Extensive Gene Diversity in Septicemic *Escherichia coli* Strains

Daphna Mokady, Uri Gophna, and Eliora Z. Ron*

*Department of Molecular Microbiology and Biotechnology, the George S. Wise Faculty of Life Sciences,
Tel-Aviv University, Tel-Aviv, Israel*

Extraintestinal pathogenic *Escherichia coli* strains (ExPEC) are the cause of a diverse spectrum of invasive infections in humans and animals, and these infections often lead to septicemia. Strains of serogroups O2 and O78 of *E. coli* are involved in human urinary tract infections and newborn meningitis and also constitute the major serotypes involved in avian colisepticemia. In the present study we compared the unique genomic sequences of two such septicemic strains, strains O2-1772 and O78-9, obtained by suppression subtractive hybridization. Evaluation of the degree of similarity between these two strains, which cause the same disease, revealed a high degree of diversity, with only a few shared genes. Subsequently, additional strains of each serogroup of human and animal origin were screened by PCR, and the results provided further evidence for the existence of a high degree of genome plasticity. These results were unexpected, in view of data showing that the two O157:H7 strains that have been sequenced are nearly identical in terms of virulence factors. Furthermore, the data obtained for the septicemic strains suggest that each step in the infection can be mediated by a number of alternative virulence factors, indicating the existence of a mix-and-match combinatorial system. Although whole-genome comparisons of *E. coli* strains causing different diseases have shown great differences in gene contents, we show that such differences exist even within strains that cause the same disease and that target the same host tissues. Moreover, in addition to the high level of genome plasticity, we show that the large pool of virulence genes in the septicemic strains is independent of the host, implying a high degree of zoonotic risk.

Pathogenic *Escherichia coli* strains cause intestinal or extraintestinal infections in many host species. Strains that cause extraintestinal infections are involved in a diverse spectrum of diseases, including urinary tract infections (UTIs), newborn meningitis (NBM), and septicemia (12, 19, 22, 25, 27, 30). Extraintestinal pathogenic *E. coli* (ExPEC) strains constitute an increasing problem for human medicine, especially due to the high incidence of drug resistance often transmissible by plasmids. In addition to the human diseases, ExPEC strains are the cause of significant economic losses on animal farms, especially in the poultry industry. The virulence factors of ExPEC strains are not yet fully identified, nor are the differences between the virulence mechanisms of the various pathogenic serotypes.

We have been studying the virulence factors of ExPEC strains involved in human disease as well as in avian colisepticemia, such as serogroups O2 and O78, which constitute two of the major serotypes involved in avian sepsis. The diseases caused by the two serotypes in poultry are symptomatically similar: they start as a respiratory tract infection, which leads to a systemic infection involving the internal organs and sepsis (7, 18, 22). However, several differences between the two serogroups have already been identified, and different virulence factors were found in strains belonging to each serogroup (17). A thorough genetic comparison of these strains could reveal the common mechanisms involved in sepsis in general and in avian sepsis in particular. Human strains of serotype O2 are

also associated with human UTIs, while strains of serotype O78 have been known to cause human NBM. Therefore, a comparison of the two serogroups may also reveal genetic factors relevant to the study of human disease.

The molecular procedure of suppression subtractive hybridization (SSH) allows the comparison of two genomes and the detection of specific sequences that are present exclusively in one of them. This technique has been shown to be extremely efficient in detecting genomic sequences and genomic islands that are strain specific (2, 26). In research on pathogenic bacteria, SSH was used to identify virulence factors by comparing pathogenic and nonpathogenic bacterial strains of the same species (15, 29).

In this work, the genome of nonpathogenic laboratory strain K-12 was subtracted from the genome of avian pathogenic strain O2-1772 and, in parallel, from the genome of avian strain O78-9. The results from these two experiments were combined and analyzed. A total of 158 unique sequences that are not present in nonpathogenic strain K-12 were identified. Interestingly, only a few pathogen-specific sequences were shared by the two strains, even though they cause clinically similar diseases.

MATERIALS AND METHODS

Bacterial strains and growth conditions. Bacteria were routinely grown at 37°C in Luria-Bertani (LB) broth or LB agar. *E. coli* K-12 MG1655 strain was used as a driver strain for the subtractive hybridization. Human UTI strains *E. coli* O2 U33041 (U33) and B18996 (B18) were kindly provided by Dan Katcoff; and human NBM O78 strains 285, 286, and 287 were kindly provided by Hedda Milch. All the avian strains were septicemic (O78 strains O78-9 [28], 786, and 787 and O2 strains 1772, YN, MAN, and SA) and were from the laboratory strain collection; strain O2 BEN2908 (BEN) was kindly provided by Maryvonne Moulin-Schouleur. Sheep O78 strain 63–1 was from our laboratory collection.

* Corresponding author. Mailing address: Department of Molecular Microbiology and Biotechnology, the George S. Wise Faculty of Life Sciences, Tel-Aviv University, Tel-Aviv 69978, Israel. Phone: 972-3-6409379. Fax: 972-3-6414138. E-mail: eliora@post.tau.ac.il.

DNA techniques. Isolation of genomic DNA was performed as described previously (23). DNA fragments obtained from the subtractive hybridization were cloned into pGEM-T plasmids of the Promega pGEM-T and pGEM-T EASY Vector Systems kit, according to the instructions of the manufacturer.

