Identification of pathogen-specific and conserved genes expressed *in vivo* by an avian pathogenic *Escherichia coli* strain

Charles M. Dozois*, France Daigle†, and Roy Curtiss III‡

Department of Biology, Washington University, Campus Box 1137, One Brookings Drive, St. Louis, MO 63130

Contributed by Roy Curtiss III, November 11, 2002

***Escherichia coli* is a diverse bacterial species that comprises commensal nonpathogenic strains such as *E. coli* K-12 and pathogenic strains that cause a variety of diseases in different host species. Avian pathogenic *E. coli* strain χ7122 (O78:K80:H9) was used in a chicken infection model to identify bacterial genes that are expressed in infected tissues. By using the cDNA selection method of selective capture of transcribed sequences and enrichment for the isolation of pathogen-specific (non-*E. coli* K-12) transcripts, pathogen-specific cDNAs were identified. Pathogen-specific transcripts corresponded to putative adhesins, lipopolysaccharide core synthesis, iron-responsive, plasmid- and phage-encoded genes, and genes of unknown function. Specific deletion of the aerobactin siderophore system and *E. coli iro* locus, which were identified by selective capture of transcribed sequences, demonstrated that these pathogen-specific systems contribute to the virulence of strain χ7122. Consecutive blocking to enrich for selection of pathogen-specific genes did not completely eliminate the presence of transcripts that corresponded to sequences also present in *E. coli* K-12. These *E. coli* conserved genes are likely to be highly expressed *in vivo* and contribute to growth or virulence. Overall, the approach we have used simultaneously provided a means to identify novel pathogen-specific genes expressed *in vivo* and insight regarding the global gene expression and physiology of a pathogenic *E. coli* strain in a natural animal host during the infectious process.**

Escherichia coli is an adaptive bacterial species that is both a commensal resident of the intestine and a versatile pathogen of humans and other animals. Diseases associated with pathogenic *E. coli* include enteric infections caused by diarrheagenic *E. coli* and extraintestinal infections, such as urinary tract infections, meningitis, and septicemia caused by extraintestinal pathogenic *E. coli* (ExPEC) (1, 2). Specific *E. coli* pathotypes cause particular pathologies in different animal species. This host adaptation and virulence capacity are attributed to the horizontal acquisition of specific genes that are generally absent from the genomes of nonpathogenic *E. coli* strains, such as the laboratory strain *E. coli* K-12. Pathogen-specific genes are located on plasmids, bacteriophages, or discrete regions of DNA that have been termed pathogenicity islands (3, 4). Regions absent from *E. coli* K-12 that have not been confirmed to contribute to virulence have been termed unique sequence islands (3). Comparison of the complete genomes of *E. coli* K-12 strain MG1655 and enterohemorrhagic O157:H7 strain EDL933 demonstrate that both strains have genomes containing distinct genetic regions (4–6). The capacity of bacteria such as *E. coli* to acquire and delete DNA regions has permitted adaptation to new niches and the evolution of diverse virulence mechanisms (4, 7). The genomes of pathogenic *E. coli* are typically larger than those of K-12 strains (8), and pathogen-specific regions have been shown to contribute directly to virulence (9–11).

Although the presence of pathogen-specific genes may dictate the pathogenic lifestyle and virulence of particular bacteria, products encoded by conserved or "core" genes undoubtedly contribute to metabolism, physiology, and adaptation to environmental changes. In *E. coli*, ≈75–90% of the genome is conserved, as demonstrated by comparison of *E. coli* strains MG1655 (K-12) and EDL933 (O157:H7) (6). Some conserved gene products are essential for cell physiology or survival under conditions of stress that may include adaptation to host environments or resistance to host defenses. In other cases, conserved genes may encode regulators that control the expression of pathogen-specific virulence-associated genes. Examples of such include RpoS (12), Dam methylase in *Salmonella* and pathogenic *E. coli* (13, 14), and LuxS in enterohemorrhagic *E. coli* (15).

Avian pathogenic *E. coli* (APEC) cause extraintestinal infections, including respiratory infection (airsacculitis and pneumonia), pericarditis, perihepatitis, and septicemia of poultry (16). Predominant serotypes of *E. coli* associated with these infections are O1:K1, O2:K1, and O78:K80. APEC most likely enter and colonize the air sacs through inhalation of feces-contaminated dust (17). In certain cases, the bacteria spread systemically and cause fatal septicemia. Certain factors have been associated with the virulence of APEC, including the aerobactin iron-sequestering system, temperature-sensitive hemagglutinin (Tsh), Type 1 and P fimbriae, and Colicin-V plasmids (16, 18). Brown and Curtiss (11) used subtractive DNA hybridization to identify regions in the genome of APEC strain χ7122 (O78:H7:K80) that are absent from *E. coli* K-12. Twelve unique sequence islands (USIs) were identified in strain χ7122. By using P1 bacteriophage-mediated transduction, four of these USIs were replaced with the corresponding region of *E. coli* K-12. Two of these USIs, the 45.0-min region encoding the O78 antigen and an uncharacterized region located at 0 min, contributed to the virulence of APEC χ7122.

Direct screening of bacterial genes expressed during infection of the host is limited, because isolation of bacterial transcripts from host tissues necessitates separation from the abundance of host RNA. Recently, selective capture of transcribed sequences (SCOTS) has been used to identify bacterial genes expressed within macrophages (19–22). SCOTS allows the selective capture of bacterial cDNA derived from infected cells or tissues using hybridization to biotinylated bacterial genomic DNA. In this report, we have used SCOTS to preferentially isolate pathogen-specific (non-*E. coli* K-12) cDNAs representing APEC genes that are expressed within the tissues of experimentally infected poultry. Consecutive blocking to enrich for selection of APEC-specific genes did not completely eliminate the presence of transcripts that correspond to sequences also present in *E. coli* K-12. These *E. coli* conserved genes are likely to be highly expressed in host tissues and contribute to bacterial growth or virulence.

