Distribution of the Serine Protease Autotransporters of the Enterobacteriaceae among Extraintestinal Clinical Isolates of Escherichia coli

Nick Parham, Samantha Pollard, M. Desvaux, Anthony Scott-Tucker, Chengjie Liu, Amanda Fivian, Ian Henderson

JOURNAL OF CLINICAL MICROBIOLOGY, Aug. 2005, p. 4076–4082 Vol. 43, No. 8
0095-1137/05/$08.00+0 doi:10.1128/JCM.43.8.4076–4082.2005

Distribution of the Serine Protease Autotransporters of the *Enterobacteriaceae* among Extraintestinal Clinical Isolates of *Escherichia coli*

Nick J. Parham, Samantha J. Pollard, Mickaël Desvaux, Anthony Scott-Tucker, Chengjie Liu, Amanda Fivian, and Ian R. Henderson*

Bacterial Pathogenesis and Genomics Unit, Division of Immunity and Infection, The Medical School, University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom

Received 21 January 2005/Returned for modification 9 March 2005/Accepted 18 April 2005

Urinary tract infections continue to be among the most common extraintestinal diseases. Cystitis in women is by far the most common urinary tract infection; pyelonephritis in both sexes and prostatitis in men are more severe but less frequent complaints. *Escherichia coli* is by far the most common cause of urinary tract infection. It is believed that uropathogenic *E. coli* is adept at colonizing the urinary tract via the production of specific virulence factors. Recently, a novel virulence determinant, Vat, was described for the prototypical uropathogenic *E. coli* strain CFT073. Vat is a member of the SPATE (serine protease autotransporters of the *Enterobacteriaceae*) subfamily of the autotransporters. Previously, SPATEs have been described for all pathovars of *E. coli*, but until recently their presence had been noticeably absent in nonpathogenic *E. coli*. In this report we describe the prevalence and phylogenetic distribution of the SPATEs among uropathogenic *E. coli* and the ECOR collection, demonstrating an association between the presence of the SPATEs, including Vat, and uropathogenic *E. coli* phylogroups. In addition, we describe the distribution of SPATEs among nonpathogenic *E. coli*.

Escherichia coli is the predominant cause of urinary tract infections (UTIs), accounting for 50% of all nosocomial urinary tract infections and 90% of infections among ambulatory patients (6). Among UTIs, cystitis in women is by far the most common, with the more severe complaints pyelonephritis and prostatitis being less frequent (34). The production of specific virulence factors is thought to be necessary for uropathogenic *E. coli* (UPEC) to colonize the urinary tract (6, 8). These factors include adhesins, capsule, aerobactin, toxins, and proteases, which may aid attachment to host mucosal tissues, allow evasion of immune defenses, and promote invasion of the normally sterile urinary tract and tissues (1, 8, 24, 32). UPEC strains have a greater prevalence of virulence factors than commensal *E. coli*, with virulence determinants being more prevalent among strains causing invasive rather than noninvasive disease (13, 22, 32). However, half of all UPEC isolates possess none, or only one, of the virulence factors characterized thus far, and hence other as yet uncharacterized bacterial factors may be important in the pathogenesis of UTI (22).

The genome sequence of the UPEC strain CFT073 has been completed, and this has propagated the identification of further potential virulence genes (35). Autotransporters are one category of secreted proteins implicated in the virulence of UPEC (17, 18). Members of the serine protease autotransporters of the *Enterobacteriaceae* (SPATE) family are proteins from *E. coli* and *Shigella* spp. which, like the immunoglobulin A1 (IgA1) proteases and Hap autotransporters of *Neisseria* and *Haemophilus* spp., possess a consensus serine protease motif (19). Since the description of the first SPATE (30), a number of investigators have described SPATE proteins in the different pathotypes of *E. coli* and in *Shigella* (2, 9, 12, 16, 27). SPATE proteins possess several common features: (i) all possess an unusual extended signal sequence, (ii) the serine protease active site is reminiscent of a chymotrypsin clan protease, (iii) unlike the IgA1 proteases, none of the SPATE family has been shown to cleave IgA1, (iv) the serine protease motif of SPATE proteins does not have a role in cleavage of the passenger domain from the β-domain, (v) the point of cleavage of passenger domains from the β-domains is identical, (vi) each SPATE member is among the predominant secreted proteins of their respective pathogens, and (vii) SPATEs are associated with pathogenic strains (18). While the full contribution of these proteins to pathogenesis remains elusive, and no universal contribution has been suggested, specific phenotypes have been reported for various members of the SPATE family (7, 17).

Several reports have identified multiple autotransporters in single strains of *E. coli* (27, 31). Hence, we used in silico analyses of the *E. coli* CFT073 genome sequence to identify uncharacterized autotransporter proteins that might be associated with uropathogenesis. We recently documented the presence of 10 members of the autotransporter family in UPEC CFT073, including three SPATEs: Sat, PicU, and Vat (27). Sat is a toxin which has vacuolating cytotoxic activity against bladder and kidney cells (12, 14). PicU, a homologue of the Pic autotransporter identified in *Shigella flexneri* and enteroaggregative *E. coli* (EAEC), demonstrates mucinase activity (27). Vat, the vacuolating autotransporter toxin recently

* Corresponding author. Mailing address: Bacterial Pathogenesis and Genomics Unit, Division of Immunity and Infection, The Medical School, University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom. Phone: 44 (0) 121-414-4368. Fax: 44 (0) 121-414-3599. E-mail: I.R.Henderson@bham.ac.uk.

described for avian pathogenic *E. coli* (APEC), causes vacuolization of chicken embryo fibroblasts in vitro and was shown to be vital for virulence of APEC strain Ec222 (28). The prolific existence of members of the SPATE subfamily of autotransporters in the prototypical UPEC strain CFT073 suggests that the ability of *E. coli* to cause UTI may be associated with the presence of SPATE proteins. Hence, we sought to screen a collection of UPEC isolates that caused invasive or noninvasive UTIs and the ECOR collection of *E. coli* strains in order to determine the association of SPATE autotransporters, and in particular Vat, with clinical disease.