Sequence analysis. Sequence analysis was carried out by using National Center for Biotechnology Information algorithms BLASTN, BLASTX, and BLASTP (3).

SSH. The SSH technique was carried out as described previously (8). We used the CLONTECH PCR-Select Bacterial Genome Subtraction kit, as recommended by the manufacturer. Genomic DNAs isolated from strains O2-1772 and O78-9 were used as tester DNAs for two separate hybridization experiments with the driver DNA of strain K-12. Two rounds of PCR were performed with the hybridization products to amplify the specific fragments unique to each of the pathogenic strains.

Confirmation of SSH libraries. Genes found only in one of the SSH libraries were further examined. PCR was performed with the sequences which were found to be virulence related or sequences for which we could not find a known function or homology. PCRs were carried out under nonstringent conditions.

RESULTS

Identification of genomic sequences specific to the pathogenic strains. SSH was performed between each of the pathogenic strains O2-1772 and O78-9 and nonpathogenic strain K-12 in two separate experiments (see Materials and Methods). The two consequent libraries contained a total of 256 sequences, 158 of which (about 105 kb in total) were found to be specific to at least one of the pathogenic strains and absent from driver strain K-12. These unique sequences, which varied between 91 and 1,473 bp in length, were cloned and sequenced.

Sequence analysis of fragments identified by SSH. A search for sequences homologous to the specific sequences was performed with the BLASTN and BLASTX algorithms (3) in order to further analyze them. The results are summarized in Table 1. The sequences were divided into several categories, as shown in Fig. 1. The first category included nine sequences of putative or verified virulence factors, such as components of iron uptake systems, genes of the type III secretion system (TTSS), and the specific pili of each strain. The second category included sequences classified as potential virulence-associated sequences, according to their predicted conserved domains. These were sequence O2-38, which encodes a predicted ATP-binding protein involved in virulence, and sequence O78-95, which encodes a putative adhesin. The third category included 21 sequences with known or putative functions whose role in pathogenicity is not clear. The fourth category included 34 sequences with no known function that could be divided into two types: fragments with similarity to sequences of unknown function (e.g., conserved hypothetical proteins) and fragments for which no similarity was found.

Thirty-seven sequences were mobility related, such as transposases, integrases, and insertion-sequences associated sequences. The last two categories consisted of fragments that are known to be bacteriophage associated or that are similar to the replication and mobilization loci of various known plasmids. The phage sequences found in strain O2-1772 were associated with phage P1, phage 21, phage H-19B, and a lambdoid phage. Sequences of O78-9 were associated with phage CP-933K, phage CP-933I, and a bacteriophage lambda-like phage. The plasmid sequences found in the strains tested showed similarity to pColE1, pRA1, p15B, F plasmid, ColIa-IHE3113, and ColIbP9 in strain O2-1772 and pColE1 and pColIbP9 in strain O78-9.

Known PAIs from other *E. coli* strains. Analysis of the two SSH libraries revealed that many of the specific sequences were homologous to sequences from pathogenicity islands (PAIs) of other pathogenic bacteria, which were presumably acquired by horizontal transfer (13). Interestingly, several sequences of the SSH libraries of the septicemic strains showed similarity to the sequences of *E. coli* strains involved in UTIs (9) and enterohemorrhagic strain *E. coli* O157:H7 (20). The data also indicate PAIs which are homologous to those of other bacteria, such as *Shigella flexneri* (the SRL, SHI-3, and SHE PAIs).

Verification of SSH libraries. Sequences that were present in the subtracted library of only one of the strains were further tested by PCR to verify their absence from the genome of the other strain. Forty-five virulence-related sequences and sequences with unknown function were examined. In general, the PCR results were in agreement with the SSH results, showing that 34 sequences were absent from one of the strains, whereas only 9 sequences not observed in the SSH libraries could be detected by PCR. Although we assume that the SSH library does not contain all the strain-specific sequences, our library appears to represent many large gene clusters which encode complex functions, such as PAIs, as well as single genes.

Comparison of the two SSH libraries. The two SSH libraries verified by PCR have several common characteristics, as can be observed from sequence analysis (Table 1). Most of the sequences showed homology to pathogenic strains of *E. coli*, as well as to other pathogenic bacteria, most frequently to the phylogenetically closely related species *S. flexneri* and different serovars of *Salmonella*. Both libraries include numerous mobility-associated sequences, such as transposases and integrases, along with phage-related sequences, plasmid sequences, and insertion sequence-associated sequences. This implies that both libraries are highly specific for areas of the genome that were acquired horizontally, some of which may be related to virulence.

Components of two major virulence mechanisms were found in both libraries. The first is iron uptake systems: both pathogenic strains were found to possess different iron acquisition systems that were absent from the nonpathogenic strain. Strain O78-9 was found to possess the yersiniabactin system, coded by the *Yersinia* high-pathogenicity island. Strain O2-1772 has components of the *sitABCD* system, known from *S. flexneri* and *Salmonella enterica* serovar Typhimurium. The second major virulence mechanism is adhesins: both libraries contained components of virulence-related pili. Two minor subunits of the AC/I pilus were found in strain O78-9, and a chaperone involved in the assembly of the P pilus was found in strain O2-1772.