Abbreviations: APEC, avian pathogenic *Escherichia coli*; ExPEC, extraintestinal pathogenic *E. coli*; SCOTS, selective capture of transcribed sequences; Tsh, temperature-sensitive hemagglutinin.

Data deposition: The sequence reported in this paper has been deposited in the GenBank database (accession no. AF449498).

*Present address: Institut National de la Recherche Scientifique–Institut Armand-Frappier, 531 Boulevard des Prairies, Laval, Québec, Canada H7V 1B7.

†Present address: Département de Microbiologie et Immunologie, Université de Montréal, CP 6128, Montréal, Québec, Canada H3C 3J7.

‡To whom correspondence should be addressed. E-mail: rcurtiss@biology.wustl.edu.

MICROBIOLOGY

Table 1. *E. coli* strains and plasmids

Strain or plasmid	Genotype and phenotype	Source
Strains		
χ289	*E. coli* K-12 W1485 λ⁻, F⁻ *glnV44*	Ref. 11
χ7122	Avian pathogenic, O78:K80:H9, *gyrA*, Nal^R	Ref. 23
χ7273	*tsh*::*tetAR*(B), Tc^R (Tsh⁻)	χ7122; ref. 18
χ7301	Δ*iucABCD iutA*::*xylE tsh*::*tetAR*(B) Tc^R (Aerobactin⁻, Tsh⁻)	χ7273
χ7303	Δ*iroBCDEN*::*nptII*, Km^R (Iro⁻)	χ7122
χ7306	Δ*iroBCDEN*::*nptII tsh*::*tetAR*(B) Δ*iucABCDiutA*::*xylE* Tc^R, Km^R (Tsh⁻, Iro⁻, Aerobactin⁻)	χ7301
Plasmids		
pACYC184	p15A replicon, Tc^R, Cm^R	Ref. 24
pBSL86	*nptII* cassette vector. Km^R, Ap^R	Ref. 25
pC6	*E. coli rrnB* in pBR322, Ap^R	C. Squires
pMEG-375	*sacRB mobRP4 ori*R6K. Cm^R, Ap^R	Megan Health (St. Louis, MO)
pMEG-685	*xylE* cassette vector, Ap^R	Megan Health
pYA3661	pACYC184::*iroBCDEN*, Cm^R	This work
pYA3662	pMEG-375::Δ*iucABCDiutA*::*xylE*. Cm^R, Ap^R	This work
pYA3663	pMEG-375::Δ*iroBCDEN*::*nptII*. Cm^R, Ap^R, Km^R	This work

Ap, ampicillin; Cm, chloramphenicol; Km, kanamycin; Nal, nalidixic acid; Tc, tetracycline.

Materials and Methods

Bacterial Strains, Plasmids, and Culture Conditions. Bacterial strains and plasmids are presented in Table 1. The virulent APEC strain χ7122 (O78:H9:K80) (23) was used for determination of *in vivo* expressed genes, infection studies, and mutant construction. Cells were routinely grown in Lennox broth (26) at 37°C. Solid medium contained 1.5% agar. For infection experiments, strains were grown for 24 h in beef heart infusion broth (Difco) at 37°C. Antibiotics, when required, were used at the following final concentrations (μg/ml): 10 tetracycline, 30 chloramphenicol, 25 kanamycin, and 12.5 nalidixic acid.

General Molecular Techniques. Bacterial genomic DNA was prepared by using a small-scale method (27). Restriction endonucleases and DNA-modifying and ligase enzymes (New England Biolabs; Promega) were used according to the manufacturers' guidelines. Standard techniques were used for bacterial conjugation and transformation (28).

Experimental Infection, RNA Isolation, cDNA Synthesis, and Amplification. Three-week-old white leghorn specific-pathogen-free chickens (Charles River Breeding Laboratories) were inoculated into the right thoracic air sac with a 0.1-ml suspension containing 1×10^8 colony-forming units of strain χ7122. Total RNA was isolated from infected tissues (pericardium and air sacs) 6 or 24 h postinfection by using TRIzol reagent (Invitrogen). RNA samples were treated with RNase-free DNase I (Ambion, Austin, TX). RNA concentrations and integrity were determined by A260/A280 spectrophotometer readings and agarose gel electrophoresis, respectively. Five 1-μg samples of total RNA obtained from tissues of five individual chickens were pooled and converted to first-strand cDNAs by random priming with Superscript II reverse transcriptase (Invitrogen). Random priming was performed as described (29) by using oligonucleotides containing terminal sequences at the 5′ end (5′-CGGGATCCAGCTTCTGACGCA-3′ for cDNAs from the air sacs and 5′-GTGGTACCGCTCTCCGTCCGA-3′ for cDNAs from the pericardium) and a random nonamer at the 3′ end (PCR primer-dN9). cDNAs were made double-stranded with Klenow fragment (NEB, Beverly, MA) as described (29). cDNA was then amplified by PCR by using the defined primers for each set of cDNA for 25 cycles.

SCOTS. Bacterial cDNA capture was done as described previously (19, 20). Briefly, denatured, biotinylated, and sonicated *E. coli* (χ7122) genomic DNA fragments (0.3 μg) were mixed with sonicated *rrnB* DNA (pC6) (5 μg) to preblock ribosomal RNA-encoding regions of the genomic DNA. After hybridization at 67°C for 30 min, denatured amplified cDNA (3 μg) from infected tissues was added, and hybridized for 18 h at 67°C. Bacterial cDNAs that hybridized to biotinylated genomic DNA were retained by binding hybrids to streptavidin-coated M280 magnetic beads (Dynal, Bethlehem, PA). Captured cDNA was then eluted, precipitated, and amplified by PCR. In the first round of SCOTS, five separate samples of cDNA were captured by hybridization to biotinylated rDNA-blocked chromosomal DNA in parallel reactions. This was done to enhance the likelihood of recovering cDNA molecules corresponding to a more complete diversity of bacterial transcripts present in host tissues. After the first round of SCOTS, the five amplified cDNA preparations for each type of tissue sampled were combined, denatured, and again hybridized to rDNA-blocked biotinylated chromosomal DNA for two successive rounds of SCOTS. The mixtures were then used to verify successful capture of *E. coli* sequences by cloning and for competitive hybridization enrichment as described below.