MATERIALS AND METHODS

Bacterial strains and growth conditions. The prototypical UPEC strain CFT073 was isolated from the blood of a woman with pyelonephritis (23). Ninety-four *E. coli* strains isolated from clinical cases of cystitis or pyelonephritis in women, or prostatitis in men, were obtained from J. Ruiz (Barcelona, Spain). These strains were previously characterized for the prevalence of nine uropathogenic virulence factors (32). The ECOR collection, a phylogenetically characterized collection of *E. coli* strains, which consists of 10 human UTI isolates, 1 asymptomatic bacteriuria human isolate, and 61 fecal isolates from healthy human and animal hosts representing all of the phylogenetic groups, was described previously (25). A collection of 43 *E. coli* blood culture isolates were obtained from the clinical laboratories at the Queen Elizabeth Hospital, Birmingham, United Kingdom, in 2003. All strains were grown overnight at 37°C on Luria-Bertani (LB) agar plates to check purity. Overnight LB broth cultures were then used for DNA preparations using the DNeasy tissue kit (QIAGEN, Crawley, United Kingdom) following the manufacturer's instructions.

Phylogenetic grouping of the blood culture isolates was performed using the method described by Clermont et al. (4). Briefly, DNA from each strain was amplified with primers for *chuA* and positive samples were subsequently amplified with primers corresponding to *yjaA*, whereas negative samples were amplified with primers corresponding to *tspE4C2*; samples positive for *yjaA* represent the B2 cluster, samples negative for *yjaA* represent the D group, samples positive for *tspE4C2* represent the B1 cluster, and samples negative for *tspE4C2* represent the A group.

Bioinformatic analysis of the CFT073 genome. To characterize the genomic context of *vat*, the *E. coli* CFT073 genome sequence was compared to other *E. coli* and *Shigella* sp. strains using *coli*BASE, an online database for *E. coli* comparative genomics (3). This contains all the complete *E. coli*, *Shigella*, and *Salmonella* genome sequences, together with preliminary data from a number of sequencing projects currently in progress at The Wellcome Trust Sanger Institute. The database contains precalculated genome alignments performed using MUMmer and PROmer and provides user-friendly tools to display pairwise comparisons between equivalent regions of different strains and to highlight chromosomal insertions, deletions, and rearrangements.

PCR detection of SPATEs. PCR primers for *vat* (5′-GAACACAGTTCATCT GATCTCC-3′ and 5′-GAATATATCAAATTGGTCCCCC-3′) were designed against open reading frame (ORF) c0393 (accession no. AAN78874) of the uropathogenic *E. coli* strain CFT073. The resultant amplicon was 419 bp in length and corresponds to nucleotide positions 375551 to 375133 of the CFT073 genome sequence. PCR primers to detect genes encoding SPATE proteins (5′-YAAYYTNAAYAARMGNATGGG-3′ and 5′-RTTRTAYYTMCCRAANGC NGA-3′) were designed against conserved regions of the β-domain as determined through multiple alignments; the resultant amplicon was 793 bp in length. Genomic DNA preparations were tested for *vat* and for SPATE-encoding genes in 50-μl PCR mixtures containing 15 pmol of each of the forward and reverse primers, 10 nmol of each deoxynucleoside triphosphate, 1 U of *Taq* DNA polymerase (Invitrogen, Paisley, United Kingdom), and 2 mM $MgCl_2$ in 1× PCR buffer (Invitrogen). The PCR conditions were as follows: initial incubation at 94°C for 2 min, followed by 30 cycles of denaturation at 94°C for 15 s, annealing at 55°C for 30 s, and extension at 72°C for 45 s. A final extension step of 72°C for 10 min was also included. Products were analyzed by electrophoresis through 0.8% (wt/vol) agarose gels with ethidium bromide as a visualization agent. The sizes of amplicons were determined by comparison to the 1-kb plus DNA ladder (Invitrogen). *E. coli* K12 (MG1655) and UPEC CFT073 were used as negative and positive controls, respectively. DNA preparations that were initially found to be negative for *vat* or SPATE-encoding genes were retested using different DNA concentrations and annealing at 50°C in order to allow for small variations in primer-site sequence between strains. The prevalence data was analyzed online (http://www.matforsk.no/ola/fisher.htm) using a two-tailed Fisher's exact test to test for statistical significance.

RESULTS

Genomic context of *vat*. Previous in silico analyses of UPEC CFT073 identified an ORF encoding the 148-kDa Vat protein (27). BLAST analyses revealed that Vat was homologous to a number of SPATE proteins from several different pathogenic strains of *E. coli* and was virtually identical at the nucleotide level (99.9%) to the Vat identified in an avian pathogenic strain of *E. coli* Ec222 (28). In UPEC CFT073, Vat was previously annotated as the SPATE proteins hemoglobin protease (Hbp) and Tsh (15, 35); however, closer inspection revealed that UPEC CFT073 Vat was in fact a separate protein and displayed only 78% identity with these two proteins. Like the other SPATEs, Vat possesses a consensus serine protease motif (G^{258}DSGSP) which is characteristic of the SPATE subfamily of autotransporters, the IgA1 proteases of *Neisseria*, and Hap of *Haemophilus*. Indeed, examination of the Vat predicted amino acid sequence revealed conserved histidine residues (H^{130}) and aspartic acid (D^{158}) which, with the conserved serine residue of the consensus region, are presumed to form the catalytic triad of an enzyme active site reminiscent of the SA (chymotrypsin) clan of serine proteases (10).