The two SSH libraries were also compared to each other by use of the BLAST algorithm. First, each sequence of the O78-9 library was run by using the BLASTN algorithm. Usually, the sequence showed similarity to a part of an open reading frame (ORF). We selected the complete ORF of the closest homolog and performed a search against the entire library of O2-1772 with the BLAST algorithm. Thus, even if two sequences from the two libraries were homologous to different parts of the same gene, the similarity between them could be determined. Some of the sequences did not show homology within an ORF, and these were run without manipulations against the whole library of O2-1772. We found two adjacent genes of TTSS: *eivC* of O78-9 and *eivA* of O2-1772. Both genes had very strong

TABLE 1. Specific SSH fragments and their homologies determined with the BLASTN and BLASTX algorithms[a]

Sequence group and clone	Homology by BLASTN or BLASTX	Length (bp)	e value	GenBank accession no.	Function
Virulence-related sequences					
O2-38	*Shigella flexneri* 2a strain 301 SF3002	596	0.0	AAN44483	Predicted ATP-binding protein involved in virulence
O2-106	*E. coli* O157:H7 EDL933 *eivA*	300	$5e^{-90}$	AAG57987	Type III secretion apparatus protein
O2-108	*S. flexneri* 2a strain, *sit* operon	530	0.0	AAN42967-9	Iron transport proteins
O2-154	*E. coli* pSFO157 SfpD	392	e^{-124}	AJ131667	P pilus assembly chaperon
O78-13	*E. coli* strain 536 PAI III *sfaG*	617	0.0	CAA34652	Minor subunit of AC/I pili
O78-74	*E. coli* CFT073 *sfaH*	817	0.0	NP_753159	Minor subunit of AC/I pili
O78-44 + O78-102	*E. coli* strain 536 PAI III	631	0.0	X16664	Enterobactin receptor *iroN*
O78-95	*E. coli* O157:H7	1,002	0.0	NP_290185	Autotransporter, putative adhesin
O78-96	*Yersinia pestis ybtS*	666	0.0	NP_405477	Versiniabactin synthesis
O78-163	*E. coli* O157:H7 *eivC*	622	0.0	NP_289427	Type III secretion system
Known and putative functions					
O2-4	*Shewanella oneidensis* MR-1	299	$3e^{-49b}$	NP_716453	Outer membrane efflux family protein
O2-16	*S. flexneri* 2a SRL PAI	314	$5e^{-72}$	AF326777	LysR-like transcriptional regulator
O2-23	*S. enterica* serovar Typhimurium genomic island I S005	591	e^{-109}	AF261825	Putative mating pair stabilization protein
O2-27	*E. coli* clone SauE4 C12, neonatal meningitis strain-specific DNA sequence	636	$1e^{-58}$	AF222135	Putative Lon protease, truncated
O2-37	*S. flexneri* 2a *she* PAI ORF24 + ORF25	553	0.0	AAK00478, AAK00479	ORF24, putative antirestriction protein ORF25, putative RadC-like protein YeeS
O2-54	*E. coli* beta-lactamase	519	0.0	AF427132	Beta-lactamase
O2-60	*E. coli cvaB*	473	0.0	X57524	Protein involved in colicin V secretion
O2-64	*Staphylococcus aureus* subsp. MW0218	493	$1e^{-10c}$	BAB94083	Maltose- and glucose-specific factor II
O2-113	*E. coli* CFT073 *c4764*	277	e^{-104}	AE016769AE014075	Carbamate kinase-like protein YahI
O2-173	*E. coli hsdS*	576	$1e^{-27}$	AJ132566	Subunit of restriction enzyme
O2-284	*E. coli* O157:H7 EDL933 Z2508	410	0.0	AAG56520	Putative efflux pump
O2-316	*S. flexneri* 2a strain SF2763	391	0.0	AE015290, AE005674	Putative transcriptional regulator protein
O2-318	*Sulfolobus solfataricus*	263	0.003^{c}	AAK4183	Glycosyltransferase
O2-323	*Vibrio cholerae traV*	547	$9e^{-31c}$	AAL59672	Sex pilus assembly
O2-333	*S. enterica* serovar Typhimurium LT2 STM2342	452	$3e^{-65}$	AE008805, AE006468	Putative inner membrane and transport protein
O78-12	*E. coli* strain 536 PAI III ORF38	1,473	0.0	CAD66186	Putative nitrile hydratase
O78-73a	*E. coli* CFT073	597	0.0	NP_753855	Putative regulator
O78-73b	*E. coli* CFT073	704	0.0	NP_753853	Putative enzyme
O78-84	*Legionella pneumophila*	516	$8e^{-28c}$	AAM08232	Putative type I restriction enzyme
O78-151	*S. flexneri* 2a strain 301 SHE PAI	572	0.0	NP_708769	Putative antirestriction protein
O78-169b	*S. enterica* serovar Typhi		$6e^{-38c}$	NP_457351	Putative regulator
Unknown function and no homology					
O2-1	*E. coli* strain 536 PAI II ORF44	367	e^{-135}	CAD42059	Conserved hypothetical protein
O2-14		766			
O2-32	*Silicibacter pomeroyi*	295	$3e^{-49b}$		Unknown
O2-42	*E. coli* O157:H7 EDL933, Z1637	463	$3e^{-40}$	NP_309406	Conserved hypothetical protein
O2-87	*S. pomeroyi*	322	$6e^{-07b}$		Unknown
O2-125		804			
O2-132		545			
O2-138	*E. coli* strain 536 PAI II, no ORF	222	$2e^{-43}$	AJ494981	Unknown
O2-144		197			
O2-157		518			
O2-158	*Klebsiella pneumoniae* plasmid, no ORF	530	3e-59	AJ276453	Unknown
O2-164		262			