Enrichment for Pathogen-Specific Bacterial cDNAs. To enrich the bacterial cDNA populations for APEC-specific (non-*E. coli* K-12) sequences, *E. coli* χ7122 biotinylated genomic DNA (0.3 μg) was first preblocked with an excess of denatured sheared genomic DNA (10 μg) from *E. coli* K-12 strain χ289. Bacterial cDNAs that hybridized to biotinylated genomic cDNA were retained by binding to streptavidin-coated magnetic beads (Dynal). Bacterial cDNAs were eluted, precipitated, and PCR amplified by using specific terminal sequences. After three rounds of competitive hybridization, bacterial cDNAs were cloned by using the Original TA Cloning kit (Invitrogen). Cloned inserts were sequenced by using the ABI Prism Big Dye primer cycle sequencing kit (PE Applied Biosystems). Sequences were compared with the complete genome of *E. coli* K-12 (5), and database comparisons were carried out with the BLAST algorithm (30).

Construction of Defined Mutations in Pathogen-Specific Genes of APEC Strain χ7122. Mutants of strain χ7122 were generated by deletion of pathogen-specific genes that were identified by SCOTS. Generation of strain χ7273 by inactivation of *tsh*, which encodes the Tsh/hemoglobin protease is described elsewhere (18). A suicide vector for deletion of the aerobactin encoding gene cluster (*iucABCDiutA*) was constructed as follows. A 833-bp fragment of the 5′ end of *iucA* was generated by PCR by using primers aero1 (5′-gctctagattatgatcctgccctctg-3′; added *Xba*I site underlined) and aero2 (5′-ttgcggccgctggtagcacagtagagg-3′; added *Not*I site underlined), and a 1,200-bp fragment of the 3′ end of *iutA* was generated by PCR by using primers aero3 (5′-ttgcggccgcactgacgggcatttga-3′; added *Not*I site underlined) and aero4 (5′- atgcatgctgaagctgagtgtacc-3′; added *Sph*I site underlined). These two fragments and a *Not*I-*Not*I *xylE* gene cassette from pMEG685 were cloned into the *Xba*I and *Sph*I sites of pMEG-375. A resultant suicide vector containing the *iucA′-xylE-′iutA* fragment was named pYA3662. A suicide vector for deletion of the *iroBCDEN* genes was constructed as follows. A 663-bp fragment of the 5′ end of *iroB* was generated by PCR by using primers iroBKO1 (5′-aggcgcgcctctctatgggc-3′; added *Asc*I site underlined) and iroBKO2 (5′-ctctagatcaaggccgtcaacc-3′; added *Xba*I site underlined), and a 609-bp fragment of the 5′ end of *iroN* was generated by PCR by using primers iroNKO1 (5′-aagcatgctcctggttgggttgaata-3′; added *Sph*I site underlined) and iroNKO2 (5′-ctctagagcattaccagccagagg-3′; added *Xba*I site underlined). These two fragments and a *Xba*I-*Xba*I *npt II* gene cassette from pBSL86 were cloned into the *AscI* and *Sph*I sites of pMEG-375. A resultant suicide vector containing the *iroB′-npt II-iroN′*

Table 2. *E. coli* pathogen-specific *ecs* clones identified by SCOTS

Functional group and locus	Homolog, source*† (% identity/% similarity)	Accession no. of *ecs* homolog	Function	Tissue‡
Adherence				
ecs-1	*pilN*, R721 (96)	AP002527	R721 pilus assembly	A
ecs-2	*pilQ*, R721 (98)	AP002527	R721 pilus assembly	A
ecs-3	Z5222, EcO157 (66/81)	AAG58927	Putative fimbrial usher	P
ecs-4	TcfD, St (52/67)	CAB51577	Putative fimbrial subunit	P
ecs-5	*tsh*, pAPEC-1 (99)	AF218073	Hemagglutinin	A
	hbp, ColV (99)	AJ223631	Hemoglobin protease	
LPS synthesis				
ecs-6	*waaO*, EcF470 (96)	AF019746	LPS core biosynthesis	A
ecs-7	*waaY*, EcF470 (94)	AF019746	LPS core biosynthesis	A
Iron-responsive				
ecs-8	*iutA* ColV (96)	X05874	Aerobactin/cloacin	P
ecs-9	*iutA* ColV (93)		DF13 receptor	
ecs-10	*iucA* ColV (98)	X76100	Aerobactin synthesis	P
ecs-11	*iucD* ColV (92)	U90207	Aerobactin synthesis	A
ecs-12	*iroC* Ec536 (97)	X16664	Putative ABC transporter	A
Plasmid function				
ecs-13 to *ecs-28*	ColE2-type plasmids (87–98)	D30059	Replication, relaxation	A, P
ecs-29	*traK* F (92)	U01159	F-plasmid transfer	A
ecs-30	*traG* F (90)	U01159	F-plasmid transfer	A
ecs-31	*traT* F (97)	U01159	Exclusion, serum resistance	A
ecs-32	SopA F (78/81)	P08866	F-plasmid partitioning	A
ecs-33	*psiA* pO157 (90)	AF074613	Inhibition of SOS response	A
Phage-related				
ecs-34	*hkaG* (96), ϕHK620	AF335538	Unknown function	P
ecs-35	*hkbV* (97), ϕHK620	AF335538	DNA stabilization	A
ecs-36	HkbQ (43/63), ϕHK620	AAK28892	Portal protein	A
ecs-37	HkbQ (61/74), ϕHK620	AAK28892	Portal protein	A
ecs-38	Z3370(51/56), ϕCP-933V	AAG57253	Unknown function	P
ecs-39	Int (46/57), ϕP2	P36932	Integrase	P
Other				
ecs-40	CC0532 Cc (62/77)	AAK22519	Putative mandelate racemase	A
ecs-41	TM0427 Tm (41/58)	C72379	Putative oxidoreductase	P
ecs-42	YPO3000 Yp (39/62)	CAC92244	Putative ABC transporter	A
ecs-43	*rhsH* Ec45 (97)	AF044501	Rearrangement hot spot element	P
ecs-44	RSp0733 Rs (46/52)	NP_522294	Hypothetical protein	A

Details regarding *ecs* homologs are published as supporting information on the PNAS web site.