The *vat* gene is present on a 7.8-kb insertion relative to the *E. coli* K12 MG1655 genome sequence (Fig. 1). The *vat* insertion is found at positions 370301 to 378080 of the *E. coli* CFT073 genome and is inserted between the genes *proA* and *yagU*, which in *E. coli* K12 contains a 39.9-kb region (262172 to 302124), rich in insertion and phage-like elements and genes of unknown function but which is not present in *E. coli* CFT073 (Fig. 1). Analyses of several other strains of *E. coli* revealed that *vat* was present in the same position in UPEC strain 536, APEC Ec222, and the neonatal meningitis *E. coli* strain RS218. Interestingly, both UPEC 536 and APEC Ec222 possessed additional DNA inserts which were not homologous. The additional sequence in UPEC 536 encoded the S-fimbriae associated with binding to sialic acid moieties on eukaryotic cells, the *iro* locus associated with iron acquisition, and an additional autotransporter termed Antigen 43; the additional sequence in APEC Ec222 was homologous to insertion and phage-like elements and genes of unknown function (5, 28). Examination of this region in the diarrheagenic strains of *E. coli* whose genomes have been completed revealed all strains contained an insertion between *proA* and *yagU*. Surprisingly, with the exception of the enteroaggregative *E. coli* strain 042 and the enterohemorrhagic *E. coli* strain EDL933, which share seven conserved genes, the sequences in this region were dissimilar when compared to each other and when compared to the extraintestinal strains of *E. coli*.

Distribution of SPATEs among the ECOR collection. *E. coli* is a highly clonal species represented by four major phylogenetic groups, i.e., A, B1, B2, and D (6, 25). To determine the phylogenetic distribution of *vat*, and specifically if it was preferentially associated with the extraintestinal phylogenetic groups (B2 and D), we used PCR with primers specific for the *vat* gene to survey the well-defined ECOR collection. Of the 57 non-B2 strains, 1 (ECOR23) was found to be positive for *vat*. We previously demonstrated that ECOR23, an elephantine

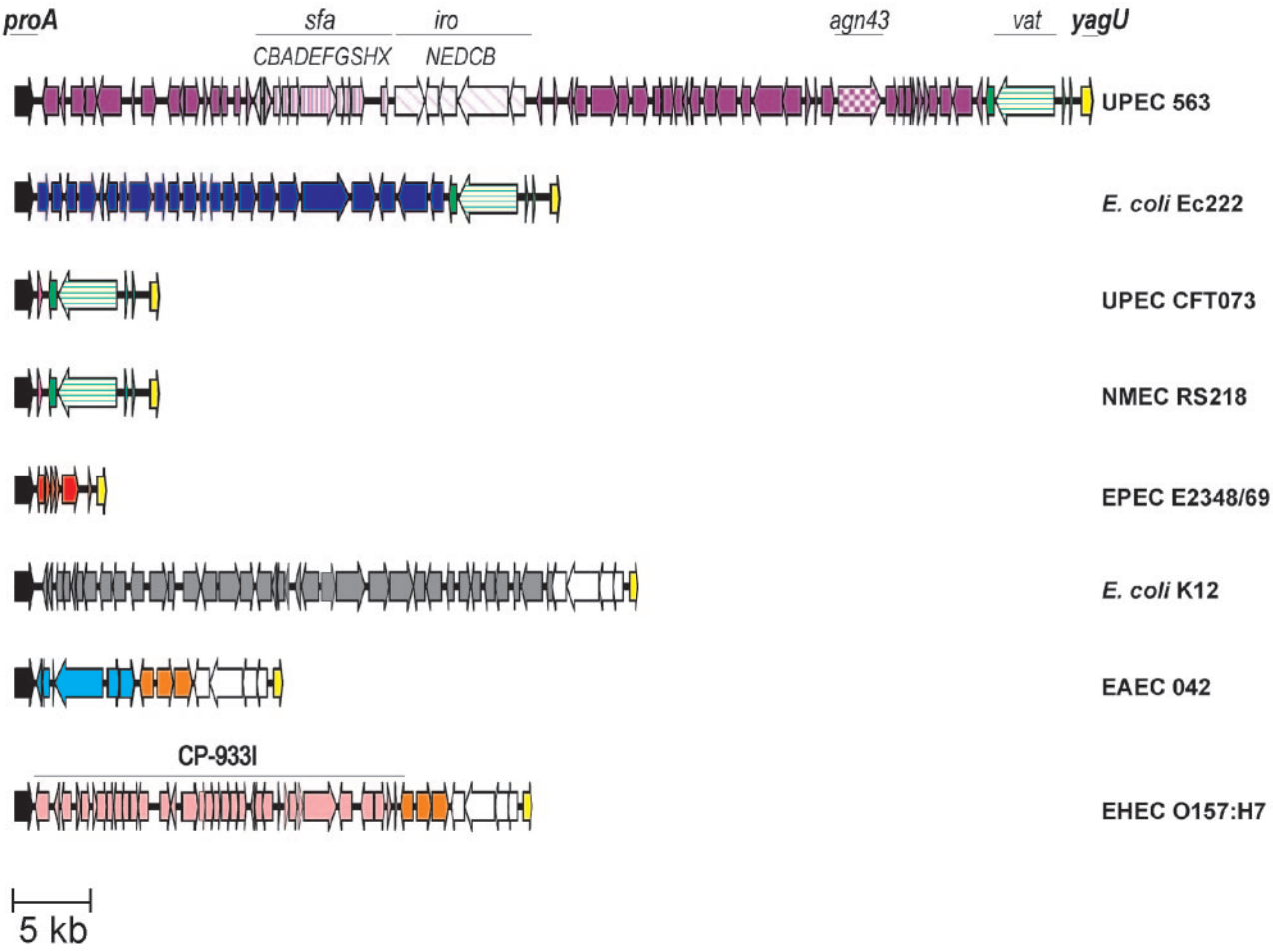

FIG. 1. Genetic organization of the *proA-yagU* region of *E. coli*. The gene organization and content of the *proA-yagU* region of genome-sequenced *E. coli* strains is depicted. The *vat* gene is found in the same position and orientation in all the extraintestinal pathogens but is absent from the intestinal pathogenic isolates. With the exception of several genes conserved among EAEC 042, EHEC EDL933, and *E. coli* K12, the genetic content of each island is remarkably distinct, indicating that this region is a hotspot for recombination. Loci shared between pathogenicity islands are highlighted in the same color. Loci previously implicated in pathogenesis are shaded and labeled. UPEC, uropathogenic *E. coli*; NMEC, neonatal meningitis *E. coli*; EPEC, enteropathogenic *E. coli*; EHEC, enterohaemorrhagic *E. coli*; EAEC, enteroaggregative *E. coli*.