Continued on facing page

TABLE 1—*Continued*

Sequence group and clone	Homology by BLASTN or BLASTX	Length (bp)	*e* value	GenBank accession no.	Function
O2-165	*Photorhabdus luminescens* strain NC19 NgrC	339	$4e^{-91}$	AAK16082	Unknown
O2-193		491			
O2-207	*E. coli* CFT073 *c2499*	274	e^{-148}	AAN80955	Conserved hypothetical protein
O2-210	*E. coli* strain 536 PAI I, ORF28 + ORF29	422	e^{-174}	CAD33738 CAD33739 CAD33739	Conserved hypothetical proteins
O2-280		153			
O2-311	*E. coli* strain 536 PAI I, no ORF	304	$5e^{-13}$	AJ488511	Unknown
O2-319		425			
O2-334		295			
O2-345	*E. coli* CFT073, no ORF	91	$2e^{-43}$	AE016770 AE014075 AE014075	Unknown
O2-349	*E. coli* strain 536 PAI I ORF73	183	$8e^{-91}$	AJ488511	Putative intergenic-region protein
O2-353	*E. coli* O157:H7 EDL933 Z1600	365	e^{-125}	AE005309 AE005174 AE005174	Hypothetical protein
O2-355	*Pseudomonas resinovorans*	445	$2e^{-15c}$	BAC41652	Hypothetical protein
O78-17	*E. coli* O157:H7, O island 7	890	0.0	AE005199	Hypothetical protein
O78-18	*E. coli* O157:H7, O island 7	659	0.0	AE005198	Conserved hypothetical protein
O78-24		634			
O78-27a	*S. flexneri* 2A strain 301, no ORF	1,338	0.0	AE015221	Unknown
O78-55		602			
O78-63	*E. coli* O157:H7	604	0.0	NP_310823	Hypothetical protein
O78-69		665			
O78-79	*Pseudomonas syringae*	757	$8e^{-08c}$	NP_791901	Hypothetical protein
O78-138		375			
O78-161	*E. coli* plasmid p1658/97 ORF50	837	0.0	AAO49623	Putative protein

[a] Homology was determined with the BLASTN or BLASTX algorithm.
[b] Microbial blast *e* value.
[c] BLASTX *e* value.

homology to the TTSS of *E. coli* O157:H7. Sequences O78-151 and O2-37 showed only limited similarity in one domain by use of the BLASTX algorithm and were clearly very distant.

Distribution of specific sequences in other septicemic O2 and O78 strains. The comparison of two strains, one of serotype O2 and one of serotype O78, indicated a high degree of variability in virulence-related sequences between the strains. In order to determine if these sequences were conserved within the same serotype, we performed PCRs to look for the presence of the specific sequences in additional strains of the same serotype. Altogether, six additional strains of each serotype were examined. All the strains were invasive, and about half were of human origin. The results (Fig. 2 and Table 2) indicate that among the 45 sequences examined, only 1 sequence (the putative autotransporter, sequence O78-95) was serogroup specific and appeared in all of the O78 strains but none of the O2 strains. Most other sequences were patchily distributed and were present in only a small subset of strains; therefore, they may have been acquired by fairly recent horizontal gene transfer. Genes such as O2-207, O2-345, and O78-96 probably represent an ancient transfer event, as they were found in all the pathogenic strains tested but not in strain K-12.

DISCUSSION

The study of the comparative genetics and evolution of pathogenic bacteria has progressed rapidly in recent years as the sequences of more bacterial genomes have been determined. Technologies for the comparison genomes constitute powerful tools for identifying major conserved virulence genes and understanding the evolutionary processes behind diversity and virulence. Comparison of genomes by computer analysis (16, 27) requires the availability of complete genome sequences. SSH is an alternative that enables the comparison of two genomes, even if they have not been sequenced. This technology has already been used and has supplied important information which is valuable in providing an understanding of several aspects of pathogenicity, such as the virulence factors used by various pathogenic bacteria (21), adherence and invasion mechanisms (14), and the acquisition and mobility of PAIs (15, 26).

