUSCLE (58). The alignments were then concatenated, and a tree was generated from this alignment using RAxML, version 7.0.4 (59), with a WAG substitution matrix and a gamma model of rate heterogeneity.

Construction and screening of a transposon mutant library of χ7122. Suicide replicons harboring 95 signature-tagged mini-Tn*5*Km2 transposons were kindly supplied by C. M. Tang, University of Oxford, and were introduced separately into strain χ7122 by conjugation from *E. coli* S17-1λ*pir* with selection for kanamycin and nalidixic acid resistance as described previously (60). Exconjugants were confirmed to be sensitive to ampicillin, indicating that chromosomal integration of pUT-mini-Tn*5*Km2 plasmids had not occurred. A total of 2,850 mutants were arrayed in 30 96-well plates, such that each mutant in a plate was distinguishable by a unique oligonucleotide tag. Twenty-three pools of 95 mutants (2,185 in total) were amplified to stationary phase in LB medium, and ca. 1.5×10^8 CFU of each pool was injected separately into the left caudal air sac of each of three 3-week-old turkey poults. Bacteria resistant to kanamycin and nalidixic acid were recovered from homogenates of the lung and liver 24 h postinoculation. Genomic DNA was extracted from the inoculum, and a suspension of at least 10,000 colonies was recovered from the lung or liver using cetyl trimethyl ammonium bromide. This output pool size is the minimum required to state at the 95% confidence interval that mutants are absent from the population owing to attenuation rather than to chance (61). Separate DNA extractions were performed using bacteria recovered from each bird and tissue; however, since pilot data revealed minimal interanimal variance and reliable negative selection of mutants, the three output pool replicates were combined to simplify analysis.

Identification of negatively selected mutants. Signature tags from cognate input and output pools were amplified separately by PCR in the presence of digoxigenin-dUTP using primers P2 and P4 (see Table S1 in the supplemental material) with a Roche DIG DNA labeling and detection kit (Roche Diagnostics Ltd., Burgess Hill, United Kingdom). Primer sequences were removed from the amplified products by HindIII digestion, and the labeled tags were hybridized separately with identical colony blots. Mutants that were negatively selected *in vivo* relative to the inoculum were identified by visual inspection of the intensity of hybridization signals derived from the input and output pool tags. The DNA flanking the site of transposon insertion from attenuated mutants was cloned by ligation of EcoRI- or EagI-restricted genomic DNA into similarly restricted pBluescript KS(+) and was transformed into chemically competent *E. coli* TOP10 cells with selection for ampicillin and kanamycin resistance. Recombinant plasmid DNA was purified using a QIAprep Spin Miniprep kit (Qiagen GmbH, Hilden, Germany), and insert sequences were obtained using the mini-Tn*5*Km2-specific primer P6 for EagI clones or P10 for EcoRI clones (Lark Technologies Inc., Saffron Walden, United Kingdom) (see Table S1 in the supplemental material). Transposon-flanking sequences were analyzed using Artemis and the BLASTN search algorithm at the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/blast/) and are available from the authors on request.

Validation of STM findings with defined χ7122 mutants. Six defined null mutants of χ7122 lacking candidate virulence factors identified by STM were generated by λRed recombinase-mediated integration of linear PCR products. Primers were designed to amplify the pKD4-encoded kanamycin resistance cassette (62) with 50 nucleotide extensions homologous to the flanking regions of the target genes (see Table S1 in the supplemental material). Strain χ7122 harboring plasmid pKD46 was electroporated with these amplicons following induction of the γ, β, and *exo* genes with 0.2% (wt/vol) L-arabinose at 30°C, essentially as described previously (62). Kanamycin-resistant transformants were selected at 37°C and were confirmed to have lost pKD46 on the basis of sensitivity to ampicillin. Recombination was confirmed to have replaced the entire coding sequence of each of the target genes (*cadA*, *evgS*, *fliC*, *fucP*, *htrA*, *yccC* [*etk*]) with the pKD4-encoded resistance cassette by PCR with a locus-specific forward (F) primer and K1 and with the locus-specific reverse (R) primer and K2, and/or by using internal gene-specific primers (see Table S1 in the supplemental material). Approximately 1.5×10^8 CFU of each mutant was administered separately to three 3-week-old turkey poults by the intra-air sac route as described above. Three birds were inoculated with the parent strain, and a further seven transposon mutants isolated in the primary screen were also evaluated separately in the same way (7B1 [*entF*::Tn], 7G2 [*exoP*::Tn], 14F1 [*traE*::Tn], 2A2 [*yadL*::Tn], 10D9 [*yddB*::Tn], 14E6 [*yejO*::Tn], 14E5 [*ynbB*::Tn]). For logistical reasons, the Δ*evgS*::kan^R and Δ*yccC*::kan^R mutants were tested relative to the wild type in four birds per strain in a separate study. Challenge strains were enumerated in the lung, liver, and spleen 24 h postinoculation by plating serial 10-fold dilutions of tissue homogenates onto MacConkey agar containing antibiotics as appropriate.

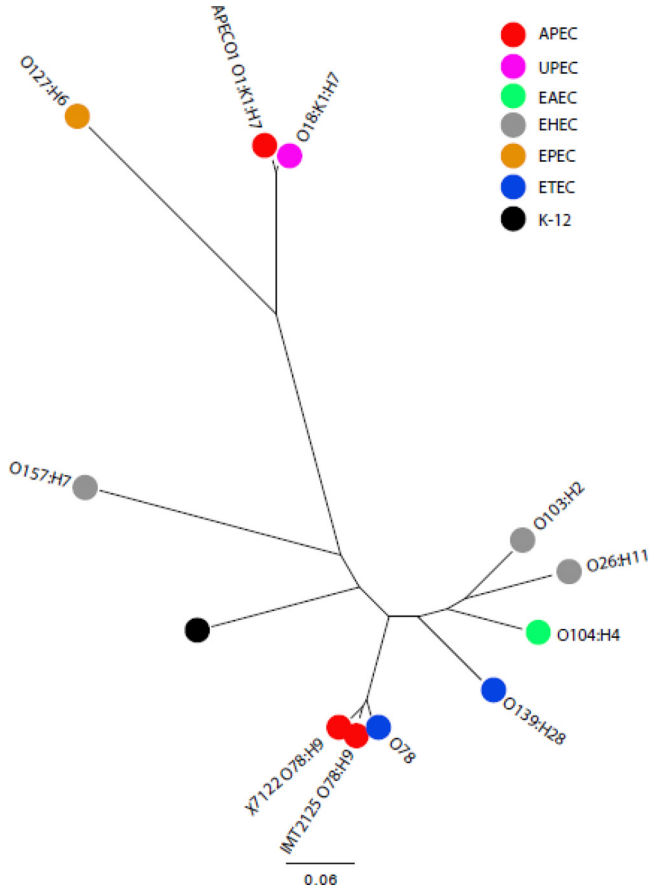

FIG 1 Phylogenetic tree generated using RAxML (version 7.0.4) from the concatenated sequences of the 3,153 proteins that are shared by the 12 *E. coli* strains analyzed (see Materials and Methods). Pathotypes are color-coded according to the key, and serotypes are given at the position of each strain on the tree.

Statistical analysis. Bacterial counts in the lung, liver, and spleen were analyzed for the effect of mutation by means of a one-way F-test after $\log_{10}$ transformation of the data using GraphPad Prism, version 5.04 (GraphPad Software, Inc., La Jolla, CA). Subsequent pairwise comparisons with the wild type were performed, and P values of ≤ 0.05 were taken to be significant.

Nucleotide sequence accession numbers. The chromosome sequences determined in this study have been submitted to GenBank under accession numbers HE962388 for χ7122 and HE964769 for IMT2125.

RESULTS AND DISCUSSION

Comparison of the genome sequences of ST23 serogroup O78 strains with that of the APEC ST95 serogroup O1 strain. Surveys of the prevalence of specific genes have indicated that APEC strains are highly heterogenous; however, the full extent of inter- and intraserogroup and sequence type diversity across genomes is ill defined. We therefore sequenced and annotated the chromosomes and plasmids of χ7122 and IMT2125, representing the prevalent ST23 serogroup O78 lineage, as described in Materials and Methods. Circular diagrams showing features of the chromosomes of the two strains were made (see Fig. S1 and S2 in the supplemental material). The sequences of the four plasmids of χ7122 have been described elsewhere (43, 44). Four contiguous plasmids of IMT2125 were also obtained. Our analysis of the

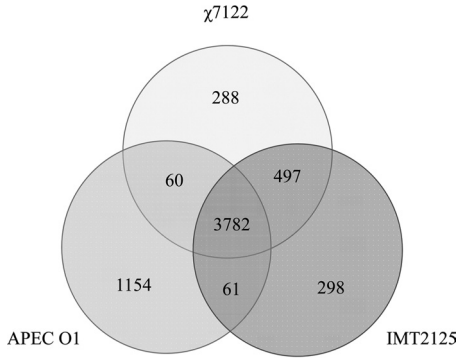

FIG 2 Venn diagram showing the numbers of conserved and strain-specific orthologous proteins predicted by OrthoMCL to be encoded by the chromosomes of strains χ7122, IMT2125, and APEC O1.