isolate of group A, has an uncharacteristic complement of UPEC-related virulence genes for this phylogenetic group (26). The prevalence of *vat* in the non-B2 group (1.8%) was significantly lower ($P < 2 \times 10^{-10}$) than for the B2 group (86.7%; Fig. 2).

To determine whether SPATE proteins in general were preferentially associated with the extraintestinal phylogenetic groups, we used PCR to determine the distribution of SPATE-encoding genes among the ECOR collection. Using a set of primers corresponding to conserved regions of the β-domain, no product could be detected in reactions with *E. coli* K-12 genomic DNA, but as expected, the presence of an appropriately sized PCR product could be detected in amplifications with *E. coli* CFT073 DNA. Similarly, products corresponding to genes encoding SPATE proteins were detected in the B2 and D groups and a subgroup of the A phylogenetic branches (Fig. 2). Nevertheless, the distribution of SPATE proteins was significantly higher ($P < 0.00001$) in the B2 group (93.3%) than in the non-B2 clusters (29.8%; Fig. 2).

Prevalence of SPATEs among extraintestinal clinical isolates of *E. coli*. As the SPATEs were associated with extraintestinal phylotypes, it was of interest to determine whether the SPATEs were associated with a particular UTI; thus, the presence of genes encoding SPATEs were detected by PCR in a collection of pathogenic *E. coli* strains isolated from cystitis, pyelonephritis, and prostatitis (Fig. 3). No statistically significant difference was observed between any of the groups of isolates. Similarly, PCR was used to test the collections of pathogenic *E. coli* isolates for the presence of *vat*. The distribution of *vat* was similar in cystitis (57.9%) and pyelonephritis (59.3%) isolates. While the prevalence of *vat* was higher in prostatitis isolates (72.4%), the increase was not significant when compared to cystitis or pyelonephritis strains. Furthermore, the distribution of the SPATEs and *vat* occurs at a similar rate in all the clinical groups when compared to the B2 phylogenetic cluster.

To investigate whether the SPATEs or *vat* was associated with *E. coli* strains which possess the capacity to invade the

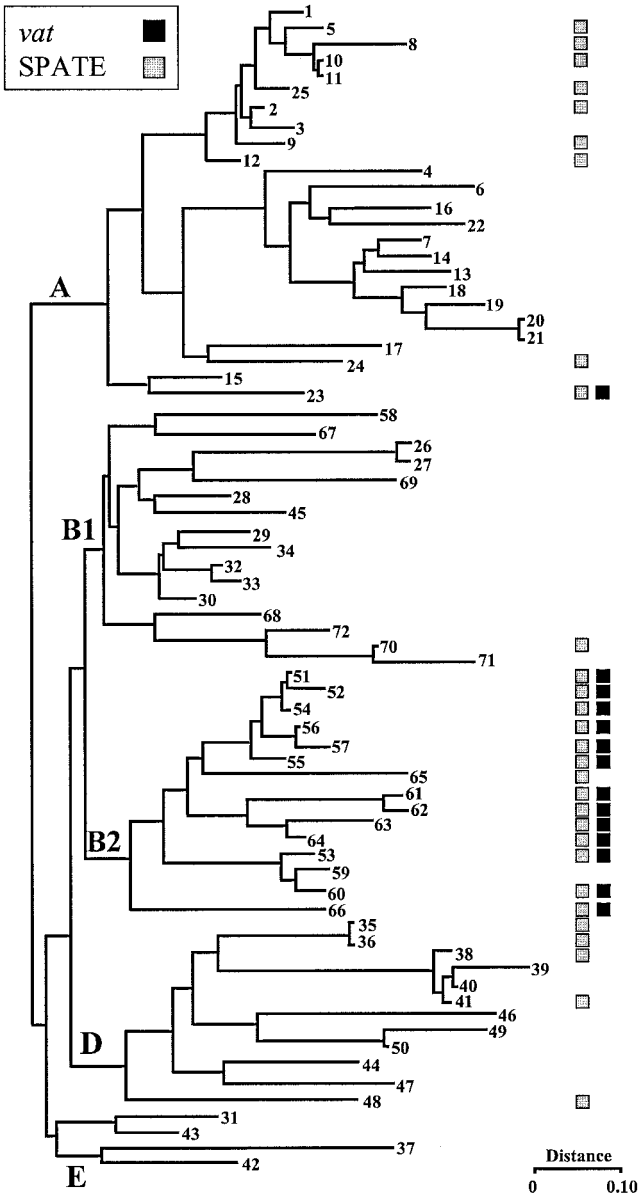

FIG. 2. Distribution of SPATEs among the ECOR collection. Phylogenetic tree of the ECOR isolates showing the distribution of *vat* and SPATE-encoding genes where each locus is represented by a shaded box as indicated in the figure. The number of the ECOR isolate is given in boldface, and each of the major phylogenetic branches is indicated. The *vat* gene is preferentially associated with the B2 phylogenetic cluster, whereas SPATE-encoding genes show a wider distribution. The figure was adapted from reference 20.