We used SSH in order to identify specific sequences of two different serotypes of septicemic *E. coli*, one of serotype O2 and one of serotype O78. The SSH libraries obtained consisted of 158 sequences, including putative new virulence factors and sequences with homology to PAIs not known to be present in septicemic strains. Although these libraries are substantial, we assume that they do not contain all the unique genes absent from strain K-12. For example, we know of the existence of the yersiniabactin iron uptake system in strain O2-1772 (11), but it was not detected in our SSH library. However, the results of the subsequent PCR experiments indicated that the SSH libraries were reliable: most of the genes absent from the SSH library were also not detected by PCR.
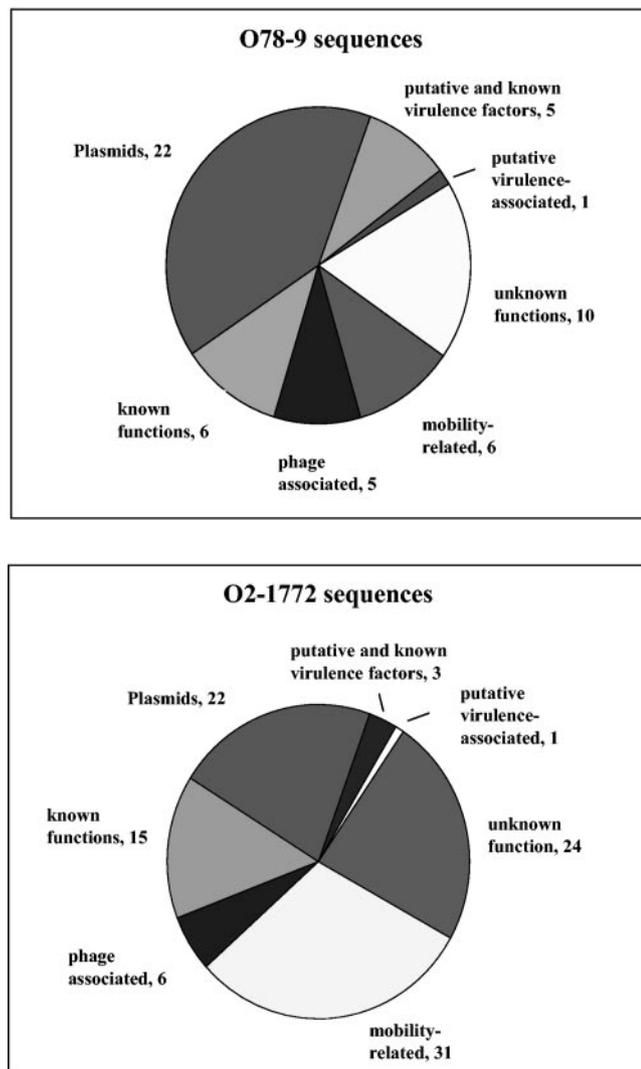
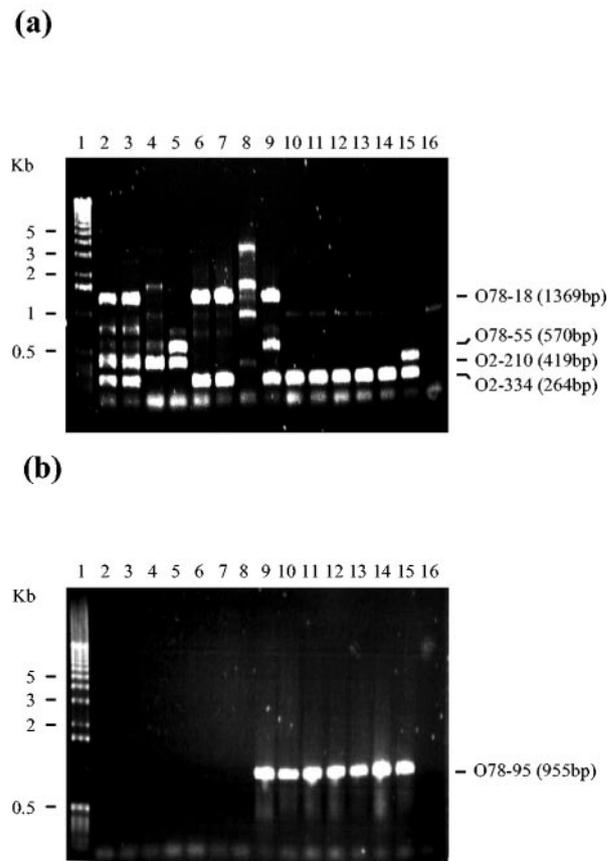FIG. 1. Breakdown of the specific sequences of each library into categories according to their homologies.

FIG. 2. Results of PCRs of specific sequences, performed with 14 septicemic strains and nonpathogenic strain K-12. Lanes 1 to 16, 1-kb DNA marker, O2-1772, O2-YN, O2-MAN, O2-SA, O2-U33, O2-B18, O2-BEN, O78-9, O78-285, O78-286, O78-287, O78-6, O78-7, O78-63-1, and K-12, respectively. (a) Multiplex PCR with primers specific for the sequences of O2-210, O2-334, O78-18, and O78-55; (b) PCR with primers specific for the putative autotransporter (O78-95 sequence).

Avian strains of serogroup O2 involved in cellulitis were previously analyzed by SSH (24). However, this study contained very few genes of significance for pathogenesis. Thus, only 62 fragments were obtained in the study of the cellulitis strains, half of which were phage sequences and only 1 of which was similar to PAI genes. Our experiments detected the presence of many more unique sequences ($n = 102$), many of which were similar to PAIs of various virulent strains. The differences in the results probably reflect the fact that the cellulitis strains are less virulent than the septicemic O2 strain tested in our study. Thus, the more virulent septicemic strains appear to have acquired a larger array of virulence factors.

A comparison of the SSH libraries of strain O2-1772 and strain O78-9 revealed that the diversity between the two strains is very high. These results are consistent with the results from a multilocus sequence type (MLST) analysis showing that the evolutionary distance between these strains is indeed large, as reflected by the sequence divergence of six housekeeping genes with adequate polymorphisms. According to the MLST results,

the various O2 and O78 strains (22 in total) are distributed on different branches of the phylogenetic tree (1). This study indicated that clonal division in *E. coli* O78 and O2 strains is host independent, and strains from the same host show high degrees of variability. It therefore follows that the ability of strains to conquer new niches and infect a new host species depends more on horizontally acquired DNA than on the vertically inherited genotype.

