Available online at www.sciencedirect.com
ScienceDirect
International Journal of Medical Microbiology 297 (2007) 83–95
IJMM
www.elsevier.de/ijmm

Characterization of four novel genomic regions of uropathogenic *Escherichia coli* highly associated with the extraintestinal virulent phenotype: A jigsaw puzzle of genetic modules

Liisa Johanna Sorsa, Friederike Feldmann, Kirsten Hildinger, Severin Dufke, Sören Schubert*

Max von Pettenkofer-Institut für Hygiene und Medizinische Mikrobiologie, Marchioninistr. 17, D-81377 München, Germany

Received 23 May 2006; received in revised form 21 September 2006; accepted 21 November 2006

Abstract

Extraintestinal pathogenic *Escherichia coli* (ExPEC) are a major cause of urinary tract infections, sepsis, and neonatal meningitis. A variety of virulence factors in these strains is encoded by mobile genetic elements, such as transposons or pathogenicity islands (PAIs). Using subtractive cloning of ExPEC genomes, we recently detected short DNA fragments, which were significantly associated with the extraintestinal virulent phenotype. In this study, we identified four novel genomic DNA regions of the highly virulent uropathogenic *E. coli* strain JS299 carrying these previously identified DNA fragments. Characterization of the partial sequences of the genomic DNA regions revealed complex DNA arrangements with variable genetic compositions regarding the G+C contents and codon usage patterns. The prevalence of 15 previously uncharacterized genes was determined in a collection of clinical ExPECs and commensal *E. coli* strains by means of DNA microarray analyses. From this, 13 novel DNA sequences were demonstrated to be significantly associated with extraintestinal virulent strains, and thus may represent new virulence traits. Beside genes predicted to play a role in metabolic functions, such as sucrose utilization (*scr*), we identified DNA sequences shared by both ExPEC and enteropathogenic *E. coli* (EPEC). These sequences were significantly more prevalent among ExPECs when compared to commensal *E. coli* isolates. Our results support the idea of a considerable genetic variability among ExPEC strains and suggest that the novel genomic determinants described in this study may contribute to the ExPEC virulence.

Keywords: Genomic variation; Virulence determinant; Extraintestinal pathogenic *Escherichia coli*; DNA microarray

Introduction

Extraintestinal pathogenic *Escherichia coli* (ExPEC) are a diverse group of strains of *E. coli* that infect extraintestinal sites, such as the urinary tract, the bloodstream, the meninges, the peritoneal cavity, or the lungs (Russo and Johnson, 2003). Urinary tract infections (UTIs) are among the most prevalent bacterial infections worldwide, and uropathogenic *E. coli* (UPEC), the major subgroup of ExPEC, accounts for around 90% of all ambulatory UTIs (Foxman and Brown, 2003).

*Corresponding author. Tel.: +49 89 2180 78 202; fax: +49 89 2180 78 207.
E-mail address: schubert@med.uni-muenchen.de (S. Schubert).

1438-4221/$ - see front matter © 2007 Elsevier GmbH. All rights reserved.
doi:10.1016/j.ijmm.2006.11.007

ExPEC strains are characterized by a broad range of virulence factors including adhesins, toxins, and iron accumulation systems, which enable them to colonize mucosal surfaces, avoid or subvert local and systemic host defense mechanisms, scavenge essential nutrients, injure and invade the host, and stimulate an injurious inflammatory response (Hochhut et al., 2005; Johnson et al., 2005; Russo and Johnson, 2003). However, unlike intestinal pathogenic *E. coli*, which have distinctive virulence determinants that result in characteristic diarrheagenic symptoms and signs (Caprioli et al., 2005; Nataro, 2005), ExPEC possess no specific virulence factors or set of factors, which are obligatory for the infection of a certain extraintestinal site.