bloodstream, their distribution in 43 strains of *E. coli* isolated from patients with septicemia was examined by PCR (Fig. 3). The distribution of the SPATE proteins within the septicemia isolates occurred at a frequency similar to the other clinical isolates and the ECOR B2 phylogenetic cluster. Surprisingly, while *vat* demonstrated a prevalence in the septicemia isolates statistically similar to the cystitis and pyelonephritis isolates, it exhibited a lower prevalence than the B2 group and prostatitis isolates ($P < 0.0025$ and $P < 0.008$, respectively). To determine whether the distribution was biased by a population of septicemia isolates which were not representative of the B2 phylogenetic cluster, the phylogenetic grouping was determined for each septicemia isolate using the method described by Clermont et al. (4). This revealed that the population was composed of 26 B2, 4 B1, 4 D, and 4 A isolates. Comparison of the B2 septicemia isolates with the B2 ECOR isolates revealed no statistically significant difference in the prevalence of *vat* or SPATEs. Comparison of the distribution of the *vat* locus among the non-B2 ECOR collection and the non-B2 septicemia isolates revealed no statistical difference in the prevalence of the loci. In contrast, the non-B2 septicemia isolates demonstrated a higher prevalence of SPATEs than the non-B2 ECOR collection (64.7% versus 29.8%; $P < 0.001$).

Phylogenetic analysis of the SPATEs. Previously, Henderson et al. (19) suggested that the SPATE family of autotransporters could be divided into two groups: group A demonstrating cytopathic activity and exemplified by Pet of enteroaggregative *E. coli* (EAEC) and group B exhibiting preference for extracellular targets, as illustrated by Pic of EAEC and *Shigella* spp. (9, 16). Recently, Dutta et al. (7) explored the phylogenetic relationships of a subset of SPATE proteins using split decomposition analyses of the complete passenger domains. Since this initial study several additional members of the SPATE subfamily have been described for *E. coli* and *Salmonella bongorii*, including Vat (28), EspI (5), EatA (29), EpeA (21), EaaA (33), EaaC (33), and Boa (accession no. AY876285; Henderson et al., unpublished results). Thus, we examined the relationships of the larger SPATE family and the functional activities ascribed to each member of the family. Using split decomposition analyses, separate comparisons were made of the functional passenger domains and of the outer membrane transporting β-domains. These analyses revealed different unrooted dendrograms (Fig. 4). The dendrogram generated for the passenger domains revealed evidence of numerous recombination events among the functional moieties of the SPATEs (Fig. 4A) and much less recombination among the transporting units (Fig. 4B). While the original bifurcating phylogenetic pattern proposed by Henderson et al. (19) continues to be valid for the passenger domains, Vat remains the only member of the group B branch of the SPATEs to demonstrate cytopathic activity. Interestingly, however, the same evolutionary pattern does not hold for the β-domains; EpeA and EspP have virtually identical β-domains, yet the corresponding passenger domains reside on separate branches of the bifurcating phylogram (Fig. 4). Furthermore, EaaA and EaaC, two autotransporters from the nonpathogenic strain ECOR9, are the most distantly related when compared with those isolated from pathogenic isolates.

DISCUSSION

Recently we described the existence of several autotransporters in the UPEC strain CFT073, including the SPATE proteins Vat, Sat, and PicU (27). Further investigation revealed that Vat was encoded on a 7.8-kb genomic island inserted at the *thrW* tRNA locus, between the *proA* and *yagU* genes. Examination of all of the genome-sequenced strains of *E. coli* revealed dissimilar sequences of varying sizes in this region, suggesting that this site is a hotspot for recombination. Interestingly, *vat* is present in the same orientation in the

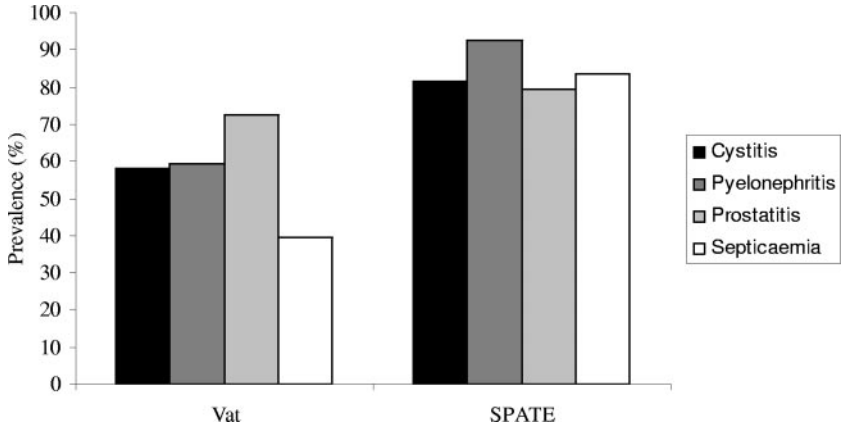

FIG. 3. Prevalence of SPATEs in extraintestinal *E. coli* clinical isolates. The *vat* and SPATE-encoding loci were detected by PCR in clinical isolates of *E. coli*. Prevalence is indicated as a percentage of the total population of strains associated with each clinical syndrome. No statistically significant difference in the distribution of SPATEs was observed among the extraintestinal clinical isolates. The *vat* gene is present at a significantly higher rate in prostatitis isolates when compared to septicemia strains; however, the prevalence of *vat* among the other groups is statistically similar.

proA-yagU region of the extraintestinal pathogens *E. coli* 536, Ec222, and RS218, despite the fact that these strains possess markedly different genetic compositions in this region. Such conservation within a subset of pathogenic strains suggests that Vat might be important in allowing *E. coli* to cause disease. This hypothesis is supported by the fact that Vat was recently demonstrated to be necessary for invasion of host tissues and full virulence of the APEC strain Ec222 (28). Furthermore, Heimer et al. (15) demonstrated by reverse transcription-PCR that *vat* was expressed in mice during experimental urinary tract infection though the contribution of Vat to infection in this model was not reported.