Many of the sequences examined in this study are of genes that have an unknown function or that have no known homologs. Twenty-seven such sequences were present only in the O2-1772 or O78-9 sequence. PCRs performed with additional strains in order to determine if these sequences were conserved within strains of the same serotype indicated that only one sequence was serogroup specific and appeared in all the O78 strains but none of the O2 strains. This sequence (O78-95) may therefore serve as a diagnostic marker for septicemic O78 strains. This is clearly one of the spin-offs of this study that is being investigated further.

The prevalence of sequences associated with genomic plasticity in both libraries supports the assumption that the patho-

TABLE 2. Distribution of specific sequences in 14 septicemic *E. coli* strains and nonpathogenic strain K-12[a]

sequence	O2 strains							O78 strains							K12
	1772 avian	YN avian	MAN avian	SA avian	BEN avian	U33 human	B18 human	285 human	286 human	287 human	63-1 sheep	786 avian	787 avian	789 avian	
O2-1 hypothetical protein	■	■	■	■							■	■			
O2-14 no homology	■	■													
O2-32 unknown	■	■													
O2-38 predicted ATP-binding protein	■	■	■		■	■	■				■		■		
O2-42 hypothetical protein	■	■													
O2-87 unknown	■	■				■									
O2-108 iron transport proteins	■	■													
O2-125 no homology	■	■													
O2-132 no homology	■	■				■									
O2-138 unknown	■	■	■												
O2-144 no homology	■	■													
O2-154 P pili assembly chaperon	■	■													
O2-157 no homology	■	■													
O2-158 unknown	■	■				■	■								
O2-164 no homology	■	■													
O2-165 unknown	■	■				■	■								
O2-193 no homology	■	■	■	■											
O2-207 hypothetical protein	■	■	■	■	■	■	■	■	■	■	■	■	■	■	
O2-210 hypothetical protein	■	■	■	■	■	■	■	■	■	■	■	■		■	
O2-280 no homology	■	■													
O2-311 unknown	■	■	■	■											
O2-319 no homology	■	■				■	■								
O2-334 no homology	■	■				■	■	■	■	■	■		■	■	
O2-345 unknown	■	■			■	■	■	■	■	■	■	■	■	■	
O2-349 putative protein	■	■			■	■	■	■			■				
O2-353 hypothetical protein	■	■						■	■	■	■				
O2-355 hypothetical protein	■	■				■	■								
O78-13/O78-74 minor subunits of AC/I pili														■	
O78-17 hypothetical protein															
O78-18 hypothetical protein	■	■				■	■							■	
O78-24 no homology														■	
O78-27a unknown															
O78-44/O78-102 enterobactin receptor iroN	■						■								
O78-55 no homology				■										■	
O78-63 hypothetical protein	■	■	■											■	
O78-69 no homology														■	
O78-79 hypothetical protein						■	■							■	
O78-95 Autotransporter								■	■	■	■	■	■	■	
O78-96 yersiniabactin synthesis	■	■				■	■	■			■	■	■		
O78-138 no homology														■	
O78-161 hypothetical protein														■	
O78-163/O2-106 TTSS proteins	■	■			■	■		■	■	■	■		■		

[a] PCR was performed for detection of virulence-related sequences and sequences with no known function or homology. The presence of a sequence in a strain is indicated by a gray square.

TABLE 3. Virulence factors of strains O2-1772 and O78-9 in the different stages of infection

Infection stage	Virulence factor(s) of strain		Reference(s) or source
	O78-9	O2-1772	
Adhesion	AC/I pilus		6
		P pilus	This work
	Type 1 pilus	Type 1 pilus	17
	Nonfimbrial adhesin		This work
Internalization	Curli fibers	Curli fibers	10
Invasion	TTSS	TTSS	D. Ideses et al., unpublished data
Persistence in the host, iron uptake	Enterobactin	Enterobactin	R. S. Adiri et al., unpublished data
	Aerobactin	Aerobactin	5, 11
	Yersiniabactin	Yersiniabactin	10; this work
	IroN receptor	IroN receptor	This work
		sitABCD	This work
Persistence in the host, immune evasion	Yersiniabactin	Yersiniabactin	4
	O antigen	O antigen, polysaccharide capsule	

genic strains evolved by processes involving genome remodeling and the horizontal acquisition of genomic regions from other pathogenic bacteria. In addition to the genes associated with genomic plasticity, many phage sequences were found in both libraries. However, these sequences differ between the two strains, with the exception of lambdoid phages, implying that the two bacterial strains have different susceptibilities to different families of phages. The differences in phage susceptibilities probably reflect differences in restriction modification systems and surface structures, such as pili or capsules, and consequently affect the genetic information that each strain can acquire from these phages.