APEC O1 sequence annotated by Johnson et al. (4) indicated, on the basis of conservation with the latest *E. coli* K-12 MG1655 annotation, that some gene predictions and start sites are erroneous. To ensure that this did not suggest a higher number of strain-specific features, we ran the same annotation pipeline as that used for the APEC O78 strains on the raw APEC O1 sequence.

For comparison, we also obtained the draft sequence of an ST23 serogroup O78 human ETEC strain (E85). We used the consistent annotation of the genomes to identify 3,153 shared core proteins. RAxML was then used to examine the phylogenetic relationships between the ST23 and ST95 strains, as well as between these strains and reference genomes in public databases for eight other *E. coli* strains representing different pathotypes. The analysis supported the clustering of APEC ST95 O1 with uropathogenic *E. coli* reported previously but clearly indicated that APEC ST23 O78 strains are distinctly and distantly clustered (Fig. 1). Indeed, the APEC O78 strains occupy positions on the tree that suggest they are closely related to the ETEC strain of the same serogroup and sequence type, even though strain E85 is associated with acute enteritis in humans rather than with systemic avian disease. These findings challenge the prevailing view that APEC strains are closely related to other ExPEC strains, and to UPEC strains in particular. In turn, the data suggest that broad classification of *E. coli* strains into diarrheagenic and extraintestinal categories is not always consistent with the phylogenetic signatures present in the chromosomal sequences of the strains.

Given the clustering of APEC ST23 strains and their phylogenetic distance from the APEC ST95 strains, we next examined the extent of conservation of genes within and between the sequence types in order to assess the extent to which they are adapted to cause the same disease via acquisition of common or unique determinants. In general, pairwise analysis of the chromosome sequences of χ7122, IMT2125, and APEC O1 revealed conservation of the sequence and order of genes (see Fig. S3 in the supplemental material). The genome sequences obtained for χ7122, IMT2125, and APEC O1 are predicted to encode 4,627, 4,638, and 5,057 genes, respectively. OrthoMCL analysis was used to search for shared and strain-specific orthologous proteins. This analysis identified 3,782 chromosomally encoded core predicted proteins shared by all three strains, but it also identified predicted proteins unique to χ7122 (n, 288), IMT2125 (n, 298), and APEC O1 (n, 1,154) (Fig. 2). A relatively small number of orthologous proteins were shared only by APEC O1 and χ7122 (n, 60), or only by APEC

O1 and IMT2125 (*n*, 61), whereas the two APEC O78 chromosomes encoded 497 orthologous proteins that were not shared by APEC O1. Taken together, these data indicate marked divergence of the O78 strains from the sequenced APEC O1 strain but a relatively high degree of similarity of the O78 strains to each other.

Because the salient features of the APEC O1 genome have been described previously (4), emphasis is placed here on further differences within and between serogroups and sequence types. A striking feature of this comparison is the low extent of conservation in χ7122 and IMT2125 of the 43 genomic islands of APEC O1 described by Johnson et al. (4). Thirty of the 43 genomic islands of APEC O1 were absent or poorly conserved in the two APEC O78 strains. These 30 genomic islands account for the bulk of the APEC O1-specific orthologs described above and include 4 predicted pathogenicity islands (PAIs) (PAI $I_{APEC\ O1}$ [~83 kb] near tRNA-Phe, carrying *tia*, *ireA*, and *pap*; PAI $II_{APEC\ O1}$ [~28 kb] near tRNA-Asp; PAI $III_{APEC\ O1}$ [~19 kb] near tRNA-Thr, carrying *vat*; PAI $IV_{APEC\ O1}$ [~87 kb] near tRNA-Asn, encoding yersiniabactin). The lack of conservation of the 43 genomic islands described previously for APEC O1 in χ7122 and IMT2125 is consistent with the variable conservation of 12 such islands among other APEC isolates in a recent survey (9). Other virulence-associated loci present in APEC O1 but absent from χ7122 and IMT2125 include the *chu* operon, involved in iron uptake; the *kps* locus, mediating group II capsule synthesis; and genes encoding cytolethal distending toxin, Auf fimbriae, and Ibe invasion-associated proteins. The *sitABCD* genes in APEC O1 island 10 were absent from this position in χ7122 but are harbored by pChi7122-1 in this strain (44). Prophages exhibited partial homology between the strains and accounted for some of the differences between the O78 strains. Some prophages were conserved in APEC O1 and one O78 strain but not the other (e.g., those between *yfdC-dsdA* and *ydfI-rspB* are mostly shared by APEC O1 and χ7122). As expected, genes associated with lipopolysaccharide core biosynthesis were conserved in all three strains but not in *E. coli* K-12, whereas differences exist between APEC O1 and the serogroup O78 strains in the region between *gnd* and *galF* that determines O-antigen biosynthesis. The few islands that were conserved across all three strains (e.g., islands 18, 26, and 37) were also conserved in other sequenced ExPEC strains (4, 9).

Many of the orthologs specific to the APEC ST23 O78 strains are encoded in clusters (see Fig. S1 and S2 in the supplemental material). Among the virulence-associated loci present in χ7122 and IMT2125 but absent from APEC O1 is the degenerate type III secretion locus ETT2, located between *yqeG* and the *glyU* tRNA locus in many *E. coli* strains (63). The ETT2 locus in χ7122 and IMT2125 is truncated from within *eivA* to the end of the locus described for *E. coli* O157:H7 strain Sakai, consistent with the truncation described for another sepsis-associated APEC O78 strain, termed 789 (30). Within ETT2, the predicted *spaS* and *eivJ* genes contain distinct frameshift mutations in χ7122 and IMT2125, consistent with the varied patterns of mutational attrition of the locus reported among *E. coli* strains (30, 63). The *ymcDCBA-yccZ-etp-etk* genes, which determine O-antigen capsule biosynthesis (located between *appA* and *yccM*) (64), were conserved in both χ7122 and IMT2125 but absent in APEC O1. The same applies to the *stg* fimbrial locus (located between *pstS* and *glmS*), which influences air sac colonization in chickens (29), and two cryptic fimbrial loci (the *sfm* locus, between *folD* and *intD*, and the *yraHIJK* genes, between *agaI* and *yraL*). Forced induction of these operons has established that they encode functional adhesins in *E. coli* K-12 (65). A ca. 62-kb region between *yjgB* and *yjhR*, carrying the *fecIRABCDE* genes for citrate-dependent iron(III) uptake, genes encoding a putative sugar uptake phosphotransferase system (PTS), and numerous genes of unknown function, was also largely conserved in χ7122 and IMT2125 but absent in APEC O1.

Many of the APEC ST23-specific functions are putatively related to metabolism and include clusters of genes predicted to mediate 4-hydroxyphenylacetate degradation (between *ygiY* and *tsr*), cyanate metabolism (between *codA* and *lacA*), use of 3-(3-hydroxyphenyl)propionic acid (*mhp* operon, between *lacI* and *yaiL*), the dioxygenolytic pathway for initial catabolism of 3-phenylpropionic acid (*hcaRACBD*, between *hcaT* and *yphA*), and fructoselysine catabolism (*yhfLMNOPQR*, between *cysD* and *yhfS*), as well as a putative 10-subunit hydrogenase complex (hydrogenase-4) encoded by the *hyf* locus (*hyfABCDEFGHIJ-hyfR-foc*, between *bcp* and *uraA*). Other metabolism-related clusters of less-defined function encoded by χ7122 and IMT2125 but not APEC O1 were also detected (e.g., between *ymjA* and *pspF*, *ydbL* and *ynbD*, *hipB* and *uxaB*, *yeaS* and *rnd*, and *yqiN* and *parE*).

Comparison of the genome sequences of APEC serogroup O1 and O78 strains extends observations made by Brzuszkiewicz et al. in relation to the evolution of virulence in another group of extraintestinal *E. coli* strains (66). These authors reported that two human uropathogenic *E. coli* (UPEC) serogroup O6 strains differed markedly in the content of the accessory genome, despite the fact that they were adapted to cause the same disease. Approximately 430 genes specific to each UPEC O6 strain were identified, and these were distributed mostly in pathogenicity islands and smaller islets (66).