In an effort to determine whether *vat* was associated with extraintestinal pathotypes we examined the ECOR collection of *E. coli* strains. The majority of extraintestinal pathogenic *E. coli* strains, including those with the most robust virulence factor repertoires and those most able to infect the noncompromised host, are found within the B2 and D phylogenetic groups. In contrast, diarrheagenic pathotypes occur almost exclusively in the A, B1, D, and ungrouped phylogenetic clusters (6). In accordance with our initial hypothesis, *vat* was found to be significantly more prevalent in the B2 cluster, suggesting an association with extraintestinal disease. However, screening of a collection of extraintestinal pathogenic *E. coli* isolates revealed that *vat* was prevalent in a statistically similar distribution among cystitis, pyelonephritis, prostatitis, and septicemia isolates despite the discovery of *vat* in several strains of *E. coli* with the ability to invade the bloodstream: CFT073, RS218, and Ec222. While these data suggest that the presence of *vat* does not contribute to invasive disease, they do not rule out the possibility that Vat contributes to initial colonization steps which subsequently lead to UTI and septicemia. Interestingly, while most extraintestinal *E. coli* clinical isolates are derived from the B2 phylogroup, not all members of the B2 phylogroup cause extraintestinal disease, thus *vat* may be represented more frequently among pathogenic isolates than this data suggests. Furthermore, the existence of multiple SPATEs suggests an element of functional redundancy such that the absence of Vat may be compensated for by the presence of other SPATEs.

Investigation of the ECOR collection revealed that SPATE proteins were clustered in the B2 and D groups and a subgroup of the A phylogenetic branches (Fig. 3). The absence of the SPATEs from the majority of the B1 strains, and the remainder of the A subgroup, appears unusual due to the fact that at least one SPATE, and in many cases several, have been identified in all of the diarrheagenic and extraintestinal pathovars of *E. coli* (17, 18). To date no SPATE has yet been characterized in a nonpathogenic organism. Interestingly, many of the strains comprising the ECOR collection have been isolated from healthy people or animals (25), which demonstrates that SPATE proteins are not exclusively associated with pathogens and therefore may not play an exclusive role in virulence. The existence of SPATEs in nonpathogenic *E. coli* is further supported by the recent description of the vacuolating toxin Sat in the probiotic strain *E. coli* Nissle 1917 (11).

Interestingly, despite their high levels of homology, the SPATE proteases demonstrate distinct substrate specificities (7). Previous studies suggested that the SPATE family of autotransporters could be divided into two groups, one demonstrating cytopathic activity, and the other exhibiting preference for extracellular targets and this was confirmed in this study for a larger group of SPATE proteins (7, 19). While these investigations were unable to establish a correlation between the phylogenetic groupings and biological activity, they did reveal evidence of significant homologous recombination among family members. Such recombination, taken in conjunction with the presence of Boa in *S. bongori*, suggests a common ancestral SPATE protein spread by horizontal gene transfer and that each strain that acquired a SPATE has adapted it to its specific niche. Thus, it is plausible that some SPATEs play a role in pathogenesis whereas others have evolved a role in conferring fitness among commensal strains of *E. coli*. Indeed, the most distantly related SPATEs are the EaaA and EaaC proteins of nonpathogenic *E. coli*, whereas those associated with pathogenic strains are more closely related. Furthermore, the statistically significant association of SPATEs with septicemic isolates of the