*% nucleic acid identity for gene homologs (italicized); % identity/similarity for translated products (raw type).

†Bacterial species: Cc, *Caulobacter crescentus*; Ec, *E. coli* strains EDL933:EcO157, F470:Ec470, 536:Ec536, and ec45:Ec45; Rs, *Ralstonia solanacearum*; St, *Salmonella enterica* serovar Typhi; Tm, *Thermotoga maritima*; Yp, *Yersinia pestis.* Plasmids: R721, pAPEC-1, ColV-K30, F, pO157. Enterobacterial bacteriophages (ϕ): HK620, CP-933V, and P2.

‡Clone derived from RNA isolated in air sacs (A) or pericardium (P).

fragment was named pYA3663. Allelic replacements were obtained as described elsewhere (18).

Virulence Studies. Three-week-old white leghorn specific-pathogen-free chickens were inoculated into the right thoracic air sac with a 0.1-ml suspension containing 1×10^7 colony-forming units of strain χ7122 or mutant derivatives. Birds were euthanized 48 h post-infection, and bacterial quantification in tissues was performed as described previously (18).

Results

Selective Capture of APEC Transcripts Expressed in Tissues of Infected Chickens. After infection, pericardial and air sac tissues were used to isolate RNA and to quantitate bacteria. Bacterial counts varied from 10^4 to 10^7 colony-forming units/gram of tissue 6 or 24 h postinfection. We first isolated total *E. coli* transcripts from infected tissues of chickens by using SCOTS. This sampling of bacterial cDNAs represented total transcripts produced by bacteria within the tissues of infected chickens. Initial screening of some of the cDNA sequences isolated from these pools demonstrated that the majority represented sequences present in *E. coli* K-12. Identified sequences corresponded to *fhuF*, *ompX*, *mdh*, *ybiL* (*fiu*), *yegE*, *yhdG*, *yfhA*, *yghU*, and the IS2 transposase gene. We therefore did not pursue further identification of sequences from these cDNA pools. These results confirmed, however, that by using SCOTS, *E. coli*-specific transcripts were successfully isolated from tissues of infected chickens.

Enrichment for Identification of Pathogen-Specific Sequences Expressed *in Vivo*. To identify transcripts that were specific to APEC strain χ7122, we subjected the total bacterial cDNA pools to three further rounds of SCOTS in the presence of a >33-fold excess of unlabeled *E. coli* K-12 DNA. An abundance of unlabeled *E. coli* K-12 DNA was expected to reduce the capture of sequences common to APEC χ7122 and *E. coli* K-12 and promote the capture of APEC χ7122-specific transcripts. After this blocking step, cDNAs were cloned. Randomly chosen cDNA clones derived from transcripts in the air sacs or pericardium were selected and sequenced. The clones, termed *E. coli* captured sequences (*ecs*), contained numerous APEC-specific sequences (Table 2). Despite three rounds of blocking to reduce the presence of sequences common to *E. coli* K-12 and APEC strain χ7122, one-third of the *ecs* clones contained fragments of genes corresponding to *E. coli* K-12 sequences (Table 3). Among the 66 distinct *ecs* clones, only one clone was identified twice from sequenced samples. Forty-four of the 66 distinct clones were APEC-specific. Pathogen-specific

Table 3. *E. coli* conserved genes expressed *in vivo* identified by SCOTS

Functional group and locus	Gene (% identity: *E. coli* K-12 MG1655)	Function	Tissue*
Metabolism			
ecs-45	*aceF*† (96)	Pyruvate dehydrogenase component Acetate requirement	P
ecs-46	*folP** (90)	Dihydropteroate synthase	A, P
ecs-47	*folP** (96)	Folic acid, vitamin auxotroph	
ecs-48	***alaS***‡ (97)	Alanyl-tRNA synthetase	P
ecs-49	*purA** (97)	Adenylosuccinate synthetase required for nucleic acid synthesis	A
ecs-50	***pyrG**** (97)	CTP synthetase; cytidine auxotroph	A
ecs-51	***pyrG**** (100)		
ecs-52	***kdsA*** (93)	KDO synthesis, LPS structure	A
ecs-53	*thrA* (97)	Amino acid synthesis	A
ecs-54	*napC-ccmA* (100)	Cytochrome c maturation	P
ecs-55	***eno**** (97)	Enolase	A
Cell growth			
ecs-56	***ecfE (yaeL)*** (95)	σ^H and σ^E specific protease	A
ecs-57	***yhhP (sirA)*** (92)	Normal cell growth, σ^S stability	A
ecs-58	***ftsI*** (100)	Cell division	P
ecs-59	***ftsN*** (97)	Cell division	A
Regulation and response			
ecs-60	*phoB* (92)	Phosphate response regulator	P
ecs-61	***topA*** (99)	DNA topoisomerase I	A
ecs-62	*uspB* (98)	Stringent response, ethanol resistance	A
Unknown/ putative			
ecs-63	*ydgD* (95)	Putative protease	P
ecs-64	*yeaN* (99)	Putative transporter	A
ecs-65	*yfiO* (98)	Putative lipoprotein	A
ecs-66	*ynfK* (98)	Putative dethiobiotin synthase	A

Details regarding these *E. coli* genes are published as supporting information on the PNAS web site. KDO, 3-deoxy-D-manno-octulosonic acid.
*Clone derived from RNA isolated in air sacs (A) or pericardium (P).
†Asterisks indicate genes in which null mutations result in auxotrophies in *E. coli* K-12. See ref. 31 and supporting information on the PNAS web site for details.
‡Genes (indicated in bold type) are essential for normal cell growth and viability.

clones contained sequences homologous to known and novel putative bacterial gene products involved in adherence, iron transport, lipopolysaccharide (LPS) synthesis, plasmid replication and conjugation, putative phage encoded products, and gene products of unknown function (Table 2). More detailed information regarding homologs of the pathogen-specific *ecs* clones and *ecs* clones corresponding to *E. coli* conserved genes is published as supporting information on the PNAS web site (www.pnas.org).