The common virulence factors, such as the TTSS and the iron uptake systems, could be important to the infection process, as they exist in both strains. The fact that not many virulence factors are common to both strains is unexpected, since phenotypically both strains cause the same disease. Our findings imply that the two strains are using different factors with similar roles in the various stages of the infection process. This possibility is illustrated in Table 3, which compares the virulence factors that are thought to participate in the different stages of infection with strains O2-1772 and O78-9. The two strains have adherence pili, but these are of different types: O2 codes for the P pilus (this work), while O78 codes for the AC/I pilus (6) and for a nonfimbrial putative adhesin (this work). Another virulence factor that varies between the strains is the capsule. While O2 possesses a K1 polysaccharide capsule, O78 lacks such a capsule but is enveloped in a loose layer of extracellular material which is chemically similar to the O antigen (unpublished results).

These results indicate that each step in the infection process can be mediated by a number of alternative virulence factors, and each strain may have a unique combination of such factors. This assortment of virulence genes is apparently made possible by the variety of genetic factors contributing to genome plasticity, such as plasmids, phages, and transposable elements. The fact that extant *E. coli* strains vary so much in their ge-

nome contents indicates that this mix-and-match combinatorial approach has been a successful evolutionary strategy for this species, which can colonize many different tissues and hosts.

The availability of sequenced genomes and of technologies such as SSH makes it possible to examine the virulence factors through comparative genomics. Previous whole-genome comparisons of *E. coli* strains have shown great differences in the gene contents of strains. However, none of the comparative genomics of *E. coli* strains that have been sequenced compared strains that cause the same disease and that target the same host tissues. The only exception is in the two genome sequences of O157:H7. In these strains, the virulence factors and the genes coding for them show a very high degree of similarity. These results are in contrast to our findings for the septicemic O2 and O78 serogroups, which showed a very high level of genome plasticity. An additional and important conclusion from these data is the indication that the large pool of variable virulence genes is accessible to septicemic bacteria, independent of the host. It therefore remains to be determined whether host specificity exists among the septicemic ExPEC strains and which factors control the zoonotic risk.

Comparison of strains causing similar types of diseases is essential for the identification of shared traits which are important for pathogenicity and for study of their conservation, transfer, and evolution.

REFERENCES

1. **Adiri, R. S., U. Gophna, and E. Z. Ron.** 2003. Multilocus sequence typing (MLST) of *Escherichia coli* O78 strains. FEMS Microbiol. Lett. **222:**199–203.
2. **Akopyants, N. S., A. Fradkov, L. Diatchenko, J. E. Hill, P. D. Siebert, S. A.**

Lukyanov, E. D. Sverdlov, and D. E. Berg. 1998. PCR-based subtractive hybridization and differences in gene content among strains of *Helicobacter pylori*. Proc. Natl. Acad. Sci. USA **95:**13108–13113.

3. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389–3402.

4. **Autenrieth, I., K. Hantke, and J. Heesemann.** 1991. Immunosuppression of the host and delivery of iron to the pathogen: a possible dual role of siderophores in the pathogenesis of microbial infections? Med. Microbiol. Immunol. (Berlin) **180:**135–141.

5. **Babai, R., G. Blum-Oehler, B. E. Stern, J. Hacker, and E. Z. Ron.** 1997. Virulence patterns from septicemic *Escherichia coli* O78 strains. FEMS Microbiol. Lett. **149:**99–105.

6. **Babai, R., B. E. Stern, J. Hacker, and E. Z. Ron.** 2000. New fimbrial gene cluster of S-fimbrial adhesin family. Infect. Immun. **68:**5901–5907.

7. **Dho-Moulin, M., and J. M. Fairbrother.** 1999. Avian pathogenic *Escherichia coli* (APEC). Vet. Res. **30:**299–316.

8. **Diatchenko, L., Y. F. Lau, A. P. Campbell, A. Chenchik, F. Moqadam, B. Huang, S. Lukyanov, K. Lukyanov, N. Gurskaya, E. D. Sverdlov, and P. D. Siebert.** 1996. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. Proc. Natl. Acad. Sci. USA **93:**6025–6030.

9. **Dobrindt, U., G. Blum-Oehler, G. Nagy, G. Schneider, A. Johann, G. Gottschalk, and J. Hacker.** 2002. Genetic structure and distribution of four pathogenicity islands (PAI I$_{536}$ to PAI IV$_{536}$) of uropathogenic *Escherichia coli* strain 536. Infect. Immun. **70:**6365–6372.

10. **Gophna, U., M. Barlev, R. Seijffers, T. A. Oelschlager, J. Hacker, and E. Z. Ron.** 2001. Curli fibers mediate internalization of *Escherichia coli* by eukaryotic cells. Infect. Immun. **69:**2659–2665.

11. **Gophna, U., T. A. Oelschlaeger, J. Hacker, and E. Z. Ron.** 2001. Yersinia HPI in septicemic *Escherichia coli* strains isolated from diverse hosts. FEMS Microbiol. Lett. **196:**57–60.

12. **Hacker, J., G. Blum-Oehler, I. Muhldorfer, and H. Tschape.** 1997. Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. Mol. Microbiol. **23:**1089–1097.

13. **Hacker, J., and J. B. Kaper.** 1999. The concept of pathogenicity islands, p. 1–11. *In* J. B. Kaper and J. Hacker (ed.), Pathogenicity islands and other mobile virulence elements. ASM Press, Washington, D.C.