***In vivo* functional annotation of the genome of APEC ST23 O78 strain χ7122.** Since analysis of genome sequences and phylogeny indicated the evolution of APEC from distinct lineages, and since preceding studies mostly used ST95 strains, we sought to identify virulence factors of APEC ST23. A library of 2,185 signature-tagged mini-Tn*5*Km2 mutants of strain χ7122 was screened in turkeys for mutants defective in colonization of the respiratory tract and systemic translocation. Pilot studies indicated that strain χ7122 produces airsacculitis, perihepatitis, and pericarditis typical of natural infections when administered to the left caudal air sac of a 3-week-old turkey poult, with recovery of ca. 10^7 CFU/g from the lung and liver 24 h postinoculation. In contrast, a Δ*rfb* O-antigen-deficient mutant of χ7122 was recovered at ca. 10^2 CFU/g from these sites 24 h after the inoculation of age-matched animals in the absence of overt pathology, in agreement with published data (16). It was not feasible to extend the time interval postinoculation beyond 24 h, because most birds infected with χ7122 presented symptoms consistent with agreed humane endpoints; however, it was evident from these pilot studies that adequate time had elapsed for attenuation to be evident. Inoculation of three separate birds with a single pool of 95 signature-tagged mutants revealed that the same mutants were reliably negatively selected, with minimal variance in the relative intensity of hybridization signals between replicates and no evidence of stochastic loss of mutants in lung or liver homogenates (which would be indicative of bottleneck effects) (data not shown). We therefore screened the entire library in pools of 95 mutants, giving each pool via the intra-air sac route to three 3-week-old turkey poults, with

recovery of mutants from lung and liver homogenates 24 h postinoculation.

The composition of input and output pools was examined by amplification and detection of signature tags, and a total of 119 mutants were inferred to be attenuated on the basis of their absence or negative selection in the output pool relative to the inoculum. We determined the insertion sites in 75 such mutants by subcloning and sequencing (Table 1). Conservation of the mutated genes in IMT2125 and APEC O1 is indicated by footnotes in Table 1. Multiple attenuating mutations were independently detected in several genes (*evgS*, *exoP*, *intQ*, 5′ of *traM*, *traM*-MM2_058, *trbB*, *yadL*, *ycbY*, *ycgK*, *ygaV*, *yjiZ*, *ynbB*), including a number of siblings with the same tagged transposon and insertion site that were negatively selected in separate pools (*evgS*, *intQ*, 5′ of *traM*, *trbB*, *ycbY*, *ycgK*, *ygaV*, *yjiZ*, *ynbB*). The reliable negative selection of such mutants provides a measure of confidence that mutants are absent owing to their genotype rather than to stochastic loss due to bottleneck effects or to sampling populations of inadequate size. This is further supported by the detection of multiple attenuating mutations in the same pathway. For example, five mutations were identified in the *yadN-ecpD-htrE-yadMLKC* cluster, predicted to encode fimbrial subunits and a chaperone-usher secretion system homologous to those involved in the production of *E. coli* type I fimbriae. A total of 10 mutations predicted to impair F-pilus assembly and conjugal transfer were identified, with most mapping to pChi7122-2 (5′ of *traM*, *trbB*, *trbF*, *traM*-MM2_058) and single insertions in *traE* on pChi7122-1 and *traJ* on pChi7122-3.

A number of attenuating mutations detected in the present screen support observations from other studies on APEC pathogenesis in poultry. Thus, a flagellin subunit (*fliC*) mutant of χ7122 was not recovered from the lung or liver (Table 1), in agreement with the finding that *fliC* mutation in APEC O78 strain EC34195 impaired intestinal persistence in 10-day-old chicks and systemic virulence in day-old chicks after oral dosing (28). A signature-tagged χ7122 mutant with an insertion in the *epaP* gene, located in the ETT2 locus, was negatively selected in the liver, a finding consistent with the reduced lethality of a null mutant of APEC O78 strain 789 lacking the putative inner membrane ring of the secretion complex (Δ*eprHIJK*) after intraperitoneal infection of day-old chicks (30). In common with the present survey, an attenuating mutation in the putative polysaccharide hydrolase YcjM was detected by STM analysis of random APEC O2 mutants 48 to 72 h after intratracheal inoculation of chickens (15). Similarly, attenuating mutations detected in *yahB* and *ykgB* in the present screen are supported by the isolation of attenuating mutations in the adjacent genes (*yahA* and *ykgC*) in APEC O2 (15). Screening of the same APEC O2 library in a model of early lung colonization identified attenuating mutations in *evgA* and *tdcA* (21), consistent with the isolation of *evgS* and *tdcC* mutants here.

A mutant with an insertion in *ybcU* was negatively selected in the lung and liver and is predicted to lack a protein with 91% amino acid identity over 97 residues to Iss, encoded on pChi7122-1 (44), and 93% identity to the Bor protein of APEC O1 and lambdoid phages. Both Iss and Bor are localized on the APEC surface (67) and influence survival in serum when cloned into *E. coli* K-12 (68, 69). A further closely related homolog of APEC O1 Iss and Bor is encoded by *orf01374* in χ7122; however, we were able to fine map the attenuating insertion to *ybcU*, since the insertion is proximal to the 5′ end of *ybcU*, and the upstream sequence captured in the transposon (Tn)-flanking region is distinct from that upstream of *iss* on pChi7122-1 and from that of the chromosomal *orf01374* gene. Gene prevalence surveys have associated *iss* carriage with APEC virulence; however, a defined *iss* deletion in pAPEC-O2-ColV did not impair lethality for chick embryos despite induction of the gene *in vivo* (70), and the *iss* gene carried on pChi7122-1 does not play a major role in the serum resistance of χ7122 (16). Isogenic *iss* and *bor* APEC mutants do not exhibit the same levels of serum sensitivity (71), and the relative contributions of this family of proteins to strain χ7122 pathogenesis merit further study.

The apparent role of plasmid-encoded conjugal transfer systems in APEC pathogenesis is supported by other studies. For example, the APEC O1 *traY* and *traQ* genes are upregulated in chicken serum (72), and *traJ* has been reported to be induced during invasion of human brain microvascular endothelial cells (HBMECs) by neonatal meningitis-associated *E. coli* K1 (73). Moreover, *traJ* mutants show impairments in the invasion of HBMECs *in vitro* (73, 74), the penetration of the blood-brain barrier after intracardiac inoculation of neonatal rats (74), and dissemination from mesenteric lymph nodes to the blood and visceral organs after oral dosing (75).

No attenuating mutations were detected in several known virulence-associated loci of APEC, including genes mediating lipopolysaccharide biosynthesis or the production of type I, Stg, or Yqi fimbriae, and the *iro*, *iuc*, and *sit* genes, associated with iron uptake. It is not clear whether this reflects the absence of such mutants from the library or features of the infection model. Similarly, the genes identified in the present screen differ from those implicated in NMEC and UPEC virulence in rodent models (76–78). A distinct set of virulence factors were identified by STM when random APEC O2 mutants were screened 48 to 72 h after intratracheal infection of chickens (15) compared to an early lung colonization model (20, 21). None of the attenuating mutations detected by STM in χ7122 occurred in APEC O78 genes reported to be upregulated in chickens in a recent promoter trap screen (79). Differences between surveys reinforce the parallel evolution of avian pathogenic strains with distinct sets of virulence-associated genes and may also reflect differences in the host, inoculation route, and dose.

Validation of mutant phenotypes. To confirm the attenuation of mutants with defects in specific loci, λRed mutagenesis was used to replace six genes identified in the primary screen (*cadA*, *evgS*, *fliC*, *fucP*, *htrA*, *yccC* [*etk*]) with a kanamycin resistance cassette. Each mutant, as well as the parent strain, was then administered separately to the left caudal air sacs of three 3-week-old turkey poults, and bacteria in the lung and liver 24 h postinoculation were enumerated. Seven transposon mutants from the primary screen (with insertions in *entF*, *exoP*, *traE*, *yadL*, *yddB*, *yejO*, or *ynbB*) were also screened again in isolation as described above. We cannot exclude the possibility that such transposon mutants have polar or second-site defects; however, the data are presented as a means of validating the negative selection detected by STM in the absence of competition between coscreened mutants.