Role of Pathogen-Specific Genes Encoded by Virulence Plasmid pAPEC-1. Plasmid pAPEC-1, a Colicin V-type plasmid, is required for the virulence of strain χ7122 in chickens (18). Genomic analyses demonstrated that the genes encoding Tsh, aerobactin, and the Iro system, which were all identified by SCOTS, are encoded by plasmid pAPEC-1. These three systems may contribute to acquisition of iron by APEC strain χ7122, because aerobactin is a pathogen-associated iron sequestering system, Tsh is a hemoglobin protease, and the uncharacterized Iro system encodes a putative siderophore receptor. To examine the role of these *in vivo* expressed pathogen-specific systems, mutant derivatives of strain χ7122 were generated and tested in the avian infection model. The *iro* locus was cloned from *Hin*dIII-digested fragments of plasmid pAPEC-1 and was sequenced (GenBank accession no. AF449498). The aerobactin, Iro, and/or Tsh encoding genes were deleted or inactivated in derivatives of pathogenic strain χ7122 and were tested for virulence in chickens. The *iro* mutant was tested to determine the individual role of this system for virulence and persistence in chickens compared with a mutant lacking Tsh and aerobactin, the other two known pathogen-specific systems implicated in APEC virulence, and a triple mutant (Iro$^-$, Tsh$^-$, Aerobactin$^-$) lacking all three systems. Bacterial numbers were significantly reduced in the lungs, livers, and spleens of chickens infected with either the Aerobactin$^-$, Tsh$^-$ mutant (χ7301) or the Iro$^-$ mutant (χ7303) compared with the wild-type parent strain (χ7122) 48 h after infection (Fig. 1). In addition, strains χ7301 and χ7303 exhibited reduced lesions of airsacculitis and only mild pericarditis and perihepatitis compared with the wild-type parent. Although χ7301 is also Tsh$^-$, Tsh alone is not required for persistence of strain χ7122 in deeper tissues of chickens (18). As well, an Aerobactin$^-$ derivative of χ7122 persists in deeper tissues at a level similar to that of the Tsh$^-$ Aerobactin$^-$ strain χ7301 (C.M.D. and R.C., unpublished data). The Tsh$^-$, Iro$^-$, and Aerobactin$^-$ mutant (χ7306) did not survive in extraintestinal tissues 48 h postinfection (Fig. 1) and caused minor lesions that were limited to the air sac site of inoculation. To confirm the role of the *iro* locus for APEC virulence, plasmid pYA3661 containing the *iroBCDEN* genes was introduced into strains χ7303 (Iro$^-$) and χ7306 (Iro$^-$, Tsh$^-$, Aerobactin$^-$). Strains χ7303 and χ7306 containing pYA3661 persisted in the liver (Fig. 1*C*) and caused lesions of pericarditis and perihepatitis to the same extent as the wild-type APEC strain χ7122.

Identification of *in Vivo* Expressed *E. coli* Genes Common to APEC and K-12 Strains. Twenty-two of the *ecs* clones identified after enrichment for isolation of pathogen-specific sequences represented transcripts that correspond to genes present in both APEC strain χ7122 and *E. coli* K-12 (Table 3). These *ecs* clones corresponded to genes required for metabolism, cell growth, stress response, or regulation, and genes of unknown function. Because these conserved transcripts remained after pathogen-specific cDNA enrichment, it is likely that they are abundantly expressed in infected tissues and were therefore not as effectively eliminated from the cDNA pools. This indicates that in addition to identifying pathogenic *E. coli*-specific sequences, the enrichment procedure favored the isolation of *E. coli* conserved genes that are likely to be highly expressed *in vivo*. Therefore, it is likely that the *E. coli* conserved genes identified by SCOTS after enrichment for pathogen-specific *E. coli* genes represent important metabolic or regulatory systems expressed *in vivo*, and many of these genes are also likely to be essential for bacterial growth or for full virulence.

Discussion

Determination of bacterial genes that are expressed during infection or that are essential or required for virulence in the host may provide valuable information to prevent and control infectious diseases. In recent years, identification of *in vivo* expressed or *in vivo* required bacterial genes has been achieved by using a number of different approaches. The most commonly used methods have included signature-tagged mutagenesis (STM), a negative-selection method that involves comparative isolation of individual specifically tagged transposon-generated mutants from pools of mutants propagated *in vitro* and *in vivo* (32), and promoter fusion methods such as *in vivo* expression technology (IVET) (33). IVET positively selects for *in vivo* promoter activity that provides transcription of a gene product required for survival *in vivo*. In addition, differential fluorescence induction (DFI), a promoter trap-*gfp* (green fluorescent protein) fusion-based differential selection method, identifies bacterial genes induced within host cells by determining comparative levels of fluorescence by flow cytometry (34). Recently, a