14. **Janka, A., M. Bielaszewska, U. Dobrindt, and H. Karch.** 2002. Identification and distribution of the enterohemorrhagic *Escherichia coli* factor for adherence (*efa1*) gene in sorbitol-fermenting *Escherichia coli* O157. Int. J. Med. Microbiol. **292:**207–214.

15. **Janke, B., U. Dobrindt, J. Hacker, and G. Blum-Oehler.** 2001. A subtractive hybridisation analysis of genomic differences between the uropathogenic *E. coli* strain 536 and the *E. coli* K-12 strain MG1655. FEMS Microbiol. Lett. **199:**61–66.

16. **Jin, Q., Z. Yuan, J. Xu, Y. Wang, Y. Shen, W. Lu, J. Wang, H. Liu, J. Yang, F. Yang, X. Zhang, J. Zhang, G. Yang, H. Wu, D. Qu, J. Dong, L. Sun, Y. Xue, A. Zhao, Y. Gao, J. Zhu, B. Kan, K. Ding, S. Chen, H. Cheng, Z. Yao, B. He, R. Chen, D. Ma, B. Qiang, Y. Wen, Y. Hou, and J. Yu.** 2002. Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity through comparison with genomes of *Escherichia coli* K12 and O157. Nucleic Acids Res. **30:**4432–4441.

17. **Knobl, T., M. R. Baccaro, A. M. Moreno, T. A. Gomes, M. A. Vieira, C. S. Ferreira, and A. J. Ferreira.** 2001. Virulence properties of *Escherichia coli* isolated from ostriches with respiratory disease. Vet. Microbiol. **83:**71–80.

18. **La Ragione, R. M., A. R. Sayers, and M. J. Woodward.** 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.

19. **Meier, C., T. A. Oelschlaeger, H. Merkert, T. K. Korhonen, and J. Hacker.** 1996. Ability of *Escherichia coli* isolates that cause meningitis in newborns to invade epithelial and endothelial cells. Infect. Immun. **64:**2391–2399.

20. **Perna, N. T., G. Plunkett III, V. Burland, B. Mau, J. D. Glasner, D. J. Rose, G. F. Mayhew, P. S. Evans, J. Gregor, H. A. Kirkpatrick, G. Posfai, J. Hackett, S. Klink, A. Boutin, Y. Shao, L. Miller, E. J. Grotbeck, N. W. Davis, A. Lim, E. T. Dimalanta, K. D. Potamousis, J. Apodaca, T. S. Anantharaman, J. Lin, G. Yen, D. C. Schwartz, R. A. Welch, and F. R. Blattner.** 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature **409:**529–533.

21. **Reidle, J.** 1999. Methods and strategies for the detection of bacterial virulence factors associated with pathogenicity islands, plasmids, and bacteriophages, p. 13–32. *In* J. B. Kaper and J. Hacker (ed.), Pathogenicity islands and other mobile virulence elements. ASM Press, Washington, D.C.

22. **Ron, E. Z., Z. Yerushalmi, and M. W. Naveh.** 1990. Adherence pili of *Escherichia coli* O78 determine host and tissue specificity. *In* S. Rottem (ed.), Microbial surface components and toxins in relation to pathogenesis. Plenum Press, London, United Kingdom.

23. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

24. **Stocki, S. L., L. A. Babiuk, N. A. Rawlyk, A. A. Potter, and B. J. Allan.** 2002. Identification of genomic differences between *Escherichia coli* strains pathogenic for poultry and *E. coli* K-12 MG1655 using suppression subtractive hybridization analysis. Microb. Pathog. **33:**289–298.

25. **Tschape, H., and J. Hacker.** 1991. Virulence patterns and long-range genetic mapping of extraintestinal *Escherichia coli* K1, K5, and K100 isolates: use of pulsed-field gel electrophoresis. Infect. Immun. **59:**2664–2672.

26. **Walker, J. C., and N. K. Verma.** 2002. Identification of a putative pathogenicity island in *Shigella flexneri* using subtractive hybridisation of the *S. flexneri* and *Escherichia coli* genomes. FEMS Microbiol. Lett. **213:**257–264.

27. **Welch, R. A., V. Burland, G. Plunkett III, P. Redford, P. Roesch, D. Rasko, E. L. Buckles, S. R. Liou, A. Boutin, J. Hackett, D. Stroud, G. F. Mayhew, D. J. Rose, S. Zhou, D. C. Schwartz, N. T. Perna, H. L. Mobley, M. S. Donnenberg, and F. R. Blattner.** 2002. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. Proc. Natl. Acad. Sci. USA **99:**17020–17024.

28. **Yerushalmi, Z., N. I. Smorodinsky, M. W. Naveh, and E. Z. Ron.** 1990. Adherence pili of avian strains of *Escherichia coli* O78. Infect. Immun. **58:**1129–1131.

29. **Zhang, Y. L., C. T. Ong, and K. Y. Leung.** 2000. Molecular analysis of genetic differences between virulent and avirulent strains of *Aeromonas hydrophila* isolated from diseased fish. Microbiology **146**(Pt 4)**:**999–1009.

30. **Zingler, G., G. Blum, U. Falkenhagen, I. Orskov, F. Orskov, J. Hacker, and M. Ott.** 1993. Clonal differentiation of uropathogenic *Escherichia coli* isolates of serotype O6:K5 by fimbrial antigen typing and DNA long-range mapping techniques. Med. Microbiol. Immunol. **182:**13–24.