All mutants were recovered in lower numbers (at least 0.5 $\log_{10}$ CFU/g) than the parent strain tested in parallel in one or more of the organs sampled, except for the *yddB*::Tn and *yejO*::Tn mutants in the lung and the Δ*cadA*::kan^R, Δ*fucP*::kan^R, and *ynbB*::Tn mutants in the liver (Fig. 3). The absence of negative selection of *yddB*::Tn and *yejO*::Tn mutants in the lung is consistent with the

TABLE 1 Loci disrupted in signature-tagged mutants of APEC O78:H9 strain χ7122 that were negatively selected following intra-air sac inoculation of turkeys

Mutant	Gene	Function	Selection in: Lung	Liver
Known or predicted surface structures				
16E6	*fliC*	Flagellin subunit	−	−
2E12	*epaP*[a]	ETT2-encoded cryptic type III secretion system protein	+	−
5B11	*htrE*	Predicted outer membrane pilin usher protein	−	−
21A7	*yadK*	Predicted fimbrial protein	+	−
2A2	*yadL*	Predicted fimbrial protein	+	−
15F1	*yadL*	Predicted fimbrial protein	+	−
21D5	*yadN*	Predicted fimbrial protein	+	−
7E3	5′ of *orf02788*	Predicted chaperone protein for FimC-like precursor	−	−
2B7	*yccC* (*etk*)	Etk protein tyrosine kinase involved in O-antigen capsule synthesis	−	−
2H9	*ybcU*[a]	Close homolog of Iss and Bor lipoproteins mediating increased serum survival	−	−
Conjugation				
14F1	*traE*	Conjugal transfer pilus assembly protein (pChi7122-1)	+	−
21H2	*traJ*	Conjugal transfer protein (pChi7122-3)	+	−
8D9	5′ of *traM*	Conjugal transfer protein (upstream; pChi7122-2)	−	−
9D9	5′ of *traM*	Conjugal transfer protein (upstream; pChi7122-2)	+	−
1H3	*trbB*	Conjugal transfer pilus assembly periplasmic protein (pChi7122-2)	−	−
2H3	*trbB*	Conjugal transfer pilus assembly periplasmic protein (pChi7122-2)	−	−
16H3	*trbB*	Conjugal transfer pilus assembly periplasmic protein (pChi7122-2)	−	−
4G8	*trbF*	Conjugal transfer protein (pChi7122-2)	+	−
7H7	MM2_057-8	Intergenic on pChi7122-2, between TraM and transglycosylase SLT[b] domain protein	−	−
12E12	MM2_057-8	Intergenic on pChi7122-2, between TraM and transglycosylase SLT domain protein	−	−
Iron transport				
7B1	*entF*	Enterobactin synthase multienzyme complex component	−	−
10D9	*yddB*	Putative TonB-dependent receptor/porin for iron uptake	+	−
Membrane and periplasmic proteins				
10E4	*htrA* (*degP*)	Periplasmic serine protease mediating response to unfolded proteins	−	−
11D9	*hdeD*	Membrane protein associated with acid resistance	−	−
Sensing and regulation				
13H1	*envY*[a]	Thermoregulation of porin biosynthesis	−	−
2G8	*evgS*	Sensor kinase in two-component system with EvgA	−	−
15G8	*evgS*	Sensor kinase in two-component system with EvgA	−	−
16G8	*evgS*	Sensor kinase in two-component system with EvgA	−	−
15H9	5′ of *gatR*	Galactitol utilization operon repressor	−	−
13F9	*rspB*	Starvation sensing protein	−	−
17A4	*yahB*	Putative LysR family transcriptional regulator	−	−
2H10	*b2667* (*ygaV*)	Predicted transcriptional regulator	−	−
6H10	*b2667* (*ygaV*)	Predicted transcriptional regulator	−	−
Central intermediary metabolism				
13E2	*bglB*	6-Phospho-beta-glucosidase	−	−
2A3	*cadA*	Lysine decarboxylase	+	−
2G10	*fucP*	L-Fucose permease	−	−
8H10	*pykF*	Pyruvate kinase	−	−
6C2	*tdcC*	Threonine/serine transporter	−	−
2D9	*yjhH*[a]	Predicted dihydrodipicolinate synthase	−	−
Putative function				
23E9	*b0703*[a]	Putative RhsC family protein	+	−
1E8	*b0833* (*yliE*)	Putative cyclic di-GMP phosphodiesterase	−	−
7D1	*exoP*[c]	Putative exonuclease family protein	+	−
7G2	*exoP*[c]	Putative exonuclease family protein	−	−
7C11	*exoP*[c]	Putative exonuclease family protein	−	−
1F9	*intQ*[c]	Putative prophage-encoded site-specific recombinase	−	−
2F9	*intQ*[c]	Putative prophage-encoded site-specific recombinase	−	−

(Continued on following page)

TABLE 1 (Continued)

Mutant	Gene	Function	Selection in: Lung	Liver
6F2	*intQ*[c]	Putative prophage-encoded site-specific recombinase	−	−
13H11	MM1_0130	Putative IS*629* taxon	−	−
1E9	*ycbY*	Putative rRNA methyltransferase	−	−
2E9	*ycbY*	Putative rRNA methyltransferase	−	−
12F1	*ycbY*	Putative rRNA methyltransferase	−	−
16F12	*ycdR*	Putative deacetylase associated with biofilm formation	−	−
13H2	*ycjM*	Putative glycosyltransferase/polysaccharide hydrolase	−	−
14E6	*yejO*	Putative autosecreted adhesin/ATP-binding component of transport system	+	−
6D5	5′ of *yfgH*	Putative outer membrane lipoprotein (overlapping hypothetical CDS on reverse strand)	−	−
16H8	*yicH*	Putative AsmA-family protein	−	−
13A8	*yibJ*[a]	Putative Rhs family protein	−	−
13D9	*yjhA*	Putative N-acetylneuraminic acid porin	−	−
5B1	*yjiZ*[a]	Putative L-galactonate transporter	−	−
8B1	*yjiZ*[a]	Putative L-galactonate transporter	−	−
15C2	*ykgB*	Putative inner membrane protein	−	−
14E5	*b1409* (*ynbB*)	Putative CDP-diglyceride synthase	−	−
21E5	*b1409* (*ynbB*)	Putative CDP-diglyceride synthase	+	−
23E5	*b1409* (*ynbB*)	Putative CDP-diglyceride synthase	+	−
2H11	5′ of *χ7122_08431* (*yaiT*)	Putative autotransporter/flagellin-like protein	−	−
1F11	*orf04372*[a]	Putative recombinase	−	−
Hypothetical				
2A4	MM3_021-2	Intergenic on pChi7122-3; functions unknown	−	−
1G3	*orf00499*[a]	Function unknown	−	−
2E6	*yagX*	Function unknown	−	−
2G9	*ycgK*	Function unknown	−	−
23G9	*ycgK*	Function unknown	+	−
2E11	*orf01369*	Function unknown	−	−
21C1	5′ of *yfdQ*[d]	Function unknown	+	−
17F3	*yjhR*	Function unknown	−	+
15G5	*ykgI*	Function unknown	−	−

[a] Ortholog present in APEC O78:H9 strains χ7122 and IMT2125 but absent in APEC O1.
[b] SLT, Shiga-like toxin.
[c] Ortholog present in APEC O78:H9 strains χ7122 and APEC O1 but absent in IMT2125.
[d] Ortholog present in APEC O78:H9 strain χ7122 but absent in IMT2125and APEC O1.

phenotypes detected by STM (Table 1). The Δ*evgS*::kan^R, Δ*yccC*::kan^R, and Δ*fliC*::kan^R mutants were recovered in significantly lower numbers from the lung than the parent strain by 24 h postinoculation (Fig. 3A) ($P \leq 0.05$). In the liver, significantly lower numbers of the *entF*::Tn, *exoP*::Tn, Δ*fliC*::kan^R, Δ*htrA*::kan^R, *traE*::Tn, and Δ*yccC*::kan^R mutants than of the parent strain were obtained (Fig. 3B) ($P \leq 0.05$). All mutants except for the Δ*cadA*::kan^R and *yejO*::Tn mutants were recovered from the spleen in significantly lower numbers than the parent strain (Fig. 3C) ($P \leq 0.05$).

Role of the group 4 capsule. A defined Δ*yccC*::kan^R mutant, predicted to lack the Etk tyrosine kinase required for the synthesis of the group 4 (O-antigen) capsule (64), proved to be highly attenuated. The group 4 capsule has been proposed to mask surface structures in *E. coli* O157:H7 and influences the colonization of the large intestines of rabbits (80), consistent with the negative selection of a signature-tagged *ymcC* mutant of *E. coli* O157:H7 in the intestines of calves (60). To the best of our knowledge, our data provide the first evidence that the group 4 capsule is required during systemic *E. coli* infection. The absence of the *ymcDCBA-yccZ-etp-etk* genes from the APEC O1 strain highlights the variable repertoire of virulence-associated genes among APEC isolates.

Role of flagella. While flagella have been implicated in the intestinal colonization of 10-day-old chickens and systemic virulence in day-old chicks after oral dosing with APEC O78 (28), our data with a defined Δ*fliC*::kan^R mutant confirm that flagella play a significant role upon respiratory challenge. Further studies are required to determine whether this reflects their activity as an adhesin *per se*, as proposed for H6 and H7 flagella of attaching and effacing *E. coli* (81–83), their role in bacterial motility, or their ability to act as agonists of innate immunity.