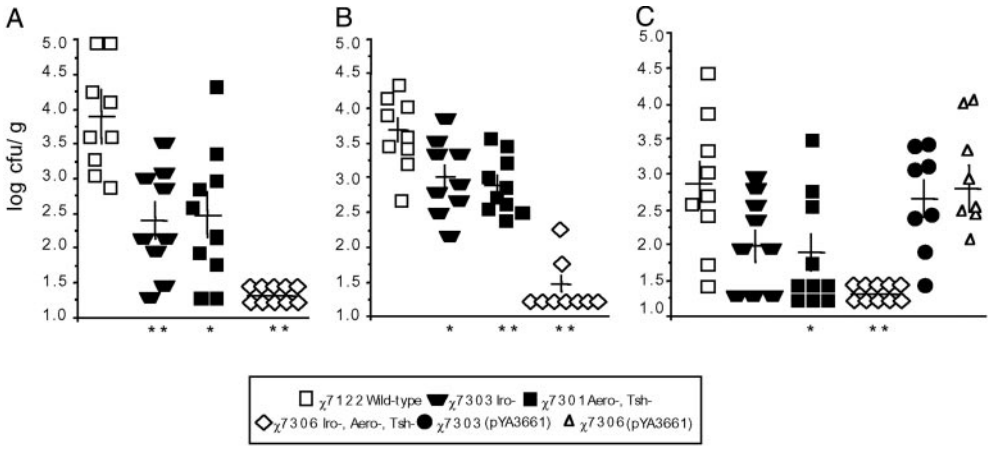

Fig. 1. Bacterial numbers present in the lungs (*A*), spleens (*B*), and livers (*C*) of chickens infected with wild-type APEC strain χ7122 or isogenic mutant derivatives. Data points represent bacterial counts from tissues isolated from different chickens ($n = 8–10$) 48 h postinfection. Horizontal bars represent the mean bacterial colony-forming units. Vertical bars represent the standard errors of the means. Statistical differences compared with the wild-type strain are noted: *, ($P < 0.05$); **, ($P < 0.01$).

bacterial transcript analysis method termed SCOTS has been used to identify genes expressed by *Mycobacterium* spp. (19, 22) and *S. enterica* in phagocytes (20, 21). In this report, we have extended the use of SCOTS to determine *in vivo* expressed bacterial genes of a pathogenic *E. coli* strain within a natural host experimental animal model. With the aim of discovering pathogen-specific genes that are expressed *in vivo*, we directly enriched the pool of *in vivo* expressed genes to contain pathogen-specific (non-*E. coli* K-12) gene transcripts and demonstrated by generation of defined-deletion mutants that the pathogen-specific aerobactin-encoding and *iro* gene clusters were required for full virulence. Notably, many of the APEC-specific and conserved *E. coli* genes or their homologs that were identified by SCOTS were identified by STM, *in vivo* expression technology, and/or DFI in different bacterial pathogens (see supporting information on the PNAS web site).

APEC-specific transcripts isolated from host tissues exhibited homologies to genes involved in adherence, iron transport, LPS synthesis, plasmid replication and conjugation, putative phage-encoded products, and gene products of unknown function (Table 2). Many of these genes represent newly identified sequences, whereas others have been previously associated with virulence in *E. coli*. Transcript *ecs-5* isolated from the air sacs corresponded to *tsh* (*hbp*). Tsh (Hbp), a Tsh/hemoglobin protease, was the first enterobacterial serine-protease autotransporter to be identified (35) and is associated with virulent avian *E. coli* (18) and human ExPEC (36). Identification of the expression of *tsh* in the air sacs correlates with our infection studies that demonstrated a role for Tsh in the development of lesions in the air sacs (18).

SCOTS clones *ecs-6* and *ecs-7* corresponded to genes involved in the synthesis of the R1-type core LPS, and *ecs-52* corresponded to *kdsA*, a conserved gene that is required for synthesis of 3-deoxy-D-manno-octulosonic acid (KDO), an essential component of the LPS inner core. LPS is of major importance to pathogenic bacteria, because it can contribute to resistance to complement and phagocytic cells. The O-antigen of strain χ7122 was previously shown to be required for virulence (11). The identification of SCOTS clones corresponding to iron-regulated pathogen-specific genes (*ecs-8* to *ecs-12*) is in line with the fact that in host tissues iron availability is limited. Clones corresponding to genes of the aerobactin and Iro systems were identified. Deletion of the *iro* gene cluster and/or aerobactin encoding systems clearly demonstrated their importance for persistence and generation of lesions in deeper tissues by strain χ7122 (Fig. 1). Aerobactin is a siderophore associated with ExPEC from humans and is also associated with virulent APEC strains. Aerobactin has also been demonstrated to contribute to the virulence of extraintestinal *E. coli* strains in the mouse urinary tract infection (37) and ovine oral infection models (38).

The *E. coli iroBCDEN* gene cluster encodes putative proteins homologous to those encoded by the *iro* locus of *S. enterica* (39). In *S. enterica,* the *iro* locus is derepressed under iron-limiting conditions (39) and in ExPEC strain CP9, *iroN* expression was shown to be repressed by iron (40). Previous to this report, a direct contribution of the *iro* locus for virulence of *E. coli* has not been demonstrated. Inactivation of the Tsh, Iro, and aerobactin systems greatly reduced *in vivo* persistence (Fig. 1) and development of lesions in respiratory and deeper tissues. Reintroduction of the *iro* genes on a multicopy plasmid restored the capacity to persist *in vivo* to the Aerobactin$^-$, Iro$^-$, Tsh$^-$ mutant to levels similar to that of wild-type strain χ7122. The increased copy number of the *iro* genes from the p15A replicon is most likely responsible for an overt regain in virulence in the Aerobactin$^-$, Iro$^-$, Tsh$^-$ mutant. However, these complementation results suggest that the pAPEC-1 encoded aerobactin and Iro systems function in concert to increase bacterial acquisition of iron within host extraintestinal tissues.

The most common pathogen-specific transcripts isolated were homologous to genes associated with ColE2-related plasmids (*ecs-13* to *ecs-28*). ColE2-related plasmids are small (5–11 kb) multicopy (10–20 copies per cell) plasmids (41). The higher copy level of a ColE2-related plasmid present in χ7122 would lead to abundant levels of corresponding transcript. This is most likely why sequences corresponding to this plasmid were frequently isolated by SCOTS. Reduced isolation of clones corresponding to the ColE2-related plasmid could have been obtained by including an excess of unlabeled ColE2-related plasmid DNA along with *E. coli* K-12 DNA during capture rounds. Transcripts corresponding to *tra* genes and other F plasmid-related genes (*ecs-29* to *ecs-32*) were also identified by SCOTS. F-pilus export and plasmid transfer comprise a type IV secretion system. Many type IV secretion systems are critical factors for the virulence of bacterial pathogens (42), although a role for F-pili in the virulence of *E. coli* has not been demonstrated. Bacteriophage-related genes were also identified by SCOTS. Bacteriophage-associated genes may code for virulence properties such as toxins and adhesins. It is currently not known whether phage-encoded genes are associated with the virulence of APEC strain χ7122. Among the APEC-specific transcripts of unknown function, *ecs-43* is highly similar to the core gene of a non-*E. coli* K-12 rearrangement hotspot element *RhsH* (43). Interestingly, in *E. coli* K-12, *Rhs* core ORFs are not expressed to a detectable extent during routine cultivation, and despite their

presence in many *E. coli* strains, the function of these elements has not been assessed.

