Available online at www.sciencedirect.com
SCIENCE DIRECT®
Plasmid 54 (2005) 149–159
PLASMID
www.elsevier.com/locate/yplas

The complete sequence and analysis of the large virulence plasmid pSS of *Shigella sonnei*

Yan Jiang, Fan Yang, Xiaobing Zhang, Jian Yang, Lihong Chen, Yongliang Yan, Huan Nie, Zhaohui Xiong, Jing Wang, Jie Dong, Ying Xue, Xingye Xu, Yafang Zhu, Shuxia Chen, Qi Jin *

State Key Laboratory for Molecular Virology and Genetic Engineering, # 6 Rongjing Eastern Street, BDA Beijing 100176, China

Received 4 January 2005, revised 2 March 2005
Available online 20 April 2005
Communicated by Stuart Levy

Abstract

The complete sequence of pSS, which is the large virulence plasmid of *Shigella sonnei*, was determined. The 214-kb plasmid is composed of segments of virulence-associated genes, the O-antigen gene clusters, a range of replication and maintenance genes, and large numbers of insertion sequence (IS) elements. Two hundred and forty-one open reading frames (ORFs) were identified, of which 117 are highly homologous to IS elements or transposases, 57 are homologous to known pathogenesis-associated proteins, and 30 are related to replication, plasmid maintenance, or other metabolic functions. Thirty-seven ORFs have no similarity to proteins with a known function, including two with no significant similarity to any hypothetical proteins. Interestingly, 10 ORFs encoding O-antigen gene clusters were identified on the plasmid and this is markedly different from most other *Shigella* spp. virulent plasmids. A novel toxin–antitoxin system, a series of stbDE homologs, was found on the plasmid immediately downstream of the replication region; the sole segregation stability system may be responsible for the instability of pSS. The pSS plasmid is a mixture of genes with different origins and functions. The sequence suggests a remarkable history of IS-mediated recombination and acquisition of DNA across a range of bacterial species.

Keywords: *Shigella sonnei*; Virulent plasmid; pSS; Sequence; O-Antigen; stbDE

1. Introduction

Bacteria of *Shigella* spp., belonging to the Enterobacteriaceae family, are Gram-negative, facultative aerobic organisms. The virulent strains cause

* Corresponding author. Fax: +86 10 67877736.
E-mail address: zdsys@sina.com (Q. Jin).

0147-619X/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.plasmid.2005.03.002

diseases ranging from diarrhea to bacillary dysentery, characterized by the fever, abdominal cramps, and bloody diarrhea. This continues to be a major health problem worldwide, causing an estimated 1 million deaths and 163 million cases of dysentery annually (Kotloff et al., 1999). The organisms are officially divided into four groups and at least 47 serotypes: *Shigella boydii*, *Shigella dysenteriae*, *Shigella flexneri*, and *Shigella sonnei* (Noriega et al., 1999). In contrast to the other groups, all virulent *S. sonnei* strains comprise a single serotype and have two forms of cells.

The essential mechanism of virulence of *Shigella* spp. resides in its ability to enter susceptible epithelial cells and to induce apoptosis in infected macrophages (Clerc et al., 1986). The virulence genes necessary for invasion of epithelial cells are contained on a large 220-kb plasmid, termed the virulence plasmid or the invasion plasmid (pINV), which is present in all pathogenic strains (Watanabe and Nakamura, 1985). This large plasmid also encodes virulence regulating genes, and genes associated with the replication, maintenance, and segregation of the plasmid.

Most work on the molecular pathogenesis of *Shigella* has been carried out in *S. flexneri*. Several plasmids from different strains of this organism were sequenced, including the virulence plasmid pWR100 of *S. flexneri* 2a (Buchrieser et al., 2000), the virulence plasmid pWR501 of *S. flexneri* 5a (Venkatesan et al., 2001), and the virulence plasmid pCP301 of *S. flexneri* 2a 301 strain (AF386526).

Similar to the other three groups of *Shigella*, the virulence plasmid of *S. sonnei*, designated as pSS, is sufficient for entering, replicating, and disseminating within epithelial cells (Sansonetti et al., 1981; Watanabe and Nakamura, 1985). However, pSS is unstable and tends to be lost at a high frequency, unlike other large unicopy plasmids (Sansonetti et al., 1981). In addition, experiments have shown that an O-antigen gene cluster, commonly found in the chromosomes of *Shigella* spp., is actually contained on this plasmid (Lai et al., 1998; Shepherd et al., 2000). Here, we present the sequence and the initial analysis of the virulent plasmid of *S. sonnei*. The complete sequence of the plasmid may help to understand the gene composition and construction of the plasmid and its role in virulence, metabolism, and gene transfer. Hopefully, this will provide a greater insight into the *S. sonnei* bacterium and allow a better comparison of the virulent plasmids of the different *Shigella* groups and serotypes.

2. Materials and methods

2.1. Bacterial strain and plasmid

The pSS is a large virulence-associated plasmid harbored by *S. sonnei*, which is a clinical strain isolated in the Chinese mainland.

2.2. Library construction and sequencing

The pSS DNA was isolated from *S. sonnei* using a Qiagen plasmid purification kit and fragmented using sonication. DNA fragments ranging from 4 to 6 kb were collected by agarose gel electrophoresis. Purified and blunt-repaired DNA fragments were subcloned into the pUC19 vector. Positive clones were sequenced using BigDye terminator chemistry (PE Applied Biosystems, USA) on an ABI3730 sequencer (Perkin-Elmer). The sequences were assembled using Phred and Phrap software with optimized parameters and the quality score set >20. All sequences were assembled into 56 contigs. The gaps among the contigs were closed by primer walking on linking clones or by sequencing the DNA products generated by regular and long PCR.

2.3. Sequence analysis and annotation

By analyzing codon usage and elements located upstream, ORFs with greater than 50 codons were identified using Glimmer 2.01 software. The predicted protein sequences were compared to sequences of proteins in the database using Blastp (NCBI blast). Significant homology was defined as >30% similarity over 60% of each query sequence, and the cluster of orthologous groups (COGs) of protein database was used to identify families to which the predicted proteins were related. IS sequences were identified by comparing to the known insertion sequences.