Role of the EvgAS system. Attenuation of the Δ*evgS*::kan^R mutant is consistent with the negative selection of three χ7122 *evgS* transposon mutants here and an APEC O2 *evgA* mutant by STM in an early lung colonization model in chickens (21). The *evgAS* genes encode a two-component sensory system that controls the transcription of numerous genes in *E. coli*, including some associated with acid resistance (84, 85). Such transcriptome studies were conducted in *E. coli* K-12, and it is feasible that APEC-specific genes may also fall under the control of this system.

Role of iron acquisition genes. The *entF* and *yddB* genes are predicted to be involved in iron uptake. Transposon mutants were negatively selected in the liver and spleen to a greater extent than in the lung in the primary screen and when tested in isolation, as

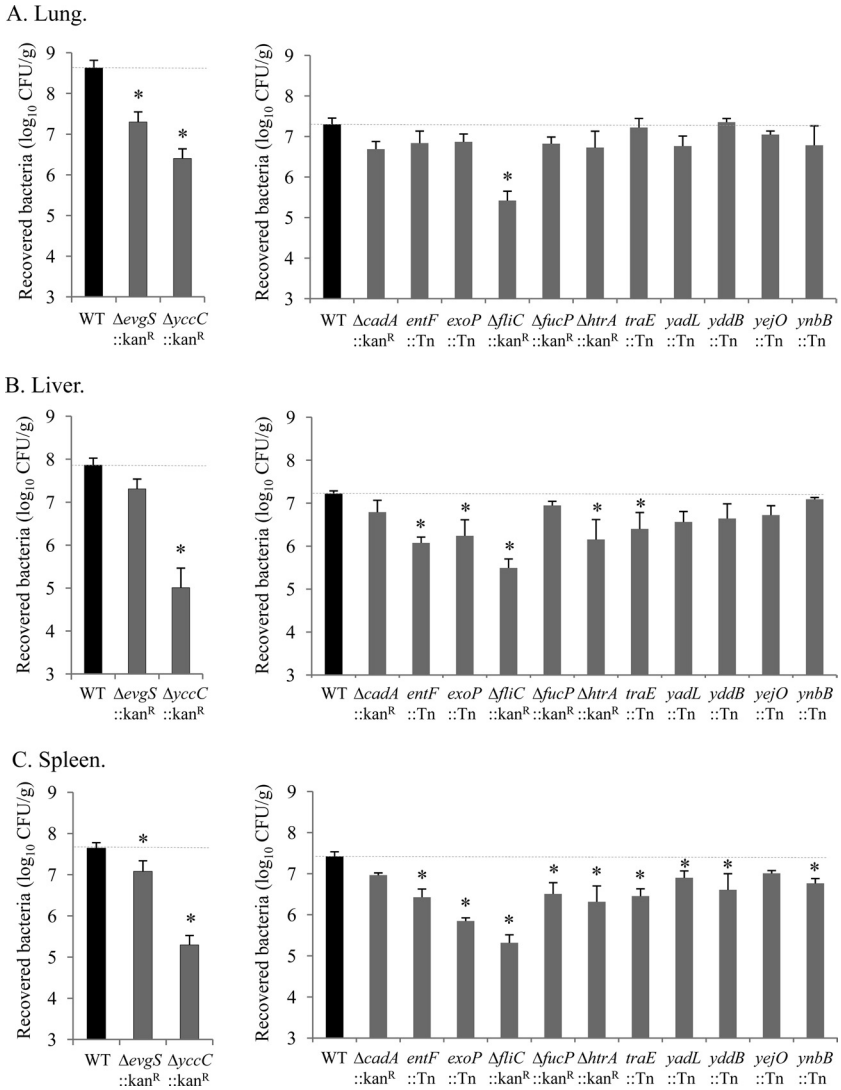

FIG 3 Recovery of APEC strain χ7122 mutants from the lung (A), liver (B), and spleen (C) 24 h after intra-air sac inoculation of turkeys. The recovery of the Δ*evgS*::kan^R and Δ*yccC*::kan^R mutants relative to that of the parent strain was tested in 4 birds per strain. In a separate study, other mutants were tested relative to the parent strain in three birds each. Asterisks indicate a significant difference ($P \leq 0.05$) between the concentration ($\log_{10}$ CFU/g) of the mutant strain recovered and that of the parent strain.

may be expected given the need to acquire iron during systemic spread and the upregulation of genes involved in the synthesis and transport of siderophores in an APEC O1 strain exposed to chicken serum (72). EntF forms part of multienzyme complex (EntDEF) that mediates amide linkage of 2,3-dihydroxybenzoic acid L-serine during the synthesis of enterobactin. A recent study indicated that enterobactin secretion, but not enterobactin synthesis *per se*, is essential for the full virulence of APEC O78 strain χ7122 in chickens: a Δ*entS* mutant was attenuated in an intra-air sac coinfection model (33). A χ7122 Δ*entD* mutant tested in that study was negatively selected in lung and liver tissue relative to a virulent Δ*lacI* strain, but not significantly so (33). YddB is a predicted TonB-dependent porin for iron uptake, although formal evidence for such a role is lacking.

Role of Yad fimbriae. The *yadL*::Tn mutant was recovered at levels at least 0.5 $\log_{10}$ CFU/g lower than those for the parent strain at each site sampled, although only the data for the spleen were significant. Taken together with the independent isolation of 5 mutants with defects in the *yadN-ecpD-htrE-yadMLKC* cluster by STM, the data suggest that Yad fimbriae play a subtle role in APEC pathogenesis. The *yad* cluster in APEC strain χ7122 is conserved in other sequenced APEC strains, *E. coli* K-12 MG1655, UPEC CFT703, and EHEC EDL933 and encodes fimbrial subunits and a chaperone-usher secretion system homologous to those involved in the production of *E. coli* type 1 fimbriae. Type I fimbriae are known to play an important role during the infection of chickens with strains of APEC serogroups O2 (27) and O78 (28). The predicted promoters of the *E. coli* K-12 *yad* operon and of the orthologous genes in *E. coli* O157:H7 (loc2; ECs0139-0145) have been reported to be inactive during growth *in vitro* (65, 86). However, forced expression of the *yad* operon in a Δ*fim* mutant of *E. coli* K-12 led to the production of long, flexible pili, sometimes in bundles, and increased biofilm formation at 30°C but not adherence to epithelial cells (65). The *yadN* gene is enriched in human

UPEC isolates relative to commensal *E. coli* strains, and Yad fimbriae are required for adherence to bladder epithelial cells and biofilm formation (87). They appear not to play a key role in bladder colonization in a murine urinary tract infection model but may act in concert with Ygi fimbriae *in vivo* (87).

Roles of other factors. Our data also support roles for the periplasmic serine protease HtrA (DegP), which mediates the response to stress and unfolded proteins, and *exoP*, which encodes a predicted prophage-encoded exonuclease family protein of ill-defined function. The significant attenuation of the *traE*::Tn mutant in the liver and spleen, but not in the lung, when tested in isolation is consistent with the phenotype of the mutant determined by STM screening. The data support a role for the pChi7122-1-encoded conjugation system, even though it is truncated relative to sequenced F plasmids (44). No significant attenuation could be detected for mutants lacking *cadA* (encoding a lysine decarboxylase that mediates stress responses in several enteric bacteria [reviewed in reference 88]) or *yejO* (a predicted autosecreted adhesin). Although both mutants were recovered in lower numbers than the parent strain at all three sites, mutants lacking FucP (L-fucose permease) or YnbB (a putative CDP-diglyceride synthase) were significantly attenuated only in the spleen.

Concluding remarks. Our analysis of the repertoire and function of APEC ST23 serogroup O78 genes in poultry reinforces the complexities of defining the APEC pathotype. Sequence typing and phylogenetic analysis of the core genomes of APEC indicate that *E. coli* strains causing the same disease of poultry arise from distinct lineages. Previous studies revealed that APEC ST95 O1 strains cluster with human ExPEC strains, prompting numerous investigations into the zoonotic potential of APEC. Our studies indicate that *E. coli* strains pathogenic for poultry are not uniformly related to ExPEC strains, since ST23 serogroup O78 strains clustered not with APEC O1 and UPEC strains but with a human enterotoxigenic *E. coli* strain of the same lineage. In turn, this challenges us to consider whether simple descriptors of *E. coli* strains as extraintestinal or diarrheagenic are supported by their phylogeny. We do not suggest that the similarity of the ST23 serogroup O78 strains studied is indicative of the potential for APEC strains from this lineage to cause human diarrheal disease, because they lack key enterotoxins. Rather, we infer that the core genome of the ST23 lineage can give rise to variants adapted to cause either avian or human disease via variation in the accessory genome. It is noteworthy that many of the attenuating mutations occurred in core chromosomal genes rather than in the variable plasmids, and saturating mutagenesis studies are needed to determine which features of "successful" backbones mediate *E. coli* fitness *in vivo*.