Some representative SCOTS clones isolated after selective enrichment for pathogen-specific genes corresponded to genes that are also common to *E. coli* K-12. Importantly, among the *ecs* clones that corresponded to *E. coli* K-12 genes, many of these genes have been identified as preferentially expressed or required *in vivo* or have been shown to be essential genes (details are published as supporting information on the PNAS web site). Notably, *purA* and *pyrG* are required for synthesis of purines and pyrimidines, respectively, in *E. coli*. The importance of synthesis of nucleotide precursors for the virulence of bacteria *in vivo* is well established. *purA* (*ecs-49*) was selected as the initial null mutation and gene fusion to be used for *in vivo* gene expression analysis by *in vivo* expression technology in *Salmonella typhimurium* (33), because it is required for bacterial survival and virulence *in vivo*. Genes implicated in regulation and stress response were also identified by SCOTS. *uspB* (*ecs-62*) (*u*niversal *s*tress *p*rotein B) is a σ^S-dependent gene that contributes to resistance to ethanol (44). *topA* (*ecs-61*) encodes DNA topoisomerase I, which contributes to DNA supercoiling, global gene transcription, bacterial survival against high osmolarity (45), and oxidative damage (46). *phoB* (*ecs-60*) encodes the global response regulator of the phosphate regulon. PhoB has been shown to regulate *hilA* and invasion genes in *S. typhimurium* (47), and a *phoB* mutant of *V. cholorae* is less able to colonize the rabbit intestine (48). Although *phoB* has not been shown to contribute to virulence in *E. coli*, a pathogenic *E. coli* strain causing septicemia in swine that is defective for high-affinity phosphate transport was reduced in virulence (49). Among the genes of unknown function identified by SCOTS, *ynfK* (*ecs-66*) encodes a putative dethibiotin synthase. Interestingly, avidin, which is a biotin-binding protein, is induced in avian tissues after *E. coli* infection (50), and therefore biotin limitation may be an innate host response against infection.

The isolation of certain *E. coli* K-12 genes after selective reduction of these sequences implicitly suggests these genes are highly transcribed *in vivo*. Higher transcription levels would result in an abundance of sequences in the SCOTS cDNA pool, and therefore removal of these abundant products would be less likely even after removal of some of these transcripts with an excess of *E. coli* K-12 genomic DNA. The pertinence of most of these conserved genes for survival *in vivo* also supports the likelihood that a greater amount of transcription of these genes may occur *in vivo*. Further experiments, such as quantitative real-time PCR or microarray analyses, will be required to confirm whether these *E. coli* conserved genes are actually expressed at high levels *in vivo* in infected tissues or during growth *in vitro*.

The current study provides only a glimpse of some of the bacterial genes that are transcribed inside the host during *E. coli* infection, because only a fraction of the clones that were isolated by SCOTS have been analyzed so far. Moreover, because the major purpose of this study was to identify pathogen-specific genes that were expressed in host tissues, we have not addressed whether these genes are expressed specifically or preferentially *in vivo* or in a particular tissue location. However, such analyses could readily be determined by selective enrichment of sequences transcribed under specific conditions, as has been the focus of other reports using SCOTS (19–21). The approach we have presented herein could also be used to identify pathogen-specific genes that are transcribed *in vivo* in other types of pathogenic *E. coli* or closely related *Shigella* spp. in different animal or cell culture models. In addition, for other bacterial species, differences in gene transcription in host cells or tissues could be established by comparative blocking between different strains belonging to the same or similar species with high overall DNA homology but that have different degrees of virulence or host specificities. For instance, comparative *in vivo* gene expression of *S. typhimurium* vs. *Salmonella typhi* or *Yersinia pseudotuberculosis* vs. *Y. pestis* in appropriate infection models. Identification of pathogen-specific and conserved bacterial genes that are expressed *in vivo* will provide further insight into the mechanisms by which bacteria colonize host tissues, cope with, or circumvent host defenses and adjust to the nutrient limitations and other stresses that occur in different host environments.

We thank C. Squires (Tufts University, Boston) for plasmid pC6, Brian Morrow for helpful suggestions, and J. Clark-Curtiss for reviewing the manuscript, and we gratefully acknowledge Daryoush Saeed-Vafa for isolation of the DNA clone encompassing the *iro* gene cluster. This work was supported by U.S. Department of Agriculture National Research Initiative Competitive Grants Program Grants 97-35204-4512 and 00-35204-9224 (to R.C. and C.M.D.) and by National Institutes of Health Grant AI24533 (to R.C.). C.M.D. and F.D. were supported by Canadian Natural Sciences and Engineering Research Council postdoctoral fellowships.