The ExPEC genomes are highly diverse mosaic structures in permanent flux (Dobrindt et al., 2003). These strains have obtained a significant amount of DNA (predictably up to 25% of the genomes) through acquisition of foreign DNA from diverse related or non-related donor species by means of laterally transferable mobile genetic elements, such as pathogenicity islands (PAIs), plasmids, phages, transposons, and insertion elements (IS) (Dobrindt et al., 2004; Fischer and Eisenberg, 1999; Ochman et al., 2005; Pal et al., 2005). The ability of ExPEC strains to cause a disease is mainly derived from this horizontally acquired gene pool. The extragenous DNA facilitates the rapid adaptation of a pathogen to changing conditions, and hence extents the spectrum of sites that can be infected (Hacker and Carniel, 2001; Lerat et al., 2005; Ochman et al., 2005; Pal et al., 2005). The strain-specific gene fraction is largely distinct from the set of established metabolic and catabolic genes constituting the core genome, and can be differentiated from the latter by both its G+C content and codon usage pattern (Ochman et al., 2005; Pal et al., 2005; Sharp and Li, 1987; Wang et al., 2001).

In a previous study, we applied a suppressive subtractive hybridization strategy to the highly virulent UPEC strain JS299, and detected four DNA fragments significantly associated with the extraintestinal pathogenic phenotype (Sorsa et al., 2004). In the present study, we characterized the DNA loci adjacent to these subtraction fragments. These DNA fragments were found to reside in four separate cosmids (pJS448, pJS666, pJS700, and pJS706). Characterization of the partial cosmid sequences revealed a patchwork of unknown genes inserted into the *E. coli* core genome. In addition, some novel genes were found to be significantly more frequent among ExPEC strains when compared to commensal *E. coli* strains isolated from stool samples. These newly described genes represent new markers for urovirulence and may even contribute to the virulence of the respective strains. The data of this study further emphasize the impact of horizontal transfer on the distribution of virulence factors and show that individual "genetic modules" influence the mosaic organization of the *E. coli* genomes.

Materials and methods

Bacterial strains and plasmids

Cosmids used in this study are listed in Table 1. Bacteria were routinely grown at 37 °C in Luria-Bertani (LB) medium. For plasmid maintenance, LB medium was supplemented with kanamycin (50 μg/ml) and ampicillin (100 μg/ml). The uropathogenic *E. coli* strain CFT073 (GenBank accession No.: NC_004431), the enterohemorrhagic *E. coli* strain EDL933 (EHEC; NC_002655), and the commensal *E. coli* K-12 strain MG1655 (NC_000913) were used as control strains. The genome sequences of these strains are available in the public databases of the National Center for Biotechnology Information (www.ncbi.nih.gov) and of the European Bioinformatics Institute (www.ebi.ac.uk). The *E. coli* strain JS299 (O6:H31) was isolated from a urine sample of a cystitis patient at the diagnostic laboratory of the Max von Pettenkofer-Institut (Munich, Germany). In addition, 88 *E. coli* strains (UPEC, newborn meningitis *E. coli*, and commensal *E. coli*) were also obtained from the Max von Pettenkofer-Institut or were kindly provided by Dr. Timo Korhonen (Department of Biological and Environmental Sciences at the University of Helsinki, Finland). All *E. coli* strains were classified into the four main phylogenetic groups of *E. coli* (A, B1, B2, and D) using the PCR method described by Clermont et al. (2000). From 53 ExPEC strains tested, 38 belonged to the phylogenetic group B2 (PG; 72%), five to the PG B1 (9%), four to the PG D (8%), and six to the PG A (11%). The commensal *E. coli* strains belonged predominantly to group A (16 strains, 43%). Ten commensal *E. coli* strains were classified as PG B1 (27%), six as PG B2 (16%), and five as PG D (14%; see Table 4). Obtained prevalence was in agreement with the previously reported phylogenetic distributions of ExPECs (predominantly B2 and D) and fecal *E. coli* strains (predominantly A and B1) (Clermont et al., 2000; Johnson et al., 2005). The UPEC strain JS299 was shown to belong to the phylogenetic group B2.

Recombinant DNA methods, DNA sequencing, and sequence analysis

DNA manipulations, such as isolation of cosmids and genomic DNA, construction of a cosmid library of the UPEC strain JS299, and selection of cosmids carrying previously identified subtraction fragments were performed using standard techniques (Ausubel et al., 1989). Cosmids were partially sequenced using the double strand sequence walking technique with sequence-derived primers. Co-linearity of the cosmid sequences was ascertained by PCR on the genomic DNA of strain JS299 using primers flanking the characterized cosmid