The evolutionary distance between APEC ST95 and ST23 is reinforced by the dramatic lack of conservation of virulence-associated genomic islands and the acquisition of distinct sets of virulence genes. This is further reflected in the lack of overlap between the virulence genes identified by STM in χ7122 and those identified in APEC ST95 by STM or targeted mutagenesis. Our analysis of signature-tagged and defined mutants supports roles for a number of factors found in the ST23 O78 strains (e.g., group 4 capsule, ETT2) but absent from the APEC ST95 O1 strain and suggests that some factors may exert their effects in a niche-specific manner.

This study represents the first high-resolution comparison of APEC genomes and provides valuable data for the rational design of strategies to control an important avian disease. It is clear from our work, and from preceding studies on the conservation of specific genes among field isolates, that extensive whole-genome analysis of strains from dominant APEC sequence types will be needed if we are to appreciate the full extent of diversity among these heterogenous pathogens. As the volume of sequence data for APEC expands, mutagenesis studies of the kind we report in target animals will be essential in order to determine which genes or combinations thereof are necessary for host adaptation and pathogenesis.

ACKNOWLEDGMENTS

We gratefully acknowledge the support of the Biotechnology and Biological Sciences Research Council (grants BB/E001661/1, BB/H001859/1, and BB/J004227/1), the British Poultry Council (Turkey Research and Development Sector), The Wellcome Trust (grants 0689 and 076964), the German Research Foundation (Indo-German Research Training Group, GRK1673), the U.S. Department of Agriculture (grant USDA 2007-35201-18519), and Aviagen Ltd.

REFERENCES

1. **Barnes HJ, Vaiillancourt J-P, Gross WB.** 2003. Colibacillosis, p 631–656. *In* Saif YM, Barnes HJ, Glisson JR, Fadly AM, McDougald LR (ed), Diseases of poultry, 11th ed. Iowa State University Press, Ames, IA.
2. **Pires-Dos-Santos T, Bisgaard M, Christensen H.** 2012. Genetic diversity and virulence profiles of *Escherichia coli* causing salpingitis and peritonitis in broiler breeders. Vet. Microbiol. **pii**:S0378–1135(12)00599-8. [Epub ahead of print.] doi:10.1016/j.vetmic.2012.11.008.
3. **Olsen RH, Chadfield MS, Christensen JP, Scheutz F, Christensen H, Bisgaard M.** 2011. Clonality and virulence traits of *Escherichia coli* associated with haemorrhagic septicaemia in turkeys. Avian Pathol. **40**:587–595.
4. **Johnson TJ, Kariyawasam S, Wannemuehler Y, Mangiamele P, Johnson SJ, Doetkott C, Skyberg JA, Lynne AM, Johnson JR, Nolan LK.** 2007. The genome sequence of avian pathogenic *Escherichia coli* strain O1:K1:H7 shares strong similarities with human extraintestinal pathogenic *E. coli* genomes. J. Bacteriol. **189**:3228–3236.
5. **Bauchart P, Germon P, Brée A, Oswald E, Hacker J, Dobrindt U.** 2010. Pathogenomic comparison of human extraintestinal and avian pathogenic *Escherichia coli*—search for factors involved in host specificity or zoonotic potential. Microb. Pathog. **49**:105–115.
6. **Ewers C, Antão EM, Diehl I, Philipp HC, Wieler LH.** 2009. Intestine and environment of the chicken as reservoirs for extraintestinal pathogenic *Escherichia coli* strains with zoonotic potential. Appl. Environ. Microbiol. **75**:184–192.
7. **Ewers C, Li G, Wilking H, Kiessling S, Alt K, Antáo EM, Laturnus C, Diehl I, Glodde S, Homeier T, Böhnke U, Steinrück H, Philipp HC, Wieler LH.** 2007. Avian pathogenic, uropathogenic, and newborn meningitis-causing *Escherichia coli*: how closely related are they? Int. J. Med. Microbiol. **297**:163–176.
8. **Johnson TJ, Wannemuehler Y, Johnson SJ, Stell AL, Doetkott C, Johnson JR, Kim KS, Spanjaard L, Nolan LK.** 2008. Comparison of extraintestinal pathogenic *Escherichia coli* strains from human and avian sources reveals a mixed subset representing potential zoonotic pathogens. Appl. Environ. Microbiol. **74**:7043–7050.
9. **Johnson TJ, Wannemuehler Y, Kariyawasam S, Johnson JR, Logue CM, Nolan LK.** 2012. Prevalence of avian pathogenic *Escherichia coli* strain APEC O1 genomic islands among extraintestinal and commensal *E. coli* isolates. J. Bacteriol. **194**:2846–2853.
10. **Kariyawasam S, Scaccianoce JA, Nolan LK.** 2007. Common and specific genomic sequences of avian and human extraintestinal pathogenic *Escherichia coli* as determined by genomic subtractive hybridization. BMC Microbiol. **7**:81. doi:10.1186/1471-2180-7-81.
11. **Mora A, López C, Dabhi G, Blanco M, Blanco JE, Alonso MP, Herrera A, Mamani R, Bonacorsi S, Moulin-Schouleur M, Blanco J.** 2009. Extraintestinal pathogenic *Escherichia coli* O1:K1:H7/NM from human and avian origin: detection of clonal groups B2 ST95 and D ST59 with different host distribution. BMC Microbiol. **9**:132. doi:10.1186/1471-2180-9-132.
12. **Moriel DG, Bertoldi I, Spagnuolo A, Marchi S, Rosini R, Nesta B,**

Pastorello I, Corea VA, Torricelli G, Cartocci E, Savino S, Scarselli M, Dobrindt U, Hacker J, Tettelin H, Tallon LJ, Sullivan S, Wieler LH, Ewers C, Pickard D, Dougan G, Fontana MR, Rappuoli R, Pizza M, Serino L. 2010. Identification of protective and broadly conserved vaccine antigens from the genome of extraintestinal pathogenic *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. **107**:9072–9077.