1. Sussman, M., ed. (1997) *Escherichia coli: Mechanisms of Virulence* (Cambridge Univ. Press, Cambridge, U.K.).
2. Russo, T. A. & Johnson, J. R. (2000) *J. Infect. Dis.* **181,** 1753–1754.
3. Dozois, C. M. & Curtiss, R., III (1999) *Vet. Res.* **30,** 157–179.
4. Ochman, H., Lawrence, J. G. & Groisman, E. A. (2000) *Nature* **405,** 299–304.
5. Blattner, F. R., Plunkett, G., III, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., *et al.* (1997) *Science* **277,** 1453–1474.
6. Perna, N. T., Plunkett, G., III, Burland, V., Mau, B., Glasner, J. D., Rose, D. J., Mayhew, G. F., Evans, P. S., Gregor, J., Kirkpatrick, H. A., *et al.* (2001) *Nature* **409,** 529–533.
7. Reid, S. D., Herbelin, C. J., Bumbaugh, A. C., Selander, R. K. & Whittam, T. S. (2000) *Nature* **406,** 64–67.
8. Bergthorsson, U. & Ochman, H. (1998) *Mol. Biol. Evol.* **15,** 6–16.
9. Jarvis, K. G., Giron, J. A., Jerse, A. E., McDaniel, T. K., Donnenberg, M. S. & Kaper, J. B. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 7996–8000.
10. Bloch, C. A. & Rode, C. K. (1996) *Infect. Immun.* **64,** 3218–3223.
11. Brown, P. K. & Curtiss, R., III (1996) *Proc. Natl. Acad. Sci. USA* **93,** 11149–11154.
12. Fang, F. C., Libby, S. J., Buchmeier, N. A., Loewen, P. C., Switala, J., Harwood, J. & Guiney, D. G. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 11978–11982.
13. Garcia-Del Portillo, F., Pucciarelli, M. G. & Casadesus, J. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 11578–11583.
14. Low, D. A., Weyand, N. J. & Mahan, M. J. (2001) *Infect. Immun.* **69,** 7197–7204.
15. Sperandio, V., Mellies, J. L., Nguyen, W., Shin, S. & Kaper, J. B. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 15196–15201.
16. Dho-Moulin, M. & Fairbrother, J. M. (1999) *Vet. Res.* **30,** 299–316.
17. Carlson, H. C. & Whenham, G. R. (1968) *Avian Dis.* **12,** 297–302.
18. Dozois, C. M., Dho-Moulin, M., Bree, A., Fairbrother, J. M., Desautels, C. & Curtiss, R., III (2000) *Infect. Immun.* **68,** 4145–4154.
19. Graham, J. E. & Clark-Curtiss, J. E. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 11554–11559.
20. Morrow, B. J., Graham, J. E. & Curtiss, R., III (1999) *Infect. Immun.* **67,** 5106–1516.
21. Daigle, F., Graham, J. E. & Curtiss, R., III (2001) *Mol. Microbiol.* **41,** 1211–1222.
22. Hou, J. Y., Graham, J. E. & Clark-Curtiss, J. E. (2002) *Infect. Immun.* **70,** 3714–3726.
23. Provence, D. L. & Curtiss, R., III (1992) *Infect. Immun.* **60,** 4460–4467.
24. Chang, A. C. & Cohen, S. N. (1978) *J. Bacteriol.* **134,** 1141–1156.
25. Alexeyev, M. F. (1995) *BioTechniques* **18,** 52–56.
26. Lennox, E. S. (1955) *Virology* **1,** 190–206.
27. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. M., Seidman, J. G., Smith, J. A. & Struhl, K. (1991) *Molecular Cloning: A Laboratory Manual* (Greene and Wiley Interscience, New York).
28. Miller, J. H. (1992) *A Short Course in Bacterial Genetics: A Laboratory Manual for Escherichia coli and Related Bacteria* (Cold Spring Harbor Lab. Press, Plainview, NY).
29. Froussard, P. (1992) *Nucleic Acids Res.* **20,** 2900.
30. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410.
31. Neidhardt, F. C., Curtiss, R., III, Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M. & Umbarger, H. E. (1996), *Escherichia coli and Salmonella: Molecular and Cellular Biology* (Am. Soc. Microbiol., Washington, DC).
32. Hensel, M., Shea, J. E., Gleeson, C., Jones, M. D., Dalton, E. & Holden, D. W. (1995) *Science* **269,** 400–403.
33. Mahan, M. J., Slauch, J. M. & Mekalanos, J. J. (1993) *Science* **259,** 686–688.
34. Valdivia, R. H. & Falkow, S. (1997) *Science* **277,** 2007–2011.
35. Provence, D. L. & Curtiss, R., III (1994) *Infect. Immun.* **62,** 1369–1380.
36. Otto, B. R., van Dooren, S. J., Dozois, C. M., Luirink, J. & Oudega, B. (2002) *Infect. Immun.* **70,** 5–10.
37. Torres, A. G., Redford, P., Welch, R. A. & Payne, S. M. (2001) *Infect. Immun.* **69,** 6179–6185.
38. Der Vartanian, M., Jaffeux, B., Contrepois, M., Chavarot, M., Girardeau, J. P., Bertin, Y. & Martin, C. (1992) *Infect. Immun.* **60,** 2800–2807.
39. Baumler, A. J., Norris, T. L., Lasco, T., Voight, W., Reissbrodt, R., Rabsch, W. & Heffron, F. (1998) *J. Bacteriol.* **180,** 1446–1453.
40. Russo, T. A., Carlino, U. B., Mong, A. & Jodush, S. T. (1999) *Infect. Immun.* **67,** 5306–5314.
41. Mock, M. & Pugsley, A. P. (1982) *J. Bacteriol.* **150,** 1069–1076.
42. Christie, P. J. & Vogel, J. P. (2000) *Trends Microbiol.* **8,** 354–360.
43. Wang, Y. D., Zhao, S. & Hill, C. W. (1998) *J. Bacteriol.* **180,** 4102–4110.
44. Farewell, A., Kvint, K. & Nystrom, T. (1998) *J. Bacteriol.* **180,** 6140–6147.
45. Bhriain, N. N. & Dorman, C. J. (1993) *Mol. Microbiol.* **7,** 351–358.
46. Tse-Dinh, Y. C. (2000) *J. Bacteriol.* **182,** 829–832.
47. Lucas, R. L., Lostroh, C. P., DiRusso, C. C., Spector, M. P., Wanner, B. L. & Lee, C. A. (2000) *J. Bacteriol.* **182,** 1872–1882.
48. von Kruger, W. M., Humphreys, S. & Ketley, J. M. (1999) *Microbiology* **145,** 2463–2475.
49. Daigle, F., Fairbrother, J. M. & Harel, J. (1995) *Infect. Immun.* **63,** 4924–4927.
50. Elo, H. A., Raisanen, S. & Tuohimaa, P. J. (1980) *Experientia* **36,** 312–313.