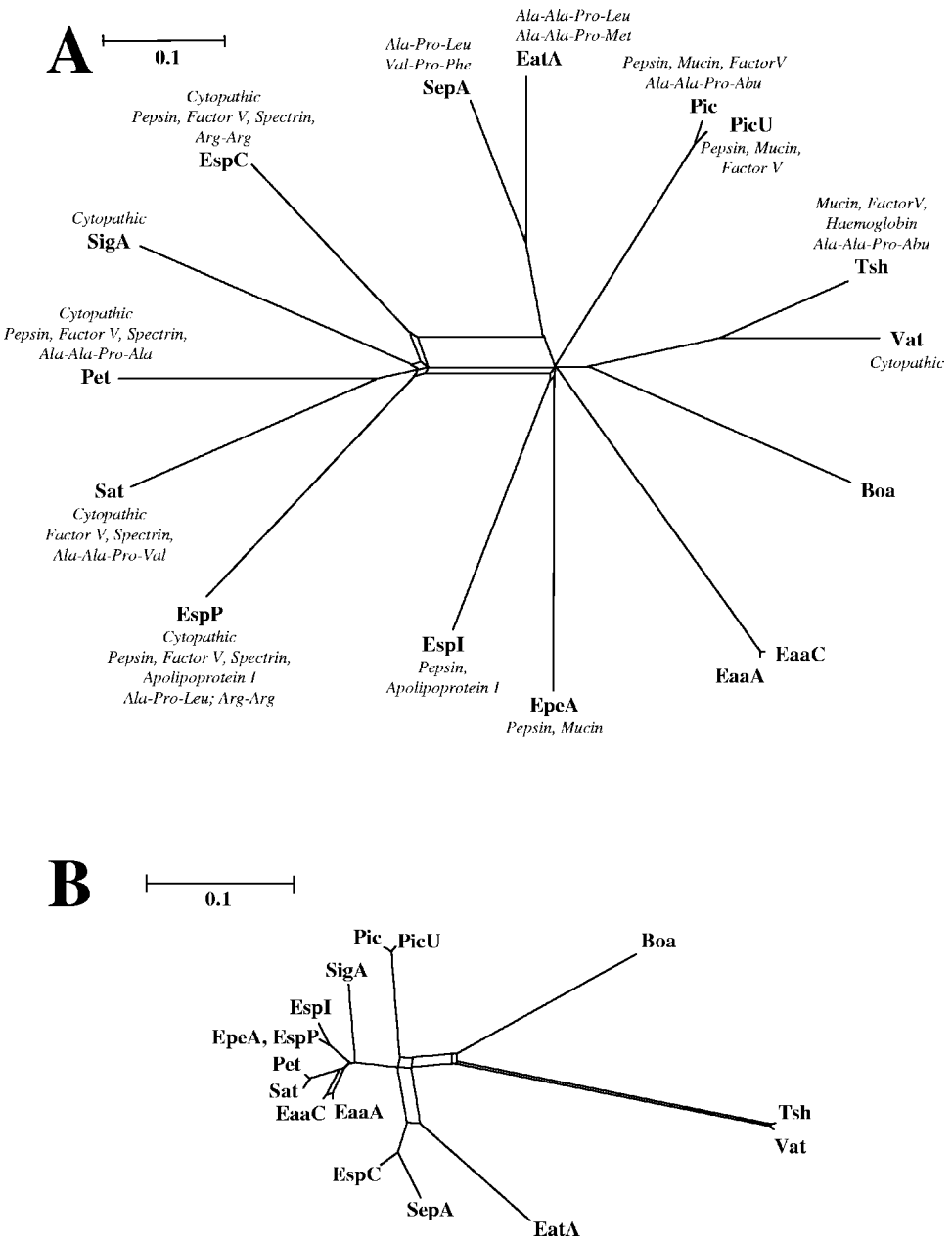

FIG. 4. ClustalX phylograms of amino acid sequence alignments of full-length SPATE passenger domains (A) or the β-domains (B). (A) The known substrates for each of the SPATEs is indicated. In addition, the oligopeptide sequences known to be recognized and cleaved by the SPATEs are also depicted. The bifurcating pattern of SPATE distribution can be observed, where with the exception of Vat the cytopathic SPATEs are found in group A and those identified as extracellular proteases are located in Group B. (B) The β-domains do not have the same phylogenetic pattern and have undergone more restricted recombination events. Trees were further tested for reliability using bootstrap analysis, yielding results of 96.9% (A) and 99.4% (B).

non-B2 phylogenetic groups suggests that in some circumstances SPATEs may contribute to pathogenicity.

In summary, we have demonstrated that *vat* is specifically associated with extraintestinal phylogenetic groups of *E. coli* but have not been able to demonstrate a correlation with a specific clinical manifestation. Furthermore, we have shown that SPATE proteins are not confined to pathogenic strains of *E. coli* and thus the full contribution of SPATEs to pathogenesis and the ability of organisms to thrive in a particular niche need further examination.

ACKNOWLEDGMENTS

We thank Guy Plunkett III for access to the prepublication sequence of *E. coli* CFT073 and RS218, Harry Mobley for providing the *E. coli* CFT073 strain, Mark Pallen for facilitating collection of the septicemia isolates, and Joaquim Ruiz for the extraintestinal *E. coli* isolates. Data from several ongoing *E. coli* genome-sequencing projects was made

available by the Wellcome Trust Sanger Institute (www.sanger.ac.uk) and *E. coli* Genome Project (http://www.genome.wisc.edu/).

This work was funded by the Biotechnology and Biological Sciences Research Council (BBSRC) grants 81/D14955 and 81/P14130.