13. **Moulin-Schouleur M, Répérant M, Laurent S, Brée A, Mignon-Grasteau S, Germon P, Rasschaert D, Schouler C.** 2007. Extraintestinal pathogenic *Escherichia coli* strains of avian and human origin: link between phylogenetic relationships and common virulence patterns. J. Clin. Microbiol. **45**:3366–3376.
14. **Tivendale KA, Logue CM, Kariyawasam S, Jordan D, Hussein A, Li G, Wannemuehler Y, Nolan LK.** 2010. Avian pathogenic *Escherichia coli* strains are similar to neonatal meningitis *E. coli* strains and are able to cause meningitis in the rat model of human disease. Infect. Immun. **78**: 3412–3419.
15. **Li G, Laturnus C, Ewers C, Wieler LH.** 2005. Identification of genes required for avian *Escherichia coli* septicemia by signature-tagged mutagenesis. Infect. Immun. **73**:2818–2827.
16. **Mellata M, Dho-Moulin M, Dozois CM, Curtiss R, III, Brown PK, Arné P, Brée A, Desautels C, Fairbrother JM.** 2003. Role of virulence factors in resistance of avian pathogenic *Escherichia coli* to serum and in pathogenicity. Infect. Immun. **71**:536–540.
17. **Brown PK, Curtiss R, III.** 1996. Unique chromosomal regions associated with virulence of an avian pathogenic *Escherichia coli* strain. Proc. Natl. Acad. Sci. U. S. A. **93**:11149–11154.
18. **Sabri M, Caza M, Proulx J, Lymberopoulos MH, Brée A, Moulin-Schouleur M, Curtiss R, III, Dozois CM.** 2008. Contribution of the SitABCD, MntH, and FeoB metal transporters to the virulence of avian pathogenic *Escherichia coli* O78 strain χ7122. Infect. Immun. **76**:601–611.
19. **Li G, Ewers C, Laturnus C, Diehl I, Alt K, Dai J, Antão EM, Schnetz K, Wieler LH.** 2008. Characterization of a *yjjQ* mutant of avian pathogenic *Escherichia coli* (APEC). Microbiology **154**:1082–1093.
20. **Antão EM, Ewers C, Gürlebeck D, Preisinger R, Homeier T, Li G, Wieler LH.** 2009. Signature-tagged mutagenesis in a chicken infection model leads to the identification of a novel avian pathogenic *Escherichia coli* fimbrial adhesin. PLoS One **4**:e7796. doi:10.1371/journal.pone.0007796.
21. **Antão EM.** 2010. Identification of avian pathogenic *E. coli* (APEC) genes important for the colonization of the chicken lung and characterization of the novel ExPEC adhesin I. Ph. D. thesis. Humboldt-Universität zu Berlin, Berlin, Germany. http://edoc.hu-berlin.de/dissertationen/antao-esther-maria-2010-02-08/PDF/antao.pdf.
22. **Li G, Feng Y, Kariyawasam S, Tivendale KA, Wannemuehler Y, Zhou F, Logue CM, Miller CL, Nolan LK.** 2010. AatA is a novel autotransporter and virulence factor of avian pathogenic *Escherichia coli*. Infect. Immun. **78**:898–906.
23. **Provence DL, Curtiss R, III.** 1994. Isolation and characterization of a gene involved in hemagglutination by an avian pathogenic *Escherichia coli* strain. Infect. Immun. **62**:1369–1380.
24. **Dozois CM, Dho-Moulin M, Brée A, Fairbrother JM, Desautels C, Curtiss R, III.** 2000. Relationship between the Tsh autotransporter and pathogenicity of avian *Escherichia coli* and localization and analysis of the Tsh genetic region. Infect. Immun. **68**:4145–4154.
25. **Parreira VR, Gyles CL.** 2003. A novel pathogenicity island integrated adjacent to the *thrW* tRNA gene of avian pathogenic *Escherichia coli* encodes a vacuolating autotransporter toxin. Infect. Immun. **71**:5087–5096.
26. **Kariyawasam S, Nolan LK.** 2009. Pap mutant of avian pathogenic *Escherichia coli* O1, an O1:K1:H7 strain, is attenuated in vivo. Avian Dis. **53**: 255–260.
27. **Marc D, Arné P, Brée A, Dho-Moulin M.** 1998. Colonization ability and pathogenic properties of a *fim*$^-$ mutant of an avian strain of *Escherichia coli*. Res. Microbiol. **149**:473–485.
28. **La Ragione RM, Sayers AR, Woodward MJ.** 2000. The role of fimbriae and flagella in the colonization, invasion and persistence of *Escherichia coli* O78:K80 in the day-old-chick model. Epidemiol. Infect. **124**:351–363.
29. **Lymberopoulos MH, Houle S, Daigle F, Léveillé S, Brée A, Moulin-Schouleur M, Johnson JR, Dozois CM.** 2006. Characterization of Stg fimbriae from an avian pathogenic *Escherichia coli* O78:K80 strain and assessment of their contribution to colonization of the chicken respiratory tract. J. Bacteriol. **188**:6449–6459.
30. **Ideses D, Gophna U, Paitan Y, Chaudhuri RR, Pallen MJ, Ron EZ.** 2005. A degenerate type III secretion system from septicemic *Escherichia coli* contributes to pathogenesis. J. Bacteriol. **187**:8164–8171.
31. **Dozois CM, Daigle F, Curtiss R, III.** 2003. Identification of pathogen-specific and conserved genes expressed in vivo by an avian pathogenic *Escherichia coli* strain. Proc. Natl. Acad. Sci. U. S. A. **100**:247–252.
32. **Caza M, Lépine F, Milot S, Dozois CM.** 2008. Specific roles of the *iroBCDEN* genes in virulence of an avian pathogenic *Escherichia coli* O78 strain and in production of salmochelins. Infect. Immun. **76**:3539–3549.
33. **Caza M, Lépine F, Dozois CM.** 2011. Secretion, but not overall synthesis, of catecholate siderophores contributes to virulence of extraintestinal pathogenic *Escherichia coli*. Mol. Microbiol. **80**:266–282.
34. **Lamarche MG, Dozois CM, Daigle F, Caza M, Curtiss R, III, Dubreuil JD, Harel J.** 2005. Inactivation of the *pst* system reduces the virulence of an avian pathogenic *Escherichia coli* O78 strain. Infect. Immun. **73**:4138–4145.
35. **Germon P, Chen YH, He L, Blanco JE, Brée A, Schouler C, Huang SH, Moulin-Schouleur M.** 2005. *ibeA*, a virulence factor of avian pathogenic *Escherichia coli*. Microbiology **151**:1179–1186.
36. **Wang S, Shi Z, Xia Y, Li H, Kou Y, Bao Y, Dai J, Lu C.** 2012. IbeB is involved in the invasion and pathogenicity of avian pathogenic *Escherichia coli*. Vet. Microbiol. **159**:411–419.
37. **Bertrand N, Houle S, LeBihan G, Poirier É, Dozois CM, Harel J.** 2010. Increased Pho regulon activation correlates with decreased virulence of an avian pathogenic *Escherichia coli* O78 strain. Infect. Immun. **78**:5324–5331.
38. **Crépin S, Lamarche MG, Garneau P, Séguin J, Proulx J, Dozois CM, Harel J.** 2008. Genome-wide transcriptional response of an avian pathogenic *Escherichia coli* (APEC) *pst* mutant. BMC Genomics **9**:568. doi:10.1186/1471-2164-9-568.
39. **Cortes MA, Gibon J, Chanteloup NK, Moulin-Schouleur M, Gilot P, Germon P.** 2008. Inactivation of *ibeA* and *ibeT* results in decreased expression of type 1 fimbriae in extraintestinal pathogenic *Escherichia coli* strain BEN2908. Infect. Immun. **76**:4129–4136.
40. **Herren CD, Mitra A, Palaniyandi SK, Coleman A, Elankumaran S, Mukhopadhyay S.** 2006. The BarA-UvrY two-component system regulates virulence in avian pathogenic *Escherichia coli* O78:K80:H9. Infect. Immun. **74**:4900–4909.
41. **de Pace F, Nakazato G, Pacheco A, de Paiva JB, Sperandio V, da Silveira WD.** 2010. The type VI secretion system plays a role in type 1 fimbria expression and pathogenesis of an avian pathogenic *Escherichia coli* strain. Infect. Immun. **78**:4990–4998.
42. **Mellata M, Ameiss K, Mo H, Curtiss R, III.** 2010. Characterization of the contribution to virulence of three large plasmids of avian pathogenic *E. coli* χ7122 (O78:K80:H9). Infect. Immun. **78**:1528–1541.
43. **Mellata M, Maddux JT, Nam T, Thomson N, Hauser H, Stevens MP, Mukhopadhyay S, Sarker S, Crabbé A, Nickerson CA, Santander J, Curtiss R, III.** 2012. New insights into the bacterial fitness-associated mechanisms revealed by the characterization of large plasmids of an avian pathogenic *E. coli*. PLoS One **7**:e29481. doi:10.1371/journal.pone.0029481.
44. **Mellata M, Touchman JW, Curtiss R, III.** 2009. Full sequence and comparative analysis of the plasmid pAPEC-1 of avian pathogenic *E. coli* χ7122 (O78:K80:H9). PLoS One **4**:e4232. doi:10.1371/journal.pone.0004232.
45. **Tivendale KA, Noormohammadi AH, Allen JL, Browning GF.** 2009. The conserved portion of the putative virulence region contributes to virulence of avian pathogenic *Escherichia coli*. Microbiology **155**:450–460.
46. **Mellata M, Dho-Moulin M, Dozois CM, Curtiss R, III, Lehoux B, Fairbrother JM.** 2003. Role of avian pathogenic *Escherichia coli* virulence factors in bacterial interaction with chicken heterophils and macrophages. Infect. Immun. **71**:494–503.
47. **Provence DL, Curtiss R, III.** 1992. Role of *crl* in avian pathogenic *Escherichia coli*: a knockout mutation of *crl* does not affect hemagglutination activity, fibronectin binding, or curli production. Infect. Immun. **60**: 4460–4467.
48. **Simon R, Priefer U, Pühler A.** 1983. A broad host range mobilization system for *in vivo* genetic engineering: transposon mutagenesis in Gram negative bacteria. Nat. Biotechnol. **1**:784–791.
49. **Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR,**

Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature **437**:376–380.

50. **Chain PS, Grafham DV, Fulton RS, Fitzgerald MG, Hostetler J, Muzny D, Ali J, Birren B, Bruce DC, Buhay C, Cole JR, Ding Y, Dugan S, Field D, Garrity GM, Gibbs R, Graves T, Han CS, Harrison SH, Highlander S, Hugenholtz P, Khouri HM, Kodira CD, Kolker E, Kyrpides NC, Lang D, Lapidus A, Malfatti SA, Markowitz V, Metha T, Nelson KE, Parkhill J, Pitluck S, Qin X, Read TD, Schmutz J, Sozhamannan S, Sterk P, Strausberg RL, Sutton G, Thomson NR, Tiedje JM, Weinstock G, Wollam A, Genomic Standards Consortium Human Microbiome Project Jumpstart Consortium, Detter JC.** 2009. Genome project standards in a new era of sequencing. Science **326**:236–237.
51. **Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B.** 2000. Artemis: sequence visualization and annotation. Bioinformatics **16**:944–945.
52. **Berriman M, Rutherford K.** 2003. Viewing and annotating sequence data with Artemis. Brief. Bioinform. **4**:124–132.
53. **Riley M, Abe T, Arnaud MB, Berlyn MK, Blattner FR, Chaudhuri RR, Glasner JD, Horiuchi T, Keseler IM, Kosuge T, Mori H, Perna NT, Plunkett G, III, Rudd KE, Serres MH, Thomas GH, Thomson NR, Wishart D, Wanner BL.** 2006. *Escherichia coli* K-12: a cooperatively developed annotation snapshot—2005. Nucleic Acids Res. **34**:1–9.
54. **Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ.** 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11**:119. doi:10.1186/1471-2105-11-119.
55. **Li L, Stoeckert CJ, Jr, Roos DS.** 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. **13**:2178–2189.
56. **Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J.** 2005. ACT: the Artemis Comparison Tool. Bioinformatics **21**:3422–3423.
57. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. **18**:821–829.
58. **Edgar RC.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32**:1792–1797.
59. **Liu K, Linder CR, Warnow T.** 2011. RAxML and FastTree: comparing two methods for large-scale maximum likelihood phylogeny estimation. PLoS One **6**:e27731. doi:10.1371/journal.pone.0027731.
60. **Dziva F, van Diemen PM, Stevens MP, Smith AJ, Wallis TS.** 2004. Identification of *Escherichia coli* O157:H7 genes influencing colonization of the bovine gastrointestinal tract using signature-tagged mutagenesis. Microbiology **150**:3631–3645.
61. **Barrow PA.** 1983. A note on a method to aid sampling populations for characteristics. J. Appl. Bacteriol. **54**:311–312.
62. **Datsenko KA, Wanner BL.** 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc. Natl. Acad. Sci. U. S. A. **97**:6640–6645.
63. **Ren CP, Chaudhuri RR, Fivian A, Bailey CM, Antonio M, Barnes WM, Pallen MJ.** 2004. The ETT2 gene cluster, encoding a second type III secretion system from *Escherichia coli*, is present in the majority of strains but has undergone widespread mutational attrition. J. Bacteriol. **186**:3547–3560.
64. **Peleg A, Shifrin Y, Ilan O, Nadler-Yona C, Nov S, Koby S, Baruch K, Altuvia S, Elgrably-Weiss M, Abe CM, Knutton S, Saper MA, Rosenshine I.** 2005. Identification of an *Escherichia coli* operon required for formation of the O-antigen capsule. J. Bacteriol. **187**:5259–5266.
65. **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.
66. **Brzuszkiewicz E, Brüggemann H, Liesegang H, Emmerth M, Olschläger T, Nagy G, Albermann K, Wagner C, Buchrieser C, Emody L, Gottschalk G, Hacker J, Dobrindt U.** 2006. How to become a uropathogen: comparative genomic analysis of extraintestinal pathogenic *Escherichia coli* strains. Proc. Natl. Acad. Sci. U. S. A. **103**:12879–12884.
67. **Lynne AM, Skyberg JA, Logue CM, Nolan LK.** 2007. Detection of Iss and Bor on the surface of *Escherichia coli*. J. Appl. Microbiol. **102**:660–666.
68. **Binns MM, Davies DL, Hardy KG.** 1979. Cloned fragments of the plasmid ColV,I-K94 specifying virulence and serum resistance. Nature **279**:778–781.
69. **Barondess JJ, Beckwith J.** 1990. A bacterial virulence determinant encoded by lysogenic coliphage lambda. Nature **346**:871–874.
70. **Skyberg JA, Johnson TJ, Nolan LK.** 2008. Mutational and transcriptional analyses of an avian pathogenic *Escherichia coli* ColV plasmid. BMC Microbiol. **8**:24. doi:10.1186/1471-2180-8-24.
71. **Lynne AM, Skyberg JA, Logue CM, Doetkott C, Foley SL, Nolan LK.** 2007. Characterization of a series of transconjugant mutants of an avian pathogenic *Escherichia coli* isolate for resistance to serum complement. Avian Dis. **51**:771–776.
72. **Li G, Tivendale KA, Liu P, Feng Y, Wannemuehler Y, Cai W, Mangiamele P, Johnson TJ, Constantinidou C, Penn CW, Nolan LK, Payne SM.** 2011. Transcriptome analysis of avian pathogenic *Escherichia coli* O1 in chicken serum reveals adaptive responses to systemic infection. Infect. Immun. **79**:1951–1960.
73. **Badger JL, Wass CA, Kim KS.** 2000. Identification of *Escherichia coli* K1 genes contributing to human brain microvascular endothelial cell invasion by differential fluorescence induction. Mol. Microbiol. **36**:174–182.
74. **Badger JL, Wass CA, Weissman SJ, Kim KS.** 2000. Application of signature-tagged mutagenesis for identification of *Escherichia coli* K1 genes that contribute to invasion of human brain microvascular endothelial cells. Infect. Immun. **68**:5056–5061.
75. **Hill VT, Townsend SM, Arias RS, Jenabi JM, Gomez-Gonzalez I, Shimada H, Badger JL.** 2004. TraJ-dependent *Escherichia coli* K1 interactions with professional phagocytes are important for early systemic dissemination of infection in the neonatal rat. Infect. Immun. **72**:478–488.
76. **Martindale J, Stroud D, Moxon ER, Tang CM.** 2000. Genetic analysis of *Escherichia coli* K1 gastrointestinal colonization. Mol. Microbiol. **37**:1293–1305.
77. **Gonzalez MD, Lichtensteiger CA, Vimr ER.** 2001. Adaptation of signature-tagged mutagenesis to *Escherichia coli* K1 and the infant-rat model of invasive disease. FEMS Microbiol. Lett. **198**:125–128.
78. **Bahrani-Mougeot FK, Buckles EL, Lockatell CV, Hebel JR, Johnson DE, Tang CM, Donnenberg MS.** 2002. Type 1 fimbriae and extracellular polysaccharides are preeminent uropathogenic *Escherichia coli* virulence determinants in the murine urinary tract. Mol. Microbiol. **45**:1079–1093.
79. **Tuntufye HN, Lebeer S, Gwakisa PS, Goddeeris BM.** 2012. Identification of avian pathogenic *Escherichia coli* genes that are induced *in vivo* during infection in chickens. Appl. Environ. Microbiol. **78**:3343–3351.
80. **Shifrin Y, Peleg A, Ilan O, Nadler C, Kobi S, Baruch K, Yerushalmi G, Berdichevsky T, Altuvia S, Elgrably-Weiss M, Abe C, Knutton S, Sasakawa C, Ritchie JM, Waldor MK, Rosenshine I.** 2008. Transient shielding of intimin and the type III secretion system of enterohemorrhagic and enteropathogenic *Escherichia coli* by a group 4 capsule. J. Bacteriol. **190**:5063–5074.
81. **Girón JA, Torres AG, Freer E, Kaper JB.** 2002. The flagella of enteropathogenic *Escherichia coli* mediate adherence to epithelial cells. Mol. Microbiol. **44**:361–379.
82. **Erdem AL, Avelino F, Xicohtencatl-Cortes J, Girón JA.** 2007. Host protein binding and adhesive properties of H6 and H7 flagella of attaching and effacing *Escherichia coli*. J. Bacteriol. **189**:7426–7435.
83. **Mahajan A, Currie CG, Mackie S, Tree J, McAteer S, McKendrick I, McNeilly TN, Roe A, La Ragione RM, Woodward MJ, Gally DL, Smith DG.** 2009. An investigation of the expression and adhesin function of H7 flagella in the interaction of *Escherichia coli* O157:H7 with bovine intestinal epithelium. Cell. Microbiol. **11**:121–137.
84. **Masuda N, Church GM.** 2002. *Escherichia coli* gene expression responsive to levels of the response regulator EvgA. J. Bacteriol. **184**:6225–6234.
85. **Nishino K, Inazumi Y, Yamaguchi A.** 2003. Global analysis of gene regulation by EvgA of the two-component regulatory system of *Escherichia coli*. J. Bacteriol. **185**:2667–2672.
86. **Low AS, Holden N, Rosser T, Roe AJ, Constantinidou C, Hobman JL, Smith DG, Low JC, Gally DL.** 2006. Analysis of fimbrial gene clusters and their expression in enterohaemorrhagic *Escherichia coli* O157:H7. Environ. Microbiol. **8**:1033–1047.
87. **Spurbeck RR, Stapleton AE, Johnson JR, Walk ST, Hooton TM, Mobley HL.** 2011. Fimbrial profiles predict virulence of uropathogenic *Escherichia coli* strains: contribution of *ygi* and *yad* fimbriae. Infect. Immun. **79**:4753–4763.
88. **Torres AG.** 2009. The *cad* locus of *Enterobacteriaceae*: more than just lysine decarboxylation. Anaerobe **15**:1–6.