REFERENCES

1. **Bauer, R. J., L. Zhang, B. Foxman, A. Siitonen, M. E. Jantunen, H. Saxen, and C. F. Marrs.** 2002. Molecular epidemiology of 3 putative virulence genes for Escherichia coli urinary tract infection-usp, iha, and iron (E. coli). J. Infect. Dis. **185:**1521–1524.
2. **Benjelloun-Touimi, Z., P. J. Sansonetti, and C. Parsot.** 1995. SepA, the major extracellular protein of Shigella flexneri: autonomous secretion and involvement in tissue invasion. Mol. Microbiol. **17:**123–135.
3. **Chaudhuri, R. R., A. M. Khan, and M. J. Pallen.** 2004. coliBASE: an online database for Escherichia coli, Shigella and Salmonella comparative genomics. Nucleic Acids Res. **32:**D296–D299.
4. **Clermont, O., S. Bonacorsi, and E. Bingen.** 2000. Rapid and simple determination of the *Escherichia coli* phylogenetic group. Appl. Environ. Microbiol. **66:**4555–4558.
5. **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.
6. **Donnenberg, M., and R. Welch.** 1996. Virulence determinants of uropathogenic *Escherichia coli*, p. 135–174. *In* H. Mobley and J. Warren (ed.), Urinary tract infections: molecular pathogenesis and clinical management. ASM Press, Washington, D.C.
7. **Dutta, P. R., R. Cappello, F. Navarro-Garcia, and J. P. Nataro.** 2002. Functional comparison of serine protease autotransporters of *Enterobacteriaceae*. Infect. Immun. **70:**7105–7113.
8. **Emody, L., M. Kerenyi, and G. Nagy.** 2003. Virulence factors of uropathogenic Escherichia coli. Int. J. Antimicrob. Agents **22**(Suppl. 2)**:**29–33.
9. **Eslava, C., F. Navarro-Garcia, J. R. Czeczulin, I. R. Henderson, A. Cravioto, and J. P. Nataro.** 1998. Pet, an autotransporter enterotoxin from enteroaggregative *Escherichia coli*. Infect. Immun. **66:**3155–3163.
10. **Fink, D. L., L. D. Cope, E. J. Hansen, and J. W. Geme III.** 2001. The Hemophilus influenzae Hap autotransporter is a chymotrypsin clan serine protease and undergoes autoproteolysis via an intermolecular mechanism. J. Biol. Chem. **276:**39492–39500.
11. **Grozdanov, L., C. Raasch, J. Schulze, U. Sonnenborn, G. Gottschalk, J. Hacker, and U. Dobrindt.** 2004. Analysis of the genome structure of the nonpathogenic probiotic *Escherichia coli* strain Nissle 1917. J. Bacteriol. **186:**5432–5441.
12. **Guyer, D. M., I. R. Henderson, J. P. Nataro, and H. L. Mobley.** 2000. Identification of sat, an autotransporter toxin produced by uropathogenic Escherichia coli. Mol. Microbiol. **38:**53–66.
13. **Guyer, D. M., J. S. Kao, and H. L. Mobley.** 1998. Genomic analysis of a pathogenicity island in uropathogenic *Escherichia coli* CFT073: distribution of homologous sequences among isolates from patients with pyelonephritis, cystitis, and catheter-associated bacteriuria and from fecal samples. Infect. Immun. **66:**4411–4417.
14. **Guyer, D. M., S. Radulovic, F. E. Jones, and H. L. Mobley.** 2002. Sat, the secreted autotransporter toxin of uropathogenic *Escherichia coli*, is a vacuolating cytotoxin for bladder and kidney epithelial cells. Infect. Immun. **70:**4539–4546.
15. **Heimer, S. R., D. A. Rasko, C. V. Lockatell, D. E. Johnson, and H. L. Mobley.** 2004. Autotransporter genes *pic* and *tsh* are associated with *Escherichia coli* strains that cause acute pyelonephritis and are expressed during urinary tract infection. Infect. Immun. **72:**593–597.
16. **Henderson, I. R., J. Czeczulin, C. Eslava, F. Noriega, and J. P. Nataro.** 1999. Characterization of pic, a secreted protease of *Shigella flexneri* and enteroaggregative *Escherichia coli*. Infect. Immun. **67:**5587–5596.
17. **Henderson, I. R., and J. P. Nataro.** 2001. Virulence functions of autotransporter proteins. Infect. Immun. **69:**1231–1243.
18. **Henderson, I. R., F. Navarro-Garcia, M. Desvaux, R. C. Fernandez, and D. Ala'Aldeen.** 2004. Type V protein secretion pathway: the autotransporter story. Microbiol. Mol. Biol. Rev. **68:**692–744.
19. **Henderson, I. R., F. Navarro-Garcia, and J. P. Nataro.** 1998. The great escape: structure and function of the autotransporter proteins. Trends Microbiol. **6:**370–378.
20. **Herzer, P. J., S. Inouye, M. Inouye, and T. S. Whittam.** 1990. Phylogenetic distribution of branched RNA-linked multicopy single-stranded DNA among natural isolates of *Escherichia coli*. J. Bacteriol. **172:**6175–6181.
21. **Leyton, D. L., J. Sloan, R. E. Hill, S. Doughty, and E. L. Hartland.** 2003. Transfer region of pO113 from enterohemorrhagic *Escherichia coli*: similarity with R64 and identification of a novel plasmid-encoded autotransporter, EpeA. Infect. Immun. **71:**6307–6319.
22. **Marrs, C. F., L. Zhang, P. Tallman, S. D. Manning, P. Somsel, P. Raz, R. Colodner, M. E. Jantunen, A. Siitonen, H. Saxen, and B. Foxman.** 2002. Variations in 10 putative uropathogen virulence genes among urinary, faecal and peri-urethral Escherichia coli. J. Med. Microbiol. **51:**138–142.
23. **Mobley, H. L., D. M. Green, A. L. Trifillis, D. E. Johnson, G. R. Chippendale, C. V. Lockatell, B. D. Jones, and J. W. Warren.** 1990. Pyelonephritogenic *Escherichia coli* and killing of cultured human renal proximal tubular epithelial cells: role of hemolysin in some strains. Infect. Immun. **58:**1281–1289.
24. **Morin, M. D., and W. J. Hopkins.** 2002. Identification of virulence genes in uropathogenic Escherichia coli by multiplex polymerase chain reaction and their association with infectivity in mice. Urology **60:**537–541.
25. **Ochman, H., and R. K. Selander.** 1984. Standard reference strains of *Escherichia coli* from natural populations. J. Bacteriol. **157:**690–693.
26. **Parham, N. J., S. Pollard, R. R. Chaudhuri, S. Beatson, M. Desvaux, M. A. Russell, J. Ruiz, A. Fivian, J. Vila, and I. R. Henderson.** 2005. Prevalence of PAI IICFT073 genes among extraintestinal clinical isolates of *Escherichia coli*. J. Clin. Microbiol. **43:**2425–2434.
27. **Parham, N. J., U. Srinivasan, M. Desvaux, B. Foxman, C. F. Marrs, and I. R. Henderson.** 2004. PicU, a second serine protease autotransporter of uropathogenic Escherichia coli. FEMS Microbiol. Lett. **230:**73–83.
28. **Parreira, V. R., and C. L. Gyles.** 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.
29. **Patel, S. K., J. Dotson, K. P. Allen, and J. M. Fleckenstein.** 2004. Identification and molecular characterization of EatA, an autotransporter protein of enterotoxigenic *Escherichia coli*. Infect. Immun. **72:**1786–1794.
30. **Provence, D. L., and R. Curtiss III.** 1994. Isolation and characterization of a gene involved in hemagglutination by an avian pathogenic *Escherichia coli* strain. Infect. Immun. **62:**1369–1380.
31. **Roche, A., J. McFadden, and P. Owen.** 2001. Antigen 43, the major phase-variable protein of the Escherichia coli outer membrane, can exist as a family of proteins encoded by multiple alleles. Microbiology **147:**161–169.
32. **Ruiz, J., K. Simon, J. P. Horcajada, M. Velasco, M. Barranco, G. Roig, A. Moreno-Martinez, J. A. Martinez, T. Jimenez de Anta, J. Mensa, and J. Vila.** 2002. Differences in virulence factors among clinical isolates of *Escherichia coli* causing cystitis and pyelonephritis in women and prostatitis in men. J. Clin. Microbiol. **40:**4445–4449.
33. **Sandt, C. H., and C. W. Hill.** 2000. Four different genes responsible for nonimmune immunoglobulin-binding activities within a single strain of *Escherichia coli*. Infect. Immun. **68:**2205–2214.
34. **Schappert, S. M.** 1999. Ambulatory care visits to physician offices, hospital outpatient departments, and emergency departments: United States, 1997. Vital Health Stat. **13(i-iv):**1–39.
35. **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.